Université Pierre et Marie Curie

Ecole Doctorale Complexité du vivant - ED515

Laboratoire de Biologie du Développement de Villefranche-sur-Mer

UMR7009 / Equipe EvoInSiDe

Retinoic acid signaling in chordates:

The evolutionary history of a morphogen-dependent signaling

Par João Emanuel Marques Carvalho

Thèse de doctorat en biologie du développement

Dirigée par Michael Schubert

Présentée et soutenue publiquement le 31 Janvier de 2017

Devant un jury composé de :

Prof. Umbhauer, Muriel	Président
Dr. Saccani, Simona	Rapporteur
Dr. Studer, Michèle	Rapporteur
Dr. Mallo, Moisès	Examinateur
Dr. Schubert, Michael	Directeur de thèse

To Elizabeth, my amazing wife for her support and unconditional love,

to my parents and brother for their inspiring example…

Acknowledgements

The process of obtaining a Ph.D. degree is a long way and most certainly without the many great people that crossed my path it would have been a less enriching experience.

First of all, I would like to thank my supervisor, Michael Schubert, for the last few years of scientific exchanges, challenges, failures and achievements.

I'm also very grateful to the present and past members of the EvoInSiDe team for the friendliness and support since my arrival in multiple different aspects. To François for all the technical help and never-ending shared nights in the lab, to Jeni for her kindness and availability to help me with any kind of problems, to Guy for his incredible knowledge in molecular biology and development, to Elisa, Nico and Laurent for sharing this Ph.D. adventure, to Yoann, and others students, that were of great help during the tormentous amphioxus spawning seasons.

Furthermore, a giant "merci" to all the members of the LBDV. In particular to Laurent and Sophie for all the help building and maintaining the amphioxus facility, to Lydia, Céline, Christian, Philipe and Faisal for always being available to solve and help with any kind of problems, to Christelle, Frédéric and Sofie for all the administrative support, to Evelyn for making possible my initial installation in Villefranche-sur-Mer.

I would also like to thank all the members of the LOV and OOV that participated directly or indirectly in my research work and in more detail all the Ph.D. students that made an effort to establish the J2CO and made it a success.

It goes without saying that a great part of this work wouldn't have been possible without the many collaborators. I would especially like to thank Ferdi, his bioinformatics skills and his constant availability, Hector, Steph and all the amphioxus Banyuls-sur-Mer crew, for proving every year new amphioxus and for hosting me with open arms.

I also would like to thank the ED "Complexité du Vivant" for giving me the opportunity to participate in training courses at the UPMC in Paris. I'm also very grateful to all the great scientists that supported me and openly discussed my project during my Ph.D., in particular the members of my thesis committee, Muriel Umbhauer, Dave Ferrier and Gerrard Benoit, and my thesis jury, Simona Saccani, Michèle Studer and Moisès Mallo.

Last but not least a huge thank you to all my family, in particular my wife, parents, brother, parents-in-law and sister-in-law.

This work was made possible by financial assistance from the FCT (POCH – MCTES; SFRH/BD/86878/2012), the Réseau André Picard and funds from ANR (ANR-09-BLAN-0262-02 and ANR-11-JSV2-002-01) and CNRS to Michael Schubert.

Résumé

L'une des caractéristiques les plus frappantes des animaux multicellulaires, aussi appelés les métazoaires, est leur étonnante diversité morphologique. Des études de type phylogénétique ont permis de mettre en relation cette abondance et cette variété observées au sein des formes de vie animale avec des différences au niveau de processus moléculaires, cellulaires, tissulaires et organiques. Parmi ces différences, celle affectant les programmes de développement apparaît comme un aspect clé de la diversité des métazoaires. Les programmes de développement reposent entre autres sur la mise en œuvre de communications entre cellules, ou entre milieu environnant et cellules, et ces communications sont assurées au niveau moléculaire par le déploiement de cascades d'activités protéiques nommées les voies de signalisation.

Une des voies de signalisation essentielles au cours du développement de nombreux métazoaires est la voie de l'acide rétinoïque (AR). Le fonctionnement et les rôles de cette voie pendant le développement animal ont fait l'objet de nombreuses recherches, ces travaux ayant aboutis à des découvertes majeures. Néanmoins des analyses supplémentaires restent requises afin de mieux comprendre l'histoire évolutive de cette voie, du métabolisme de l'AR à la transmission de son signal, en passant par l'identification de ses gènes cibles, ses interactions avec d'autres voies de signalisation, et ses fonctions au cours du développement, le tout au sein du règne animal. Dans ce contexte, étudier cette voie chez un métazoaire tel que le céphalochordate amphioxus, qui possède une voie de signalisation de l'AR équivalente à celle des vertébrés, mais avec une redondance moléculaire moindre, représente une étape importante pour identifier l'architecture et les fonctions ancestrales de cette voie parmi les chordés.

Chez l'amphioxus, la voie de signalisation de l'AR est étudiée depuis plus de 20 ans, mais peu de choses sont encore connues sur la régulation de la biodisponibilité de l'AR pendant le développement et sur la nature de ses gènes cibles et du réseau régulateur qu'ils définissent. Le but de mon travail de thèse a par conséquent été d'étudier ces deux aspects fondamentaux de la voie de signalisation de l'AR chez l'amphioxus. Au cours de mon projet de recherche, j'ai utilisé comme modèle d'étude l'amphioxus européen, *Branchiostoma lanceolatum*, pour lequel j'ai également mis en œuvre de nombreuses améliorations du

système de culture ainsi que développer des techniques d'analyses *in vivo*, comme la microinjection d'ARNm dans des œufs..

Afin de caractériser le mécanisme de biodisponibilité de l'AR au cours du développement de *B. lanceolatum*, j'ai mené une étude sur les enzymes responsables de la dégradation de l'AR : les protéines de la sous-famille des CYP26. Durant ce travail j'ai décrit la diversification évolutive des gènes CYP26 chez les deutérostomiens ainsi que leur patron d'expression spatiale et temporelle et leur fonction respective. J'ai également mis en évidence l'existence d'une boucle de rétroaction transcriptionnelle positive entre la voie de l'AR et les protéines CYP26, pendant le développement de l'amphioxus.

Lors d'un second projet, j'ai ensuite identifié les gènes cibles de la voie de signalisation de l'AR pendant le développement de *B. lanceolatum*. Pour ce faire, j'ai combiné des traitements pharmacologiques utilisant des rétinoïdes avec deux techniques différentes de séquençage à haut débit : l'ARN-seq qui permet de déterminer à un temps t choisi le catalogue complet de gènes exprimés ainsi que de quantifier leur niveau d'expression, et l'ATAC-seq (Assay for Transposase-Accessible Chromatin) qui permet d'identifier également à un temps t choisi les régions accessibles de la chromatine (c'est à dire les régions ouvertes). En combinant l'ensemble des données obtenues par ces techniques, j'ai ainsi pu révéler un nouvel ensemble de gènes dont la transcription est régulée par la voie de signalisation de l'AR, mais aussi si ces gènes sont des cibles directes ou indirectes de la voie ainsi que de nouvelles boucles de rétroaction transcriptionnelle faisant intervenir spécifiquement RAR, l'un des deux effecteurs nucléaires de la voie. Par ce travail j'ai ainsi établi une méthodologie puissante permettant de caractériser rapidement, à grande échelle et de façon non-biaisée, les réseaux de gènes agissant au cours du développement en amont d'une voie de signalisation, une méthodologie qui pourra plus largement être appliquée à d'autres voies de signalisation et d'autres organismes.

Abstract

One of the most striking features of multicellular animals, the metazoans, is their amazing morphological diversity. Even though phylogenetic research has made remarkable progress towards revealing how the abundance and variety of animal life forms relates on the molecular, cellular, tissue, and organismal level, the alteration of developmental programs has been revealed as a key aspect in this process. During development, a rather limited number of signaling pathways has been shown to be instrumental for generating metazoan diversity. The retinoic acid (RA) signaling pathway is one of these instrumental signaling cascades.

A significant amount of time and work has been used to characterize the functions and roles of RA signaling during development, although further work is required to better understand the evolutionary history of the RA signaling network, from metabolism to signal transduction passing by the interactions with other signaling cascades and its developmental functions and how they evolve with time. In this context, model organisms with representative, vertebrate-like RA signaling cascades, such as the cephalochordate amphioxus, will be an important case-study in order to identify the blueprint of an ancestral RA network.

The amphioxus RA signaling pathway was initially studied about 20 years ago, even though not much is known about the bioavailability of RA during development. Moreover, the target genes of the RA signaling pathway and their hierarchical relationship during amphioxus development represent an interesting open question. Therefore, this work aimed at providing a detailed description of two fundamental aspects of the RA signaling pathway during amphioxus development: (1) the regulation of the bioavailability of RA in the developing embryo and (2) the target genes under the control of the RA signaling pathway together with their hierarchical regulatory relationship. To address these questions, the European amphioxus, *Branchiostoma lanceolatum*, was used as a model system.

During my research project, not only these questions were fundamental, but also the implementation of amphioxus as a reliable model system and thus the establishment of multiple aquaculture improvements as well as in vivo techniques, such as the microinjection of mRNAs into amphioxus eggs. Furthermore, to characterize the bioavailability of RA during development of amphioxus, I focused on the study of the enzymes that mediate the catabolism of RA endogenously, the CYP26 subfamily proteins. I thus described the evolutionary diversification of CYP26 genes in deuterostomes as well as their expression,

their function and the mechanisms that govern the feedback loop controlled directly by RA during amphioxus development.

Additionally, to shed light on the target genes under the control of the RA signaling pathway during amphioxus development, I combined pharmacological treatments using retinoid-specific drugs with two different techniques of high throughput sequencing: RNAseq, that revealed the entire RNA profile and thus the genes being expressed at a given moment in time, and ATACseq (assay for transposase-accessible chromatin) that provided a global overview of accessible regions of the chromatin (i.e. open chromatin regions). By combining the data obtained by these techniques, I revealed a new set of genes that are under the control of the RA signaling pathway as well as new potential regulatory loops driving RAR-mediated expression. Moreover, I established a framework to characterize gene hierarchies during development that can be widely applied to other signaling pathways and organisms.

Table of contents

Introduction

One of the most striking feature of multicellular animal organisms, the metazoans, is their amazing morphological diversity (Fig. 1) that ranges from the small, flattened form of *Trichoplax* and the highly specialized colonial organization seen in the hydrozoan Portuguese man o' war (*Physalia physalis*), passing by the highly marked segmented form of annelids (e.g. *Platynereis dumerilii*) and the pentaradial organization of echinoderms (e.g. crinoids such as *Oxycomantus bennetti*), to some incredible and detailed adaptations made by vertebrates (e.g. *Phycodurus eques* and *Vulpes zerda*). When seen from a body organization perspective, these organisms seem to have little in common and thus it is quite difficult to imagine that all metazoans originated from one unique ancestor in evolution.

Phylogenetic research has made remarkable progress towards revealing how the abundance and variety of animal life forms relates with the molecular, cellular, tissue and organismal level. Accordingly, the alteration of developmental programs has been revealed as a key aspect to the appearance of new characters during evolution. By the means of differentially modulated signal inputs individual cells or cellular subsets are induced into a specific fate during development, ultimately assuming a given set of properties and functions. Despite the huge variety of metazoan cell types, cell organization, specialized tissues and organs, as well as larval and adult forms only a rather limited number of signaling pathways have been shown to be instrumental for generating this diversity. Between the most important of these signaling cascades we find, for instance, the wingless related (WNT), the transforming growth factor-beta (TGF-beta), the hedgehog (HH), the fibroblast growth factor (FGF), the notch delta signaling and the retinoic acid (RA) signaling.

Considerable time and effort has been invested during the past 100 years in the analysis of metabolism and functions of RA during development. It is clear that the signaling cascade controlled by this morphogen functions in various tissues at different time points during embryogenesis. Thus, due to this multitude of biological functions, a comprehensive understanding of the developmental roles played by the RA pathway is still quite unclear. Moreover, the regulation of RA signaling activity during development seems to be extremely complex, since it not only depends on a very fine regulation of endogenous RA availability, but also on the complex expression of receptors and co-factors, all of which are controlled by multiple interactions of RA signaling with other signaling cascades creating a great variety of regulatory loops.

Fig. 1 – Diversity of forms in the metazoans. *Species shown from the top left:* Trichoplax adhaerens, Physalia physalis, Platynereis dumerilii, Oxycomantus bennetti, Phycodurus eques *and* Vulpes zerda.

Further work is thus required to better understand the various aspects of the RA signaling network, from metabolism to signal transduction passing by the interactions with other signaling cascades and its developmental functions. In this context, model organisms characterized by a relatively simple anatomy and lacking genetic redundancy, such as the cephalochordate amphioxus, will be an important case study to identify the blueprint of an ancestral RA network. Expanding the scope of RA-centered research to non-conventional animal models thus holds the potential to uncover the evolutionary origins of the RA signaling network and the ancestral functions of retinoids during development of early metazoan animals.

Overview of the Introduction:

- Chapter I: *Evolutionary and developmental biology and animal phylogeny*
- Chapter II: *Amphioxus as an evo-devo model*
- Chapter III: *Bringing amphioxus to the bench-side*
- Chapter IV: *Retinoic acid*
- Chapter V: *Amphioxus and the study of retinoic acid signalling*

Chapter I: Evolutionary and developmental biology and animal phylogeny

Origins of evolutionary and developmental biology (evo-devo)

Evolution and more concrete ideas about evolution may be found among ancient philosophers as well as in the works of Buffon[1] but the first clear formalization of this concept dates back to Lamarck[2]. The basic idea of Lamarck was that similarity was more easily explained by a common ancestral origin[2]. Lamarck proposed that phenotypic changes, induced in parents, would be passed on to offspring, a mechanism called inheritance of acquired characters[2]. Some time later, Geoffroy St. Hilaire, comparing different animal phyla, tried to establish a single plan of organization and, thus, homologous structures between distinct species[3]. The concept of homology is still a central point in modern evolutionary biology studies. The idea of Evolution was extended three decades later by Darwin with the great and extensive piece of work called "On the origin of species by means of natural selection"[4]. The merit of Darwin, however, was to suggest and argue for a mechanism of evolution, which is natural selection[4]. Still now natural selection, with its multiple aspects, appears as the central mechanism of biological evolution. The evolutionary biology discipline itself only emerged much later with the evolutionary synthesis, or modern synthesis, where contributors from several previously unrelated fields of biological research, including genetics, ecology, systematics and paleontology, reconciled Darwin's theory with the facts of genetics[5].

Globally, evolutionary biology studies the evolutionary processes that explain how life was originated on Earth starting from a single origin of life, and how and when the different groups now observed in nature originated[5]. Such inferences and reconstruction (Fig. I.1) are mainly based on two very different kinds of data: the fossil record and molecular phylogenies. The fossil record itself is composed by rare and isolated deposits, leading to a fragmentation of the paleontological data and thus creating problems in the inference of the ancestral forms. Additionally, the fossilization process is very biased and dependent in the environmental

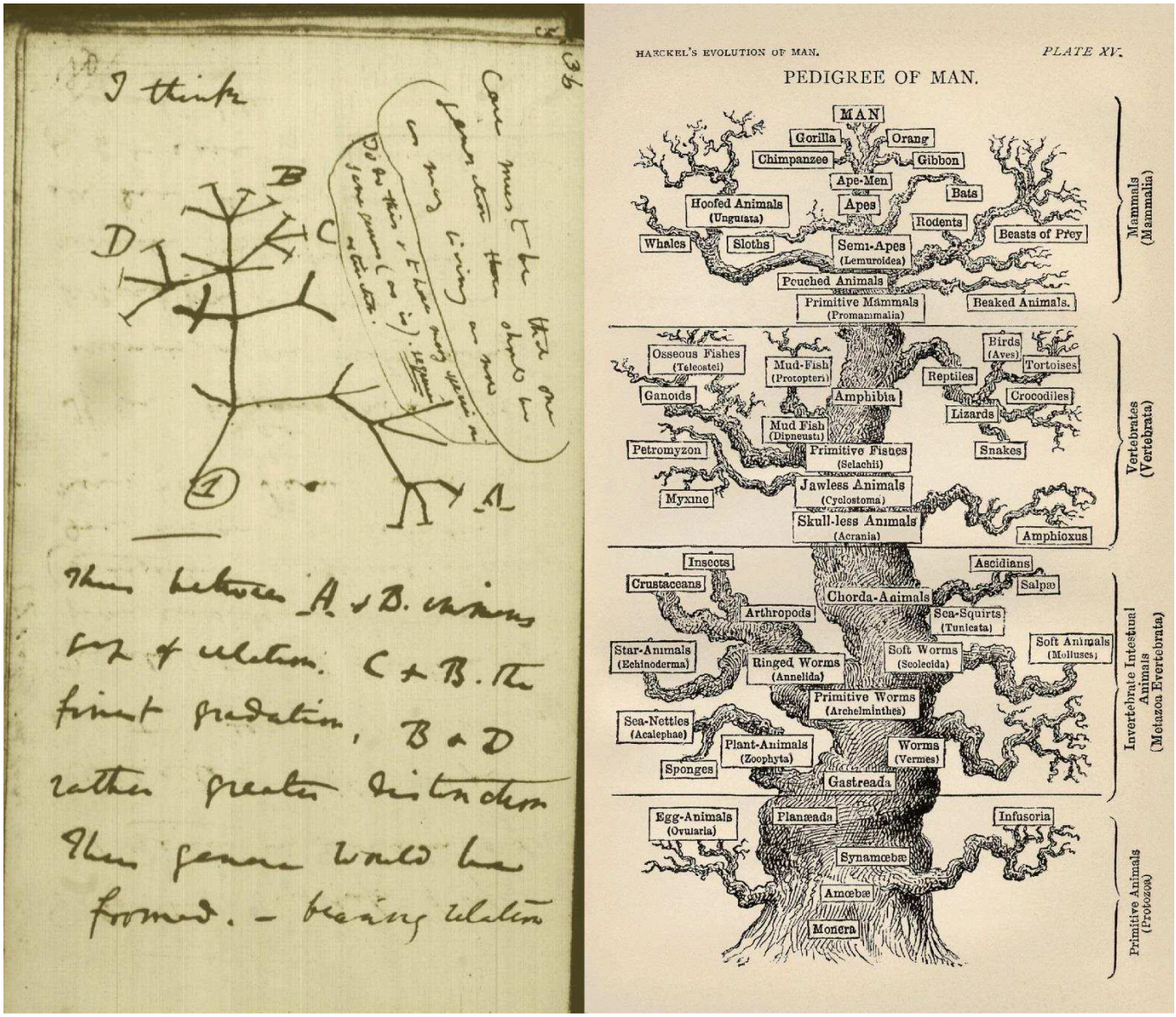


Fig. I.1 – Tree of life as seen in the 19th century. *The first-known sketch by Charles Darwin of an evolutionary tree describing the relationships among groups of organisms (left) and inspired by this idea the representation of the "Pedigree of Man" by Ernst Haeckel (right)*[6].

conditions, where most of the shapes and forms of organisms containing hard structures are recovered. Although, soft-bodied organisms are very rarely recovered, or the interpretation of the fossils obtained under these conditions should be taking very carefully[7,8]. Thus, data based only in the fossil record makes difficult to decipher the evolutionary origins of many animal groups, such as bilaterians, deuterostomes and chordates[9]. More recently evolutionary biology has widened to include molecular evolution and genetic architecture of adaptation, and it also considers very different forces that contribute to evolution, such as natural selection, sexual selection, biogeography and genetic drift. Nowadays, the concepts of evolutionary biology expanded to so many other disciplines (i.e. ecology, developmental biology, medicine, microbiology, genomics, anthropology, etc.[10]) that the well known phrase of Dobzhansky "nothing makes sense in biology except in the light of evolution" cannot be overlooked anymore.

Multicellular organisms arise by a rather slow and complex process of progressive change that is called development. In general, the development of a multicellular organism starts with a single cell, the fertilized egg, or zygote, which divides mitotically to produce all the cells of the body. But development does not stop at birth, or even at adulthood, since most organisms never stop developing. Each day multicellular organisms undergo a multitude of biological processes, such as replacing old cells by new ones (i.e. blood cells). Furthermore, some animals can regenerate body parts or even undergo metamorphosis. The study of embryonic and other developmental processes is called developmental biology. Developmental biology studies focus on two highly complex questions: how to create an entire adult organism from a single cell (Fig. I.2) and how does that adult body produce yet another body?[11]

Classically and due to the complexity inherent to these questions, they have been approached from different perspectives: differentiation, morphogenesis, growth and reproduction. Differentiation, once an initial fertilized egg gives rise to different germ layers, which by themselves will derive to different specialized cell types. Morphogenesis since it is, by definition, the biological process responsible for the rearrangement of these cells to form functional structures (organs) that later will be structured to create an entire functional organism. Growth is the paradigmal driving force of expansion of structures, which exhibit a very tight regulation, both spatial and temporal. And last but not least, reproduction that focus in the study of the origin of sperm, egg and their progenitor cells, since only they can transmit the instructions for making an organism from one generation to the next.

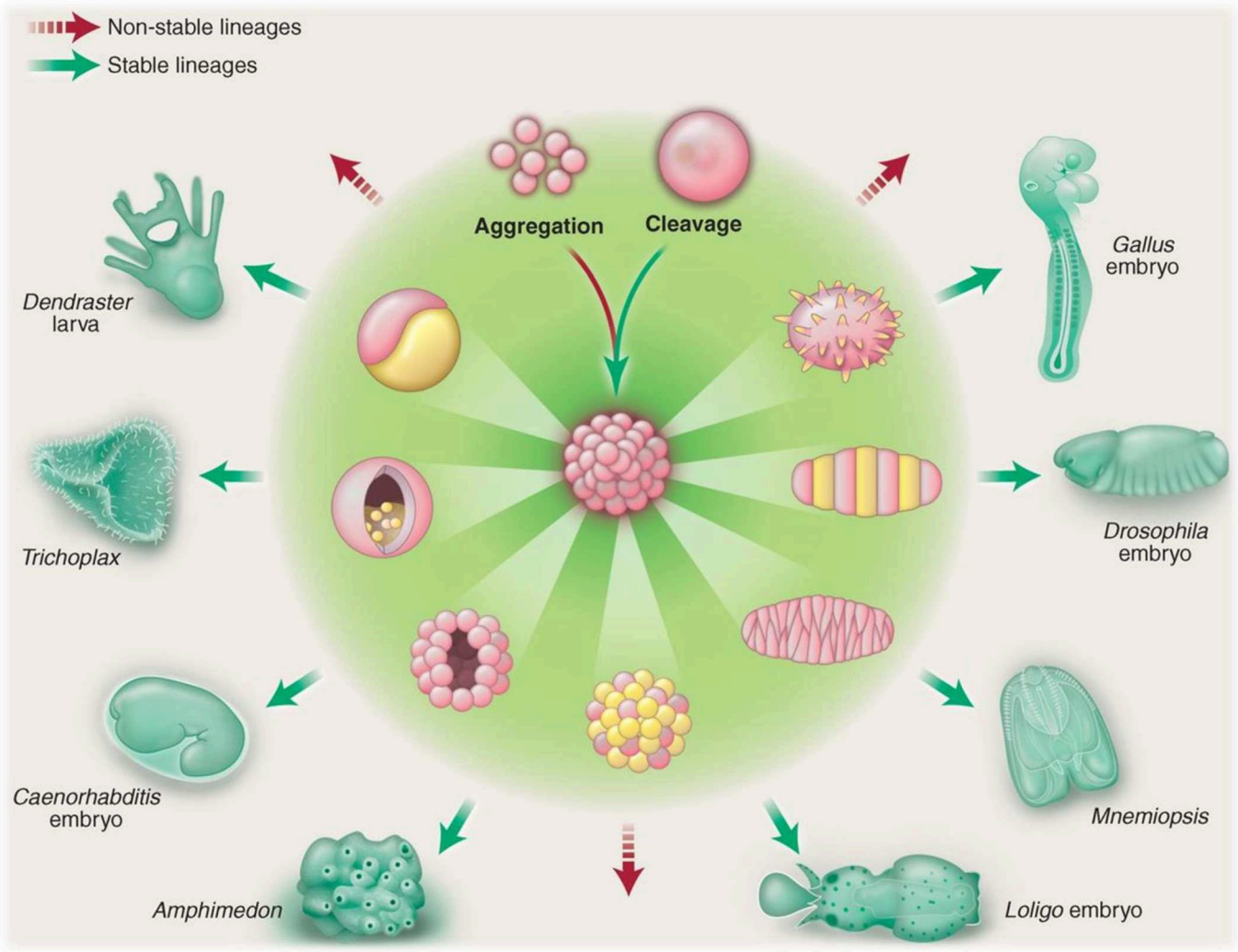


Fig. I.2 – Example of forms originated by development of multicellular animals. *Multicellular entities can be formed by the aggregation of unicellular organisms (red curved arrow) or the cleavage of enlarged cells (green curved arrow). The green inner circle shows some morphological motifs generated during development and at the exterior some late embryonic, larval and adult stages. Clockwise from top right: vertebrate (*Gallus*) embryo, arthropod (*Drosophila*) embryo, ctenophore (*Mnemiopsis*), cephalopod (*Loligo*) embryo, demosponge (*Amphimedon*), nematode (*Caenorhabditis*) embryo, placozoan (*Trichoplax*), and echinoderm (*Dendraster*) larva.*

Figure from Newman, 2012[12].

A step forward in the understanding of the remarkable process of embryogenesis was the adoption of the emerging tools of molecular biology. These new tools revealed the involvement of a relatively small number of different signaling pathways. The morphogens, receptors and transcription factors included in these cascades are considered the master regulators of development. Family members of the fibroblast growth factor (FGF), wingless/Wnt (Wnt), hedgehog (Hh), transforming growth factor beta (TGF-β) and retinoic acid (RA) signaling pathways are correlated differentially in different cellular backgrounds, yielding specific functions. In general, embryological processes are very tightly controlled, mostly through elaborate autoregulatory loops in combination with refined interactions between these master signaling cascades.

Recently, these two historical disciplines, the evolutionary biology and development, have coupled their questions and the many interacting points. This fusion built a bright new area named evolutionary developmental biology (evo-devo) that uses comparative embryology and developmental genetic as a system to study morphological evolution. The evo-devo field finds its roots at the end of the nineteenth century in the works of von Baer and Haeckel in comparative embryology[6,13,14]. Even though there are two different versions concerning the main events that originated the field of evo-devo. One view, argues that researchers such as Gould and Alberch were pioneers by coupling the study of paleontology with embryology and by arguing that changes in developmental timing can lead to major evolutionary changes[15,16]. An alternative view, claims that a major landmark in the origin of evo-devo is the discoveries of homeotic genes[17] and developmental genetic toolkit genes explaining how very different forms evolved from common Bauplans[18]. Both views recognize that in its earlier stages evo-devo had a very strong focus on the evolutionary biology and on the organism: Gould and Alberch were mainly interested in understanding the relationship between organismal evolution and organismal development, and the discoveries of homeotic genes started a movement on the characterization of the links between conserved genetic toolkit genes, Bauplans, and organismal diversity [19].

Evo-devo constitutes a research field that studies a large variety of questions but it should not be disconnected from other areas such as paleontology or zoology. Contradictorily, new theories emerging from evo-devo studies are often defied by paleontological data, and in order to propose robust evolutionary scenarios it is essential to combine elements from different and complementary fields[5,11]. In general and when possible, reconciliation of paleontological and developmental evidence supported by genetic data provides fundamental

advances in the understanding of morphological evolution. Nowadays, the main questions addressed by evo-devo studies can be grouped in the following categories[20]:

- developmental basis of homoplasy and homology;
- origin and evolution of embryonic development;
- production of novel features based on modifications of developmental processes;
- adaptive plasticity of development in life-history evolution;
- impact of ecology on development that modulates evolutionary change.

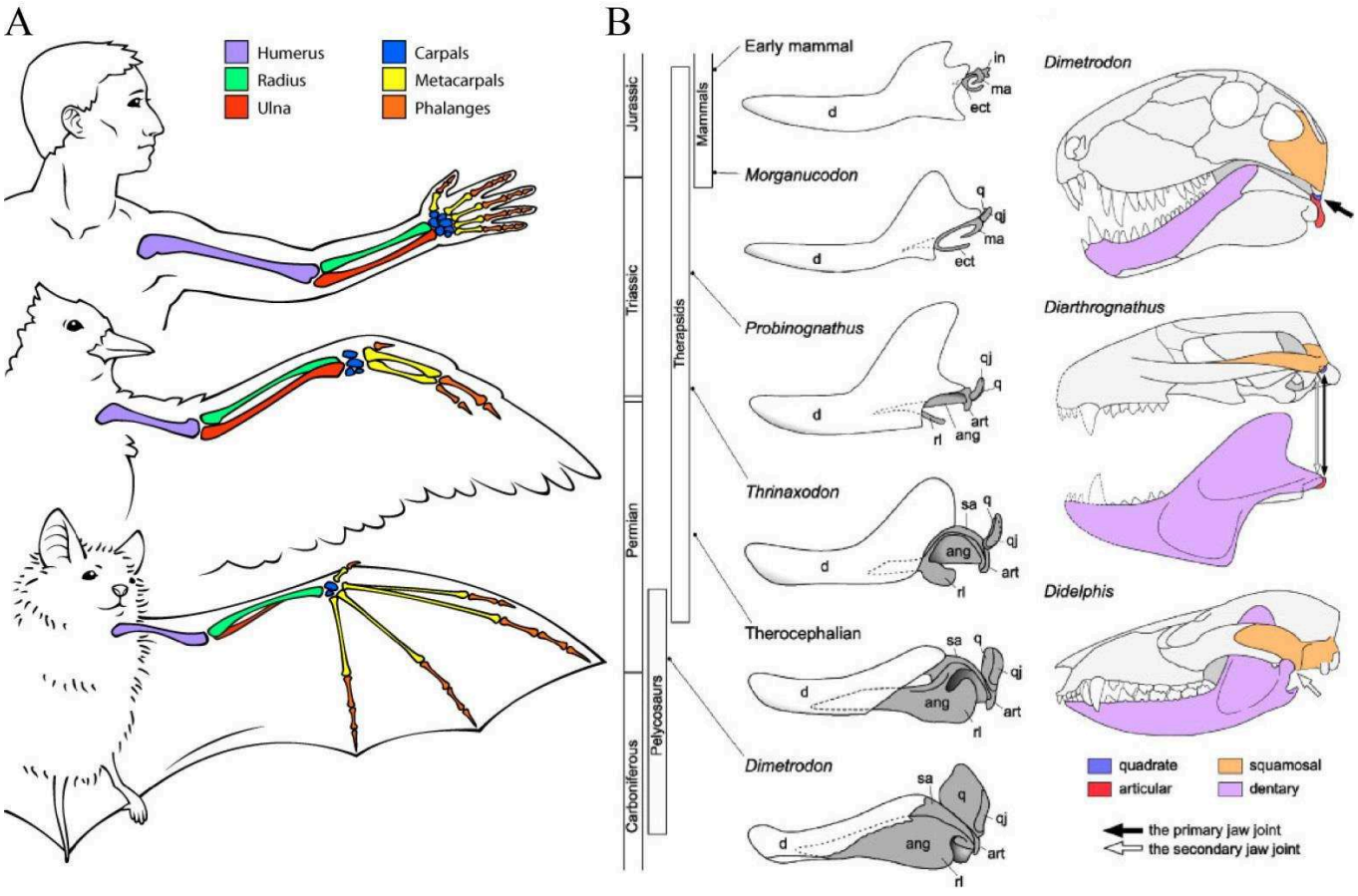


Fig. I.3 – Example of homologies. *(A) Vertebrate forelimbs bones homology in human, birds and bats. (B) Diagrams of lateral views of jaw skeletal elements showing modifications leading to the mammalian condition correlated with the occurrence of each animal in the geological record; and changes in jaw articulation during mammalian evolution in a pelycosaur,* Dimetrodon *(top), an ''advanced'' cynodont,* Diarthrognathus *(middle), and in an extant marsupial,* Didelphis *(bottom). Abbreviations: ang – angular bone; art – articular bone; d – dentary bone; ect – ectotympanic bone; ma – malleus; q – quadrate bone; qj – quadratojugal bone; rl – reflected lamina.*
(A) taken from https://askabiologist.asu.edu (B) adapted from Takechi and Kuratani, 2010[21]

The concept of homology in evo-devo

As referred above, the concept of homology is still a key principle in evolutionary biology and it also has a primordial importance in the evo-devo field. In evolutionary biology, historically, homology defines the resemblance due to inheritance from a common ancestry, meaning that homologous features in two or more organisms are those that can be traced back to the same feature in the common ancestor of these organisms[22]. For instance, bird wings and bat wings are homologous as vertebrate forelimbs, but not as wings (Fig. I.3 A). Differently, the so called biological homology concept accounts for the mechanistic rather than the genealogical aspect and it includes the idea of continuity thus, structures descended from a common ancestor are homologous even if they have diverged and have no clear morphological similarity[23]. In the light of this concept, for example, the reptilian jaw bones are homologues to the auditory ossicles of mammals (Fig. I.3 B)[21]. In evo-devo research, biological organisms are hierarchical structures, and this brings new insights into the concept of homologies. Different levels, be it body parts, organs, tissues, germ layers, cellular lineages, cells, signaling pathways, proteins, gene sequences, etc. may be used to hypothesize on homology and then analyzed within the phylogenetic framework[24]. Furthermore, specific terminology is employed in certain levels used to study homology, for instance the concept of orthology (relationship caused by the divergence of organismal lineages) and paralogy (homology arising from gene duplication) are often used when classifying the relationship between homologous genes and proteins.

The term homology is often confused with others, such as homoplasy, similarity and identity: Homoplasy is defined as the similarity between two characters that did not arise from an ancestral character. There are three types of homoplasies: convergence, parallelism and reversion[5]. Differently, similarity or identity are terms used when comparing molecular sequence: while it is possible to characterize a percentage of identity (or of similarity) between two protein or gene sequences, these same proteins or genes either are or not homologous. The capacity of defining homologies is sometimes very challenging, but it constitutes one of the main pillars of evo-devo.

Developmental basis of morphological novelties

Modifications of the embryonic development can trigger changes in the body plan of animal species and therefore have an instrumental role in the emergence of morphological

novelties. One major question in this topic concerns the evolution of differences in gene expression. Such differences can arise from *cis*-regulatory or *trans*-regulatory changes. *cis*-regulatory mechanisms affect the initiation of transcription, the rate of transcription and even transcript stability in an allele-specific manner, differently *trans*-regulatory changes can modify the activity or expression of factors that function in a *cis*-regulatory manner. Two different extreme theoretical models can be considered to explain the numerous differences in gene expression between species: a few *cis*-regulatory changes affecting regulatory genes could have widespread *trans*-acting effects on the expression of many downstream targets (*trans* modifications only), or each structural gene expression could be altered by its own *cis*-regulatory genetic changes (*cis* modifications only). Alternatively, and in agreement with is observed in empirical studies, both *cis*- and *trans*-regulatory changes contribute to divergent gene expression, although their respective contributions remain to be clarified[25]. Furthermore, it has been shown that not only these two mechanism function together to change gene expression levels, but also that they work in a compensatory way to avoid a large drift induced by structural modifications both in *cis* or in *trans*[26–28].

It is now well accepted that repeated amplifications in gene repertoires during evolution have been instrumental in the diversification of animal body plans. In the early 1970s, Susumu Ohno put forward the idea that two rounds whole genome duplication (WGD) events would be the most plausible explanation for the evolution of novelty and complexity in a short time during early vertebrate history (2R hypothesis)[29]. This hypothesis has been strengthened by data that shows that vertebrates often have multiple (generally as many as four) duplicated genes corresponding to a single set of invertebrate orthologs[30,31]. Furthermore, a great amount of studies in genomics and molecular phylogeny have pointed out the key role of genes and genome duplication in molecular and morphological evolution[32]. Even though, recent sequence analyses have not categorically clarified the timing of these genomic events during vertebrate evolution[33,34]. Within this framework, a current model called DDC (for duplication-degeneration-complementation)[35,36] proposes that recently duplicated paralogs can have different fates:(i) nonfunctionalization: after duplication, one copy is lost and one retains the function of the ancestral gene; (ii) subfunctionalization: following duplication, the two copies share the functions of the ancestor gene; and (iii) neofunctionalization: after duplication, one copy retains the functions of the ancestral gene and the other one acquires a new function[35]. Thus, this model explains how new functions can be acquired following a gene duplication event, both at the whole genome level as well as at

the level of single gene families. So, at the scale of a genome, it is easily understandable how such massive duplication processes can fundamentally facilitate the diversification of body plans in a given species or group of animals.

Alternatively to the 2R hypothesis, there are also other visons that explain the diversification of genomes and the existence of a large repertoire of duplicated gene families in vertebrates, when compared to their orthologs in invertebrate species. One of these alternatives views argues that a single, ancient WGD might have occurred in the vertebrate lineage, before the split between cyclostome (jawless vertebrates, i.e. lampreys and hagfishes) and gnathostome (jawed vertebrates). Along with it, independent segmental duplications might have occurred either pre- or post-dating the WGD event[37]. Another alternative view to the 2R hypothesis claims that the vertebrate genome evolved by relatively small-scale (segmental), regional duplication events at widely different time points in animal history[38–40]. This hypothesis is also in accordance with data confirming recent segmental duplications in primates[41] and mice[42], suggesting that the mechanisms underlying recent and ancient vertebrate genome evolution are similar[40].

One of the classical examples of mutations in gene complements in embryonic development that alter body organization and, thus, lead to morphological and anatomical changes is the case of homeotic (*Hox*) genes. In bilaterians, Hox genes are responsible for the establishment of the anteroposterior (A-P) axis. A specificity of these genes is their organization in clusters on the genome of many species[43–45]. Additionally, *Hox* genes are characterized by a collinearity of expression, both spatial and temporal[46]. *Hox* genes located at the 3' end of the cluster are usually expressed in anterior regions of the embryo and earlier during development, while genes in 5' are expressed in more posterior regions and at later stages of development. Combinatory expression of *Hox* genes in time and space confers a positional identity along the A-P axis and thus establish a specific identity to different body segments[47]. Furthermore, *Hox* genes are also involved in the patterning of organs and in establishing territories of induced cell death and cell proliferation[48]. A dramatic example of the mutation of *Hox* genes is seen in *Drosophila* with the mutation of, for instance, the gene *antennapedia* (i.e. the orthologue of *Hox6* class of genes), while loss-of-function results in the development of the second leg pair into ectopic antennae, the gain-of-function convert antennae into ectopic legs[49,50].

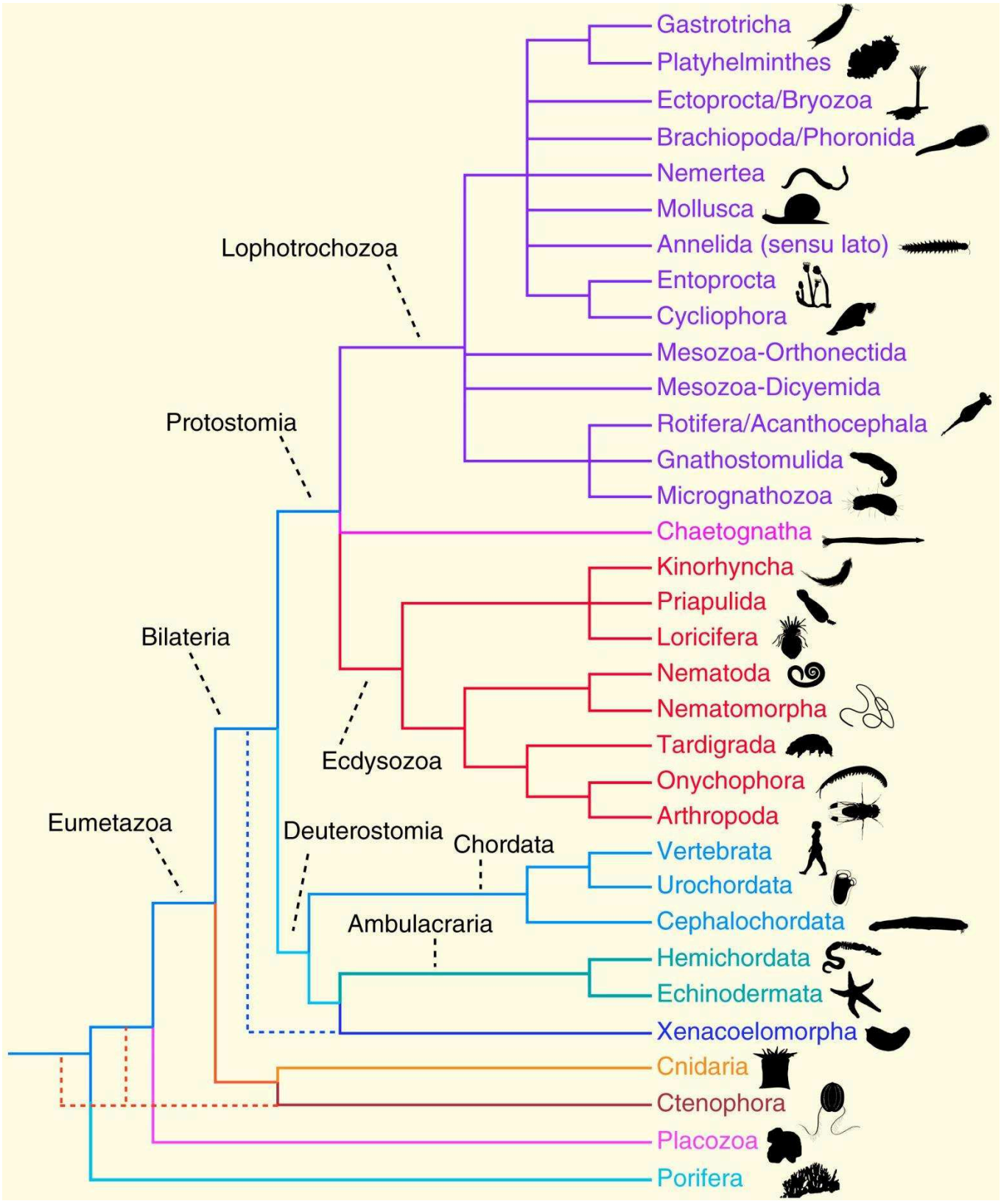


Fig. I.4 – Current view of the Metazoan tree of life. *Some of the major clade names are specified. Alternative possible positions for groups with the most contentious positions are indicated by colorful dashed lines. Figure adapted from Telford et al. 2015*[51].

Metazoan phylogeny

Since the publication of Haeckel's work containing his preposition of the animal tree of life[6,14] (Fig. I.1), scientists have made remarkable progress in understanding Metazoan (animal) evolution relationships. In the last decade, the phylogenetic analyses focus in genomes and transcriptomes have played a fundamental role to the current understanding of the tree of life. The current consensus indicates that the animal kingdom is monophyletic and composed of a little more of 30 different phyla (Fig. I.4)[51,52].

The first fundamental and parsimoniously accepted division in the tree of life is the boundary between non-bilaterians and bilaterians. Non-bilaterian animals comprise organisms from the phyla Porifera (sponges), Placozoa (small group of creeping animals), Ctenophora (comb jellies) and Cnidaria (corals, anemones, jellyfishes and their relatives)[53], while bilaterians are a major group of animals including the all the remaining phyla. These non-bilaterian phyla are also referred to as "early-branching" or "basal" metazoans and their relative position in the base of the animal phylogeny represents one of the most difficult and controversial challenges in the field[54]. These difficulties might be associated with the lack of molecular sampling and/or with an inadequate sampling of the taxa[55]. Traditionally, Porifera has been thought to be the sister group of all other animals[56], although recent studies raised the hypothesis that this position is occupied by the phylum Ctenophora[57,58].

Bilateria

Bilateria is common divided into two major groups: Protostomia and Deuterostomia. The names of these groups are due to an embryological feature that is the direction in which the mouth develops in relation to the blastopore. Protostomia (from the greek πρώτος + στόμα, "mouth first") mouth generally derives from the blastopore during embryonic development, while in deuterostomes the blastopore gives rise to the anus, and the mouth is thus secondarily formed (δεύτερος + στόμα ; "second mouth")[59]. Although, these terms are still used to differentiate these two groups, not all the protostomes phyla actually respect this process of development. For instance, the phyla Chaetognatha, Priapulida and potentially Onychophora develop under a deuterostomy model[59–61], while cases of amphistomy (blastopore as a furrow that gives rise to both mouth and anus) can be identified in some species of the phyla Nematoda and Annelida[59]. Nevertheless, in Protostomia the main neurite bundles localized ventrally are considered a synapomorphy (shared derive character or trait state that

distinguishes a clade from other organisms), whereas synapomorphies in Deuterostomia include the origin of coelom by enterocoely, pharynx with ciliated gill slits and the existence of three distinct coelomic cavities (protocoel, mesocoel and metacoel), also know as archimery[52]. Despite the position of these two groups (protostomes and deuterostomes) being fully accepted, it exist a controversy that discusses whether the phylum Xenacoelomorpha (comprised by *Xenoturbella* and two related groups of acoelomorph flatworms, Acoela and Nemertodermatida[62]) branches at the base of the Bilateria[63–65], or differently, if this phylum branches inside deuterostomes as sister group of the Ambulacraria clade(Echinordermata + Hemichordata)(Fig. I.4)[66,67].

Protostomia

Presently, Protostomia consist of three different clades: the phylum Chaetognatha as well as the clades Ecdysozoa and Lophotrochozoa, also referred as Spiralia (Fig. I.4). Chaetognats, or "arrow worms", are a small group of marine predators that have a very specific set of peculiarities[68]. In general, both the genome and the morphological characters of chaetognats denote many autopomorphies (distinctive features or derived traits unique to a given taxon) and at the same time are characterized by a mix of features from protostomes and deuterostomes[69]. This suggests that Chaetognata is an early branching phylum in the protostome linage, making it a sister group of both Ecdysozoa and Lophotrochozoa clades[69]. Ecdysozoa are characterized by animals that have a molted cuticle and a trilaminated epicuticle[52]. This clade encompasses 8 different phyla (Fig. I.4) and in terms of species numbers and niche diversity, the Ecdysozoa is from far the most significant clade of animals since there are over million described species and an estimated total of more than 5 million living ones[70]. Probably due to the high diversity found in ecdysozoans the evolutionary relationship between the composing phyla is still a matter of debate, although the division of these eight phyla in three different clades is rather supported fact[51,52]. Lophotrochozoa or Spiralia is a phylum rich clade and it represents probably one of the most poorly resolved branches of the animal tree of life (Fig. I.4). The main synapomorphy of lophotrocozoans is the cleavage mode in spiral, but a great part of them can be grouped accordingly to the type of larvae, throcophore-like[51,52,71]. Possibly, the difficult resolution of the relationship between the different clades composing Lophotrochozoa is a consequence of a likely rapid divergence of phyla from one another by dissemination of new body plans[51].

Deuterostomia

The composition of Deuterostomia clade has changed through time, with some of the lophophorate groups (especially Brachiopoda and Phoronida) being considered deuterostomes, although molecular studies now place them with the protostome clade[52]. Likewise, deuterostome affinities of chaetognaths have lost importance after the emergence of phylogenomic analyses that clearly place "arrow worms " as part of Protostomia[68,69]. It is now widely accepted that deuterostome species are distributed into two main groups: (Xen)Ambulacraria which encompass the phyla Hemichordata and Echinodermata (and eventually Xenacoelomorpha) and Chordata that consists of three phyla: Cephalochordata, Urochordata and Vertebrata (Fig. I.4). As mentioned above, pharyngeal gill slits have emerged as ancestral deuterostome characters but theses structures were subsequently lost in extant echinoderms and aminiotes[72].

The group of the ambulacrarians was named after the ambulacrarian system (or "tube feet"), which is involved in nutrition, locomotion and respiration, in for example sea urchins and sea stars. Within this clade, Hemichordata (from the Greek ήμισυ + χορδή, "half cord") represents a phylum of exclusively marine organisms and are divided into two major groups: the paraphyletic solitary enteropneust worms (the acorn worms) and the colonial and tube-dwelling pterobranchs[73,74]. The global morphological affinities (i.e. the presence of gill slits) and close phylogenetic relationship to the chordates make this phylum a good model to gain in insights into the early origins of the chordate body plan[75,76].

Echinodermata animals are often morphological characterized by pentameric (five fold) symmetric adult body plan and by the presence of a water vascular system that consist of a structured fluid-filled channels and reservoirs used for locomotion and internal transport[77,78]. Among Echinodermata ("spiny skin" in greek) morphologically diverse animals can be found, like sea urchins, sea stars, sea cucumbers and crinoids, which present non-conventional characteristics within Bilateria. Due to these non-conventional body plans Echinodermata species have largely been excluded from broad body plan comparisons because of the difficulties of establishing a rigorous basis for comparisons[77]. Even though, and since their development is relatively simple and rapid, gene function can be relatively easily perturb and manipulate, echinoderms, mainly some sea urchins and sea stars species, have being widely used to study early development mechanisms[77,78]. The phylogenetic relationship within this phylum is now rather consensual, based in both morphological and molecular data there are

two main clades: Pelmatozoa (Crinoidea) and Eleutherozoa (all the remaining extant classes)[78,79].

Finally, as mentioned above a last group of animals has recently been proposed to be part of ambulacrarians, based on molecular analyses: the phylum Xenacoelomorpha. This phylum was first associated to platyhelminths, and then to mollusks due to a contamination with their main source of food (bivalves)[80]. Ever since this phylum has been associated with ambulacrarians and thus the name of this clade (Xenacoelomorpha + Ambulacraria) was proposed to be changed to Xenambulacraria[81]. Alternatively, recent publications have proposed that Xenacoelomorpha is the sister group of all animals in the clades Deuterostomia and Protostomia[63–65].

Chordata

As the sister group of Ambulacraria, the chordates represent the second main deuterostome clade (Fig. I.4). Chordata ("chord" in greek) owe its name to the presence of an embryonic structure called notochord which originates from the mesoderm and plays instrumental roles in the establishment of the A-P axis of the embryo and the differentiation of several chordate-specific structures[82]. During development the notochord is resorbed and is completely absent in most adult chordates, being the noticeable exception the cephalochordates[83]. The origin of this Chordata synapomorphy is a matter of debate and the running hypotheses consider either the notochord as a chordate novelty or a legacy from an ancient annelid (or even an evolutionary derivative from a still unknown mechanism)[83]. Beside the notochord, chordates exhibit (during developmental or in its adult form) other synapomorphies, such as a dorsal hollow neural tube, pharyngeal slits, a postanal muscular tail and an endostyle that is considered the homolog of the thyroid gland in vertebrates. The endostyle homology found is a particular interesting example of chordate synapomorphy as it has been proposed that thyroid-controlled metamorphoses may be homologous in chordates[84]

The clade Chordata comprises three phyla with various degrees of species diversity: Vertebrata, Urochordata (also known as Tunicata) and Cephalochordata (Fig. I.4)[85]. The uniting relationship between vertebrates, tunicates and cephalochordates, were first suggested by Haeckel based on shared developmental features[86]. Within Chordata, phylogenetic relationships have been updated and despite the striking similarities observable in terms of the body plans of cephalochordates and vertebrates, it is the phylum Urochordata that is in fact the sister group of vertebrates[87]. Indeed, while both phyla Cephalochordata and Vertebrata

have segmental muscles, a dorsal nerve cord and pronepheric kidney[88], these elements are absent in the derived body plan of urochordate adults. The monophyletic group comprising Vertebrata and Urochordata is often referred to as Olfactores, a name that alludes to the presence of ectodermal placodes in these animals.

Vertebrata (or Craniata) are the largest phylum in the clade Chordata. The name Vertebrata derives from the bones of the spinal column, also referred to as vertebrae, although this definition is not completely appropriate because hagfishes are popularly believed not exhibit any strictly vertebral structure[89,90]. Rather, vertebrates are most easily distinguished from all other chordates by having an unequivocal head with paired sensory organs. Also, the presence of neural crest, segmented brain, endoskeleton and other special sensory organs are considered general synapomorphies of the phylum Vertebrata[85]. The diversification of this phylum largely starts during the Cambrian explosion and vertebrate extant species include many variable forms, such as hagfishes, lampreys and all kinds of other fishes as well as amphibians, reptiles, birds, and mammals (including humans)[85]. Vertebrates can be divided into two main groups: jawless vertebrates (Cyclostomata), a monophyletic group composed by hagfishes and lampreys, and jawed vertebrates (Gnatosthomata). Although cyclostomes only account for a small fraction of all vertebrates, they still represent fundamental organisms to study the emergence of the vertebrate body plan, as they exhibit typical craniate features but also lack several synapomorphies of gnathostomes[91]. Within the Gnathostomata tree, phylogenetic relationships are now consensually established and one major dichotomy in clade is the division between cartilaginous fishes (Chondrichthyes) i.e. sharks, rays and chimera, and the group of Osteichthyes (bony fishes) that comprise all other jawed vertebrates. Osteichthyes are then divided into ray-finned fishes (Actinopterygii) and lobe-finned fishes (Sarcopterygii) i.e. coelacanths and tetrapods [91].

Urochordata or Tunicata is the second largest phylum of the chordates and the etymology (tail + chordates) is somewhat ambiguous since some species have lost their tailed tadpole-like larvae and thus the term Tunicata is more adequate. A distinctive feature for this animal phylum lies in their capacity of directly producing cellulose, a biological function mostly associated with bacteria and plants. In general, the entire adult body of Tunicata animals is enveloped in a thick covering, made of a specific type of cellulose (tunicin), so this structure is called the tunic[92]. In these marine animals specialized in filter-feeding, commonly the seawater enters in the pharynx through an inhalation or oral siphon, food particles are then

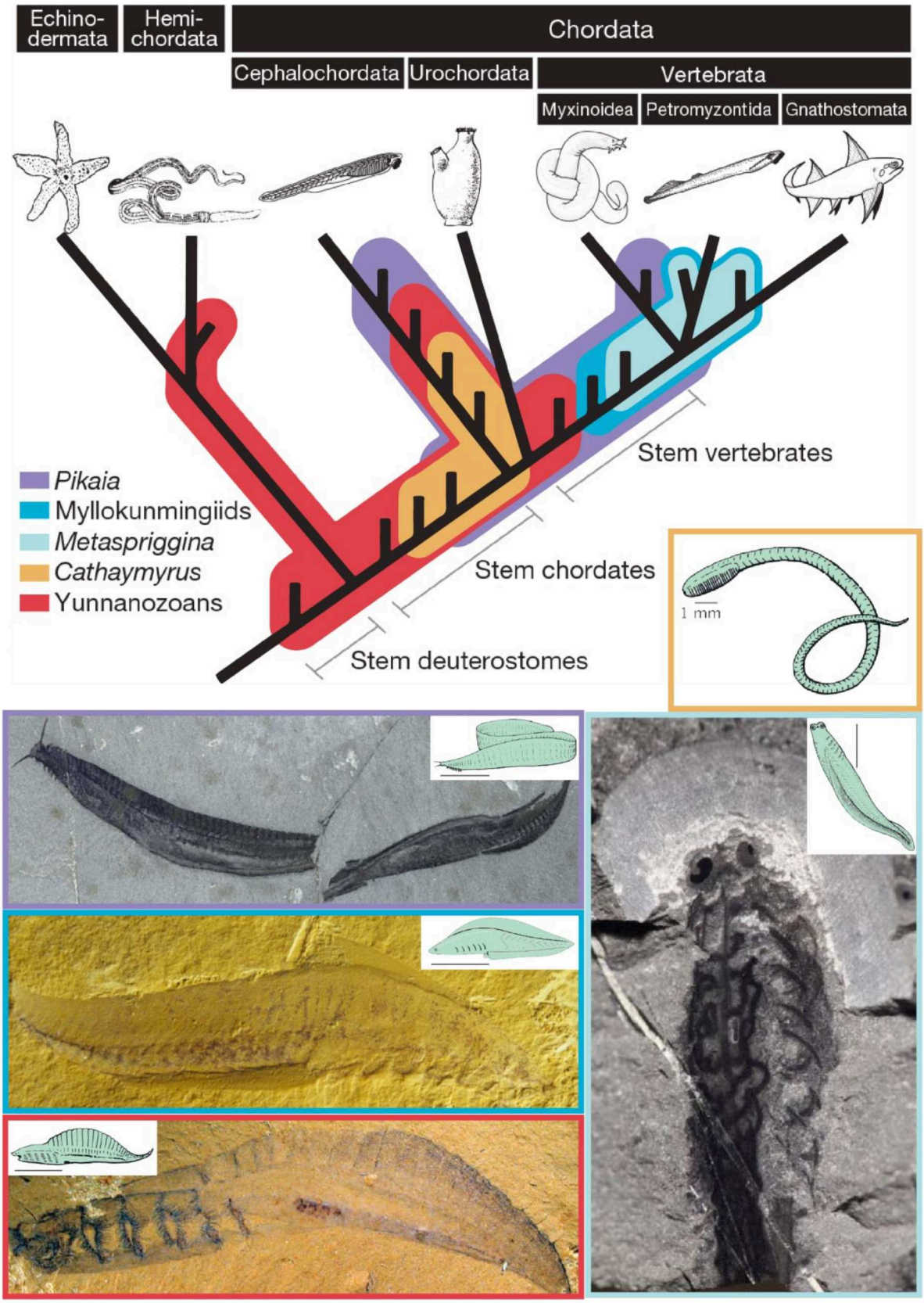


Fig. I.5 – Phylogeny and potential placement of extinct deuterostomes. *Tree and relative placement of different extinct deuterostomes adapted from Sansom et al. 2010[7], drawing representations of the fossils from Janvier 2015[9] with scale 10mm unless stated otherwise, and fossils pictures were obtained from "The Virtual Fossil Museum".*

trapped on a mucous net secreted by the endostyle, and the water and waste is expelled through an exhalation or atrial siphon[93]. Tunicata species have a great variety of life styles, varying from benthic to pelagic forms that can be colonial or solitary, and thus their evolutionary relationship remains controversy. New recent molecular analysis supports the division of tunicates into three major clades: 1) Phlebobranchia + Thaliacea + Aplousobranchia, 2) Appendicularia, and 3) Stolidobranchia, although the position of Appendicularia its still not resolved within this phylum[94]. The particular genome organization and molecular evolutionary patterns represent another particularity of the phylum Tunicata. For instance, *Oikopleura dioica* and *Ciona intestinalis* exhibit great genome plasticity, noticeable by extreme genome reduction and compaction as well as by the high rates of molecular evolution. When compared to vertebrates genomes, the evolutionary rate is three times faster for *Oikopleura* and twice as fast for *Ciona*)[95].

The phylum Cephalochordata comprises only a few dozens of species that have striking resemblances between them[96,97]. These animals have fish-like morphology and spend their lives borrow in sand where they filter food particles from the water column. Nevertheless, when needed cephalochordates can produce a very rapid fish-like locomotion[85]. Despite the lack of many characteristics of vertebrates, like for instance eyes or heart, cephalochordates (also called lancelets or amphioxus) have typical features, e.g. the dorsal hollow nerve cord and notochord, the perforated pharynx with pharyngeal gill slits[98,99]. Globally, cephalochordates can be divided in three different genera: the most basal one *Asymmetron* and the derived ones *Epigonichthys* and *Branchisotoma*[96]. At the phenotypic and genomic level cephalochordates are characterized, in contrary to tunicates, by a slow evolutionary rate, making Cephalochordata an interesting proxy for the chordate ancestor[100].

Fossil record for chordates is generally defined by their hard parts, which resumes essentially to neural crest derivatives such as odontogenic tissues of the dermal and branchial skeleton, and therefore fossils of organism completely lacking hard parts (i.e. cephalochordates or tunicates) are rare[101]. There are a few important types of fossil (Fig. I.5) that can potentially account for chordate and for vertebrate evolution: the Burgess Shale in Canada fossils *Pikaia*, *Banffia*, *Metaspriggina* and the fossils form Chengjiang in China, Yunnanozoans (i.e. *Yunnanozoon* and *Haikouella*), Vetulicolans (i.e. *Vetulicola*, *Xidazoon*, *Didazoon* and *Pomatrum*), *Cathaymyrus*, *Cheungkongella*, *Shankouclava* and Myllokunmingiids (*Haikouichthys* and *Myllokunmingia*)[9,101–103]. A lot of controversy and discussion has been placed in the information from afore mentioned fossils but, for instance *Pikaia* and Yunnanozoans are now considered as potential stem deuterostomes or stem

vertebrates. *Banffia* and Vetulicolans are possible stem deuterostomes, stem chordates or stem protostomes. Differently, *Cathaymyrus* is a likely stem cephalochordate and *Cheungkongella*, *Shankouclava* are considered as tunicates. *Metaspriggina* and Myllokunmingiids in other hand are potential stem vertebrates[9]. Taken together, the paleontological data from Cambrian fossil deposits tentatively opens a window into the early evolution of chordates, although the problematic interpretation of some structures and the unclear phylogenetic position of these soft-bodied fossils do not allow the inference of supported evolutionary scenarios.

Chapter II: Amphioxus as an evo-devo model

Model and "non-model" animals in evo-devo research

Metazoan species have a huge diversity of shapes, habitats and modes of life, with 8-10 millions being the estimated number of existing species belonging to the animal kingdom, and it is assumed that only 11-14% of that diversity has been already characterized[104]. Understanding how is generated the diversity of development strategies, of germ layers and their respective derivatives, of the overall body symmetry as well as of organs, tissues and cell types constitute a major challenge of the modern evo-devo and it can only be achieved by understanding development in the context of a broad Metazoa sampling. Despite this needy idea evo-devo started by focusing on the analysis of model organisms already established in the field of developmental biology. Even though, in the last two decades this initial state has been greatly expanded, a lot is still to do in order to have an assertive and complete vision of how morphology traits are affected by different evolutionary constraints (Fig. II.1). The investigation in evo-devo of several concepts such as modularity, redundancy, developmental constraints, evolutionary novelties and phenotypic plasticity, have created the need to establish suitable models that can "fill the blanks"[105,106]. However, sometimes evo-devo suffers from a misguided selection of "non-model" organisms, often with a limited availability of technical tools[107]. Furthermore, characterization of early branching metazoan, such as Porifera, Placozoa, Ctenophora and Cnidaria can be especially useful to understand the origin of animal development. Nonetheless, a broad sample of other metazoan taxa is required to verify the homology as well as to understand until which extent the known developmental mechanism and genetic toolkits are used and/or modified.

Thus, it came as no surprise that a given set of "non-model" organisms (e.g. hemichordates, cephalochordates, tunicates and some unusual vertebrates) are more adequate to study a given question (for instance, the origin of A-P patterning CNS organizers in vertebrates, see annexes) than a different group of animals[108]. This notion is therefore one of the key principles to select and to develop an evo-devo "non- model" animal, into an established and exploitable animal model. Furthermore, with the advent of genomics[109] and

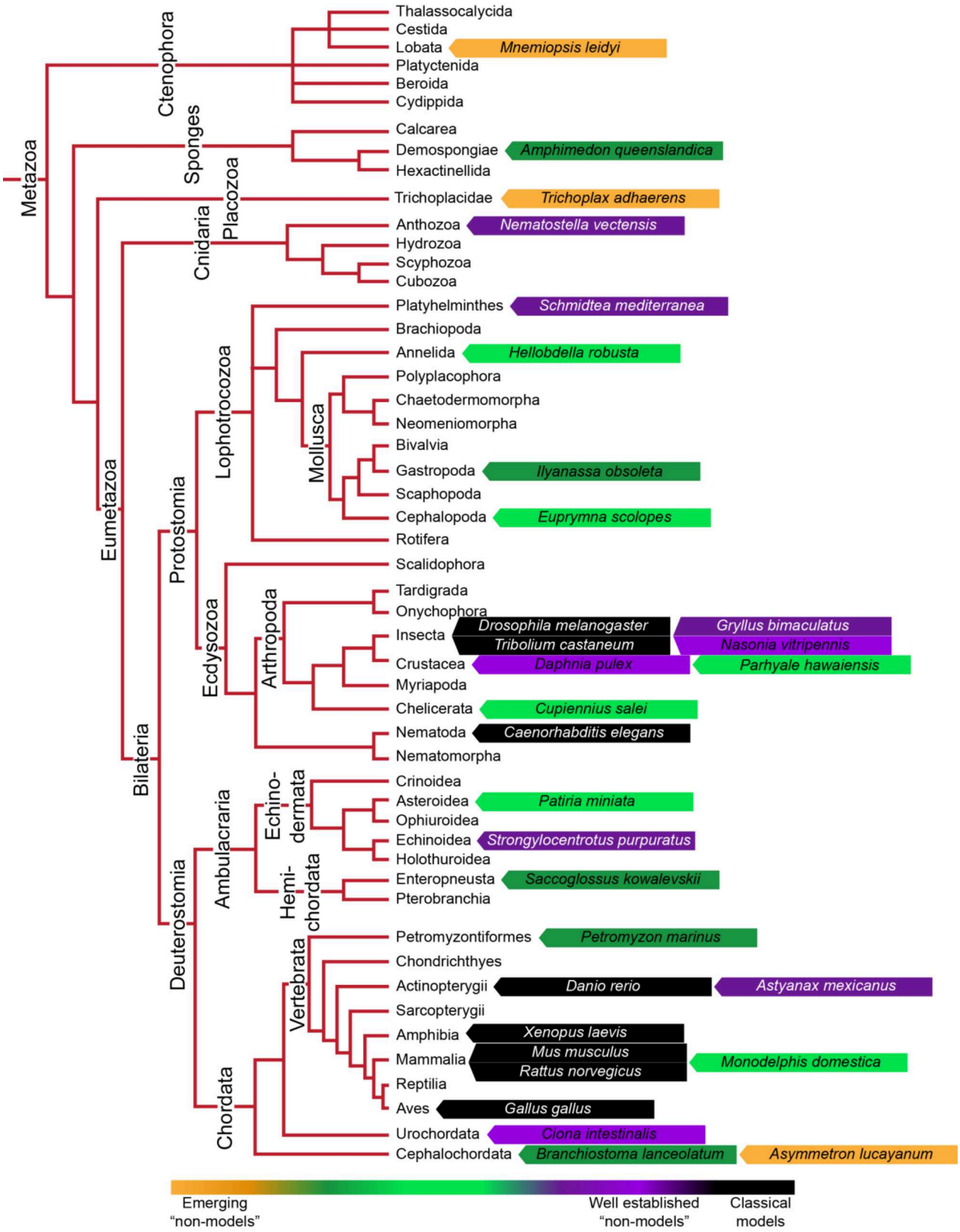


Fig. II.1 – Diversity of "non-model" and model animals. *Colored flags on the right indicate species within the indicated clade that are being used for developmental or evo-devo work. Colors established accordingly to the original image*[106].

transcriptomics[110] along with the huge potential of site-specific gene editing techniques, such as CRISPR-Cas9[111,112], it is now virtually possible to start studying any kind of non-conventional model. Although, especially in our days, another driving factor of the choice of model species is related to the potential insights that one can obtain for the improvement of human health. This biomedical-driven evo-devo thus has a very strong potential and has greatly helped to better understand, for instance, the biology underlying birth defects, cancers and even the role of the environment in development[113].

Amphioxus, a model to study vertebrate evolution

Since the 19th century, when clear affinities between amphioxus (cephalochordates) and vertebrates were first identified, amphioxus was classified as part of Vertebrata. This notion was first celebrated in 1852 by Saint-Hilaire[114,115] and for over 50 year the amphioxus were considered as vertebrates (Fig. II.2)[116]. For example, during their life cycle both amphioxus and vertebrates possess Chordata synapomorphies: a dorsal hollow nerve cord, a notochord, a postanal tail, pharyngeal gill slits and an endostyle. Moreover, contrarily to tunicates, amphioxus and vertebrates exhibit myomeres (muscle segments) and a metanephric kidney. Although, amphioxus also lacks some major vertebrate innovations such as bony and cartilaginous structures (skull, vertebrae, pharyngeal arches), paired body appendages, paired sensory organs and a morphologically differentiated brain. In addition, amphioxus exhibits some phylum specific features like segmented gonads and the persistence of the notochord in the adult, which is normally lost during development in vertebrates[98,114]. Most of the morphological differences observed between amphioxus and vertebrates can partially be explained by divergences during embryonic development. For instance the absence of definitive neural crest cells in amphioxus development can account for the absence of any bony and cartilaginous structures[117]. Also, the absence of development of ectodermal placodes can explain the lack of some sensory structures presented in most vertebrates, like paired sensory organs[118]. Furthermore, and differently to what is observed in vertebrates, during early embryonic development amphioxus exhibits a gastrulation mode by invagination, much alike sea-urchin gastrulation[119,120]. The differences in the mechanism of gastrulation were suggested to be related to the presence of yolk in the embryo, and thus probably a vertebrate innovation that arose in connection with changes in developmental processes, i.e. gastrulation by extensive involution[120]. Accordingly, tunicates lack key developmental genes

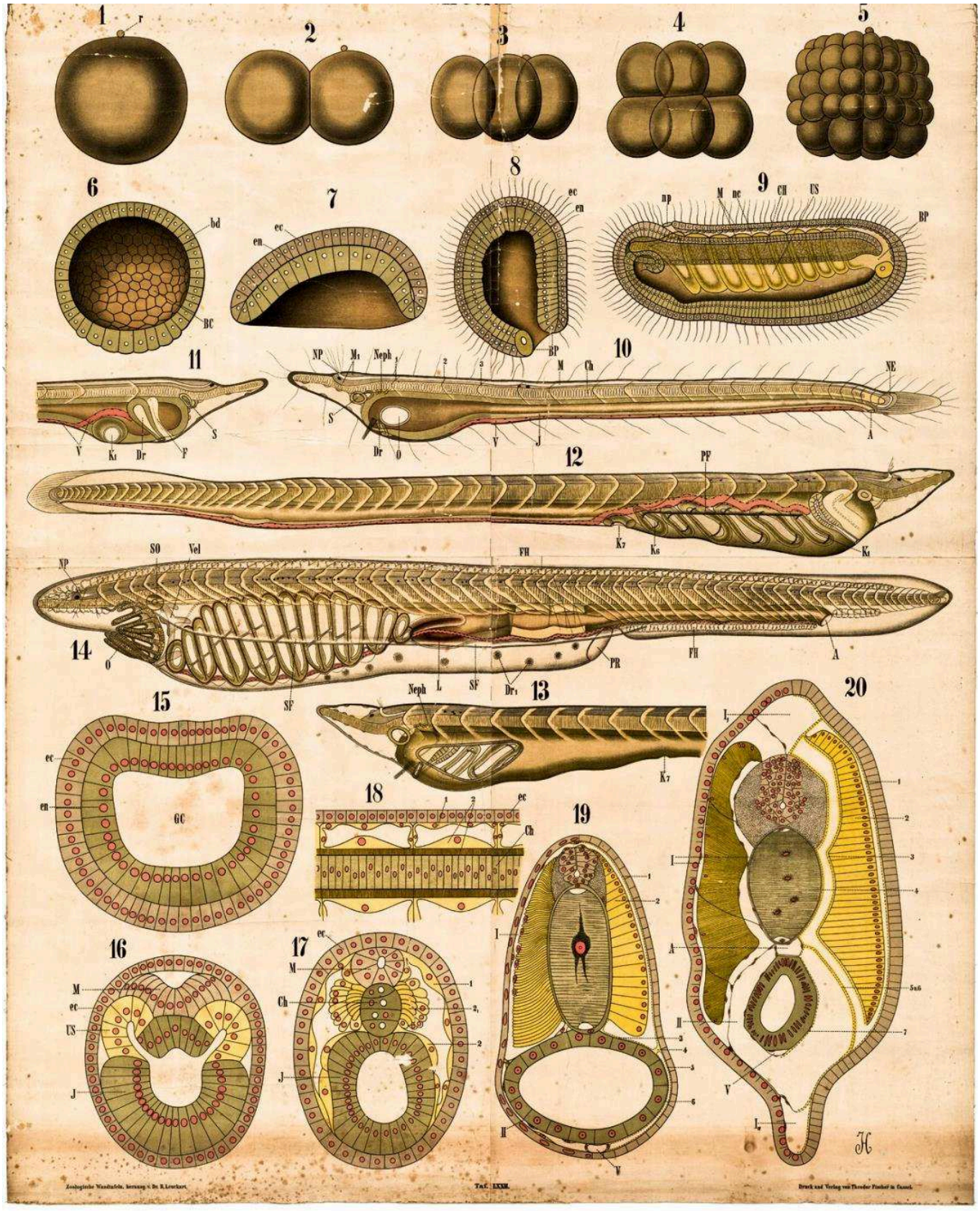


Fig. II.2 – Leuckart's chart of "Amphioxus lanceolatus".[121]

and switched to a determinant development (cell fate is fixed early in development) also, the constraints on their evolution differ greatly from cephalochordates and vertebrates, which might explain their highly divergent and diverse body plans[93]. In summary, despite some important differences between amphioxus and vertebrate early development, they nonetheless share a great number of very specific features, with cephalochordate embryos hence representing a great proxy for the stem chordate embryo[122].

Gene content and rate of evolutionary diversification

The first amphioxus genome, from *B. floridae*, was published in 2008 and it represented a great tool for the understanding of chordate evolution as well as vertebrate diversification. In general, the *B. floridae* genome contains a basic set of chordate genes involved in development and cell signaling, and its detailed analysis served to identify several large genome segments that are conserved in both amphioxus and vertebrate genomes, suggesting the chordate ancestry of these protochromosomes[31,123]. Macrosynteny analyses between amphioxus and vertebrates also brought new insights on the quadruple conserved synteny in vertebrates and reinforced the idea of a potential 2R WGD as a major instrument driving vertebrate genome evolution [31]. Furthermore, putative primitive enhancers with conserved functions between amphioxus and vertebrates were identified and thus suggest an ancient chordate origin of the regulation by some non-coding elements[123,124]. However, the amphioxus genome also includes lineage specific expansions of particular gene families, for instance, the impressive expansions of the innate immune complexity genes toll-like receptors, NOD/NALP-like receptors, and scavenger receptors[125], the vast expansion of some tyrosine kinase families[126], as well as extensively amplifications in the short-chain dehydrogenase/reductase superfamily[127].

The publication of the genome of another amphioxus species, *B. belcheri*, suggests that ancient vertebrates, existing before the split from jawed and jawless vertebrates, went through high rates of protein evolution, genome rearrangement and domain shuffling. This conclusion arose from the fact that modern vertebrates have less protein diversity and nucleotide polymorphisms, fewer domain combinations and conserved non-coding elements, as well as lost substantial transposable element diversity, when compared to both amphioxus genomes[128]. Furthermore, the *B. lanceolatum* genome will soon be released and a draft version is already available to be used by members of the genome consortium[129].

Recently, the study of the transcriptome of larvae and adults of another amphioxus species, this time not form the genera *Branchiostoma*, from *Asymmetron lucayanum*, showed that cephalochordates are evolving more slowly than the slowest evolving vertebrate[130]. Although, genes associated with innate immunity stand out as evolving relatively quickly in cephalochordates. In summary, the slow evolving amphioxus genome, despite some minor exceptions, appears to be a good surrogate for the ancestral chordate genome, not only in terms of gene content and exon–intron gene structure, but also in terms chromosomal organization.

Amphioxus general morphology

The name amphioxus originates from its characteristic shape and it means sharp at both ends (amphis = both, oxys = sharp). This morphological shape is marked by to the presence of a rigid, yet flexible dorsal notochord that is present along the A-P axis (Fig. II.2, II.3)[114]. The amphioxus notochord is made of specialized mesoendoderm discoidal cells, stacked like coins and each cell contains transverse myofilaments. During notochord formation, this arrangement is attained rapidly by inconspicuous cell movements [83,131]. On the dorsal side the notochord contacts with the neural tube and it serves as a mechanical swimming organ[96]. In contrast to vertebrates, where the notochord is lost during development, it represents the only skeletal structure of amphioxus[114].

The amphioxus epidermis consists of a single layer of prismatic secretory epithelial cells, covered by a very thin layer of muco-polysaccharids and proteins[132]. Globally, the mouth is located at the anterior end of the animal where it is surrounded by about 20 buccal tentacles (cirri) and the anus is located near to the posterior end, opening ventralwards but left to the mid-line (Fig. II.3 A-B)[114]. The mouth itself consists of two portions, an outer vestibule and an inner *apertura oris*; the latter is surrounded by a sphincter muscle, which forms the so-called velum, and the inner surface is characterized by a specialized epithelium, called the wheel organ which consists of tall strongly ciliated cells that generate a water current carrying food into the mouth[133]. Anteriorly, the wheel organ has a dorsal invagination known as Hatschek's pit (originated from the larval pre-oral pit) that might be homologous to the vertebrate adenohypophysis[96,133]. The pharynx is perforated by numerous pharyngeal gill slits and possesses two main grooves, dorsally the hyperpharyngeal groove and ventrally the endostyle, which transport food particles and secrete mucus, digestive enzymes and other proteins[114,134]. The endostyle has been proposed as an homologous structure to the vertebrate

thyroid gland[84,97,135]. The amphioxus digestive system also comprises a hepatic caecum located near the mid-gut, which is probably related to the vertebrate liver[136,137]. The mentioned organs of the digestive system as well as the gonads are enclosed in a body cavity called the atrium, which runs along two thirds of the animal and has a distal atriopore opening (Fig. II.3 B-C)[138].

The muscular system of amphioxus (Fig. II.3 A and C) is composed mostly by striated muscle fibers that are divided in two main groups: parietal muscles constitute the myotomes or muscular segments of the body and visceral muscles that are non-segmented muscles. V-shaped myomeres that are arranged along the lateral flanks of the body, with a slight left-right offset[138]. Despite the absence of skeletal tissue in amphioxus, collagen-based connective tissues with few or no cells are found in the positions where one would find the axial skeleton in vertebrates. Myomeres are separated by collagenous myosepta and as myomeres contract sequentially during the swimming movement, the generated force is transmitted to the notochord via these axial connective tissues[139]. On both sides of the trunk myomeres and myosepta extend ventrally to the metapleural folds (folds of the body wall that contain coelomic space) and run from the oral region to a position just posterior to the gonadal region[114]. There are no paired muscular fins in amphioxus, but running along the whole length of the back, starting from rostral tip of the notochord, there is a median ridge known as dorsal fin. At the posterior end, the dorsal fin becomes enlarged to form the so called tail fin and the ventral domain of the tail is continued round the hinder extremity of the body past the anus until the atriopore. Globally, the fin is supported by gelatinous fin-rays enclosed in their own chamber, except in the dorsal region of the tail where two fin-rays can be identified per chamber[134].

Despite the absence of a heart-like pumping organ in amphioxus, the vascular system (Fig. II.3 D) is rather complex and it includes a real circulatory irrigation of most organs, much alike the vertebrate vascular system[140]. Blood is colorless due to the absence of respiratory pigment and corpuscles. It seems that the main function of the closed amphioxus circulatory system is related with the transport of food and excretory products rather than gaseous components. There are no structural differences between the so called arteries or veins. The ventral aorta runs mid-ventrally to the pharynx wall bellow the endostyle giving rise to a series of paired branchial arteries that run through each primary gill bar of the pharynx. Dorsally, they are connected to the lateral dorsal aorta on each side, but before doing so, branchial vessels split into a small capillary network that is associated with the nephridia.

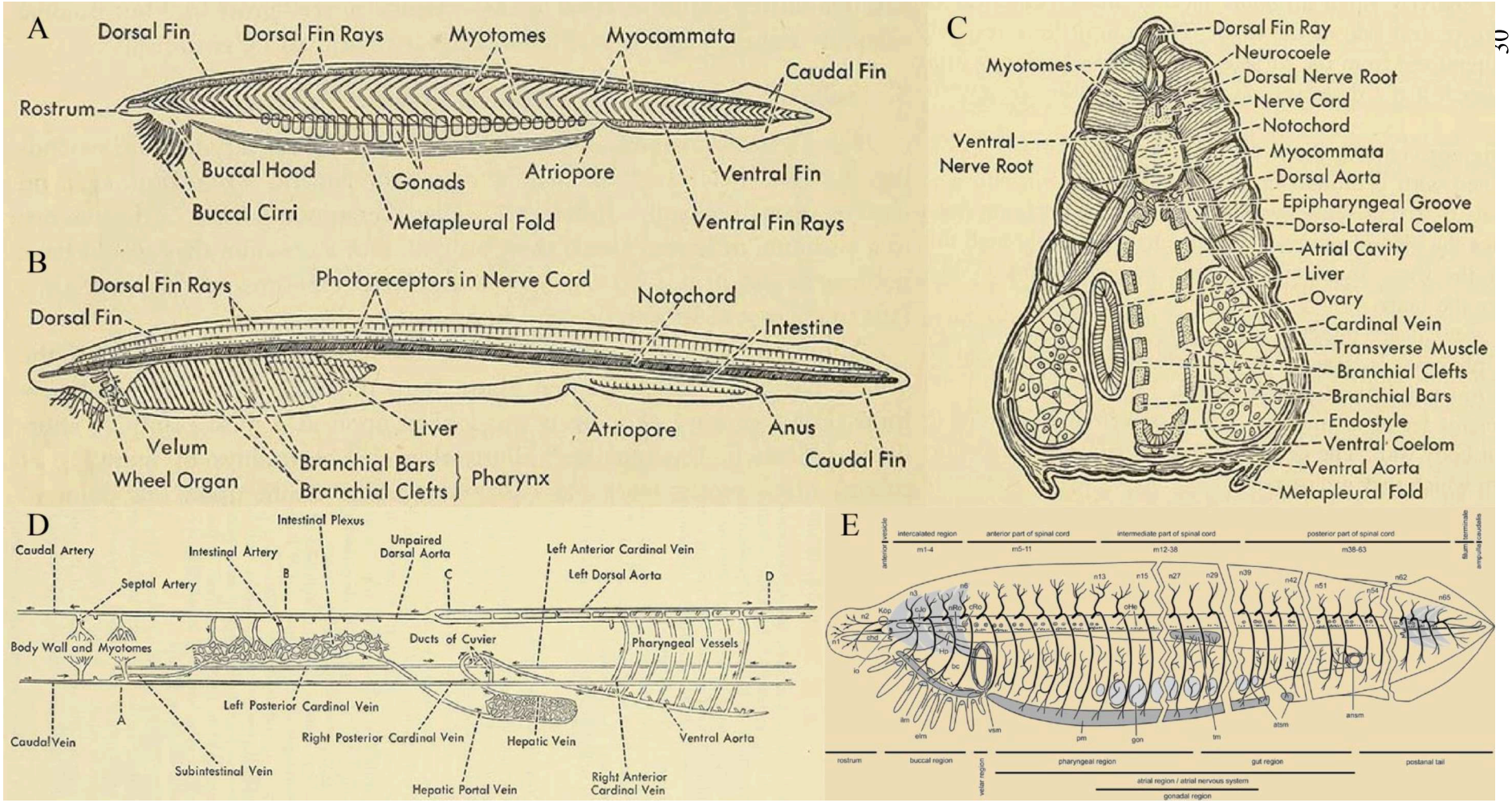


Fig. II.3 – Diagrams representing the general morphology of amphioxus. *(A) side view of the entire animal; (B) side view showing internal features (C) cross section at the anterior part of an adult amphioxus; (D) schematic representation of amphioxus circulation; and (E) schematic lateral view of amphioxus central and peripheral nervous systems organization in relation to peripheral structures. Abbreviations: alm- anterolateral migrated cell group; ansm- anal sphincter muscle; bc- buccal cavity; chd- notochord; cJo- Joseph cells; cRo- Rohde cell; elm- external labial muscle; gon- gonad; Hp- Hatschek's pit; ilm- inner labial muscle ►*

► *io- infundibular organ; Köp- Kölliker's pit; m1, m2, m3…- myomere 1, 2, 3…; n1, n2, n3…- dorsal nerve 1, 2, 3…; nRo- nucleus of Rohde; oHe- organ of Hesse; pm- pterygeal muscle; tm- trapezius muscle of atriocoelomic funnel; vsm- velar sphincter muscle. (A-D) from Walter and Sayles 1949[141] and (E) adapted from Wicht and Lacali 2005[142].*

Dorsal aorta then runs on both sides of the pharynx and anteriorly it continues as internal carotid arteries supplying the oral hood. Posteriorly both sides of the dorsal aorta fuse just behind the pharynx and enter the tail region as the caudal artery. In addition, the dorsal aorta also supplies the intestinal arteries that form a capillary network in the intestine wall (intestinal plexus). The blood from the caudal region is collected in the so called caudal vein that runs into the subintestinal vein. This vein also collects the blood from the intestinal plexus. The subintestinal vein then continues anteriorly into the midgut diverticulum as a network of capillaries that join thereafter into the hepatic vein. Four other main veins, two cardinal and two parietal, collect the blood from ventral-lateral and dorsal regions of the body respectively. The sinus venosus, located in the posterior end of the pharynx, collects all the blood coming from numerous veins in the body and pumps the blood forward to the ventral aorta[140,143].

The cephalochordates have segmental excretory organs that are positionally similar to the excretory tubules of primitive vertebrates. Instead of opening into a common archinephric duct, however, each nephridium opens independently via a short tubule into the atrium[144].Globaly the excretory system of amphioxus is composed by the Hatschek's nephridium (long tube in the roof of the oral hood), protonephridia (simple, closed ciliated, sac-like tubes located above each gill slits), brown funnels (two sac-like structure located t the posterior end of the pharynx, dorsally) and renal papillae (groups of cells located on the floor of the atrial cavity)[143]. The unpaired Hatschek's nephridium is the largest and most anterior of the nephridia in the body and its single collecting tubule lies on the left side of the notochord along the left anterior aorta. It develops from the first somite on the left side, in association with the larval mouth, and represents a homolog of the pronephros of basal vertebrates[145,146]. The protonephridia are richly supplied with blood vessels and probably excretion occurs by simple diffusion, these structures probably have an origin that derives from ectoderm and also mesoderm[131,144]. Brown funnels and renal papillae probably have an excretory function[143].

Located directly dorsally to the notochord, the hollow nerve cord of amphioxus begins at the anterior limit of the first myomere and extends to the tip of the caudal fin (Fig. II.3 E). This nerve cord shows no external structuring, apart from a caudal ampulla and serially

repeated dorsal nerves that lack ganglia[142]. Although, during embryonic development and at the larval stage a transient thickening of the cerebral vesicle (brain) is visible. In adult amphioxus, the cerebral vesicle corresponds to the portion of the neural tube that is rostral to the first myomere and stretches to myomere four. The cerebral vesicle mainly consists of small ciliated epithelial cells and includes, amongst others, GABAergic and serotonergic neurons as well as pigment cells of the frontal ocellus[142,147]. On the dorsal left side of the cerebral vesicle a vestige of the larval neuropore (Köllikers pit) still connects the neural tube to the outside, by means of a narrow channel clogged with cilia[142]. The infundibular organ, at the ventral posterior end of the anterior vesicle, comprises several rows of columnar cells that secrete Reissner's fiber. This fiber extends though the lumen of the neural tube into the caudal ampulla[142]. Despite the lack of external segmentation, the amphioxus CNS contains regions with distinct internal organization and cell type composition, such as the intercalated region that is marked by the presence of putative photosensory Joseph cells and lamellar cells[142,147]. Projecting from the anterior nerve tube the first two nerve pairs (nerve 1 and 2) have clusters of primary sensory cells on their branches enclosed in capsules, called "corpuscles of the Quatrefages". Most likely these corpuscles serve as mechanosensory organs and occur only in the rostral connective tissue that surrounds the notochord of amphioxus[142]. Nerve pair 1 further innervates the tip of the rostrum, which is thought to contain special sensory cells. The posterior-most dorsal nerves pass laterally through the myomeres septa and project their branches dorsally and ventrally towards the sub-epidermal nerve plexus and the atrial nervous system of amphioxus[142]. Nerves 3 to 6 are connected to the buccal and the velar plexus, two peripheral nerve rings that innervate the region of the buccal cavity as well as muscular structures in the velum, and the posterior-most nerves 51 to 53 innervate the anal sphincter muscle[114,142].

In amphioxus sexes are separated, although there is no sexual dimorphism between males and females. Gonads are arranged metamerically in two rows and they are situated ventro-laterally from the middle of the pharynx up to the atriopore (Fig. II.3 A). Gonads are simple hollow sacs of mesodermal origin. Mature ovaries and testes can be identified under a stereomicroscope by their different inner textures[134,148]. Mature amphioxus sperm is among one of the smallest of the chordates with only 18µm, while the eggs have a diameter of around 100-140µm. Mature gametes are discharged directly in the atrium, being ejected at the atriopore by the action of the water current[143].

Amphioxus development (Fig. II.4)

Fertilization, cleavage and gastrulation

Mature oocytes undergo the first meiotic division with formation of the first polar body and arrest at second meiotic metaphase[119]. The second meiotic division of the oocytes is completed within 10 min after fertilization, with the formation of the second polar body and with the migration of the maternal chromosomes to one side of the animal pole[149]. The egg is semi-opaque due to the high quantity of granules uniformly distributed throughout the entire egg, and it is surrounded by a cellular membrane called the vitelline layer[114]. The sperm can enter anywhere in egg surface, but preferentially at the vegetal pole. Independently of the entry point, the sperm nucleus will first migrate to the vegetal half and only then will join the maternal chromosomes in the animal pole[96]. As soon as the fertilization occurs the vitelline layer detaches from the body of the egg and largely expands, giving rise to the fertilization envelope[150]. Cleavage, gastrulation and neurulation prior to hatching occur within the fertilization membrane[96]. Also, very early after fertilization, the already present sheets of endoplasmic reticulum associated with RNA become condensed into a whorl. This structure was proposed to constitute the gem plasm of amphioxus, since the germ cell markers *nanos* and *vasa* are expressed in this territory[151].

Amphioxus cleavage is radial holoblastic. The first cleavage starts from the animal pole (identifiable by the polar body position) and gives rise to identical shaped blastomeres (2-cell stage). Each of the first two blastomeres can give rise to a complete larva, but only one will inherit the germ plasm[119]. The second division is also meridional and at right angles from the first one, originating four identical blastomeres (4-cell stage). The blastomeres are not tightly adherent resulting sometimes in twins and even quadruplets. The cleavage continues by an equatorial division that forms four animal blastomeres and four vegetal ones (8-cell stage), with the latter being visibly larger than the former[152]. The 16-cell stage is obtained by a meridional cleavage, followed by an equatorial cleavage of each blastomere (32-cell stage). At this stage, the embryo is constituted of a single layer of cells that surround a central cavity named blastocoel. From this point on the blastomeres will keep dividing in a more or less regular pattern, giving rise to 64-cell and 128-cell stages. The 8th cycle of division is marked by the initiation of asynchronous cell division and by this stage the blastula is fully formed (B stage)[134,152].

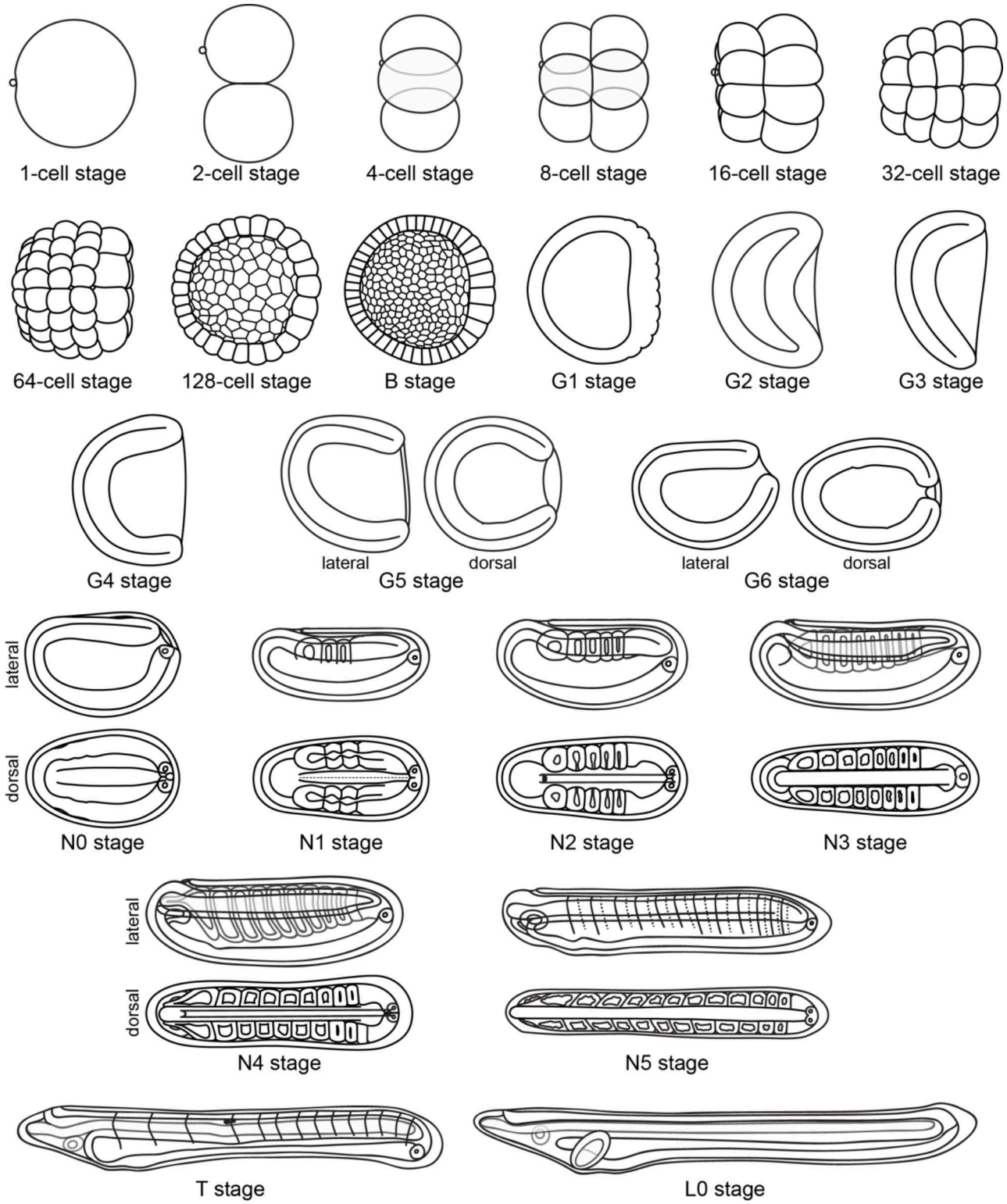


Fig. II.4 – Schematic representation of amphioxus development. *Animal pole/anterior side presented in the left. Drawings adapted from Hatscheck's original description of* B. lanceolatum *embryonic development*[153].

The lower vegetative cells of the hollow blastula of amphioxus are somewhat larger than the rest of the blastula-wall[96,114]. It is exactly this side of the embryo that become flat at the onset of gastrulation (G1 stage) and that slowly will be invaginated into the blastocoel, forming a depression at the vegetative level of the developing early gastrula (G2 stage)[154]. The invaginating layer of cells represents the presumptive endomesoderm, while the outer layer is mainly ectoderm. This cell progression will continue compressing the blastocoelic cavity and ultimately will originate a cap-shaped like embryo without blastocoel, with a deepened archenteron and distinct blastoporal lips. At the mid-gastrula stage (G3 stage), the embryo is a double-layered sac and the blastopore has a diameter of approximately 50% of the embryonic body[154]. As gastrulation progresses archenteron cavity becomes deeper, the blastoporal opening narrows down and thus the blastoporal lip is more sharply inflected, forming a cup-shaped (G4 stage), followed by a vase-shaped gastrula (G5 stage)[154]. In the lateral view of the vase-shape embryo a small difference between the dorsal and ventral side is now observable, with the former having a more flattened form and the latter exhibiting a slight convex shape[114]. These differences between dorsal and ventral side become more pronounced as the gastrula elongates and, as development proceeds, the blastopore also becomes very small (G6 stage). This effect is due to the downward and backward growth of the dorsal blastopore border, while the ventral lip remains stationary. At this late gastrula stage, the embryo possesses a bottle-shape and the blastopore is inclined towards the dorsal side of the embryo, which seems to be a synapomorphy of the Chordata[114].

Expression patterns of marker genes such as *Hox1*, *Hox3*, *Otx* and *Foxq2* determined that, with the exception of the tissue immediately around the blastopore, most of the gastrula is destined to give rise to the amphioxus head. This comprises the dorsal diencephalic forebrain, small midbrain and hindbrain, the anterior 8 to 12 somites, the pharynx, mouth and gill slits as well as the anterior section of the notochord[119]. Furthermore, transplantation experiments suggested that there might be a gastrular organizer in amphioxus, similar or equivalent to the Spemann-Mangold's organizer in vertebrates[155,156] Fact that is further supported by studies addressing Spemann's organizer homolog genes in amphioxus (e.g. *Nodal*, *Lefty* and *FGF8/17/18*) and showing a correlation of their expression and a specific location at the dorsal lip of the blastopore[157,158]. Furthermore, in developing amphioxus addition of vertebrate BMP4 induced loss of notochord and nerve chord, upregulation of *Nodal* converted ectoderm to neuroectoderm and inhibition of FGF signaling at the onset of neurulation inhibited somite formation[119,158,159].

The neurula (Fig. II.5)

Following the late gastrula stage all ectodermal cells develop cilia and the embryo starts to rotate within the fertilization envelope before the formation of the somites, marking the onset of neurulation (N0 stage)[96,160]. At this point the embryo shows a typical dipoblastic organization, where the dorsal ectoderm, now flattened and converted in neural plate, starts to show indications of a shallow longitudinal groove. The early neurula stage (N1 stage) is characterized by the establishment of the first somites (i.e. 1 to 3). Mesoderm that is initially dorsal starts forming three folds, a medial one that develops into the notochord and two dorso-lateral ones associated with the archenteron wall that give rise to the anterior somites, which at this point start to pinch off in an anterior to posterior direction. The dorsal epidermal ectoderm starts to dissociate from the neural plate and ectodermal cells from either side start migrating over the neural plate via lamellipodia and will fuse at its dorsal midline, except at a small region located at the anterior-most position of the neural plate (i.e. neuropore)[119,153,161].

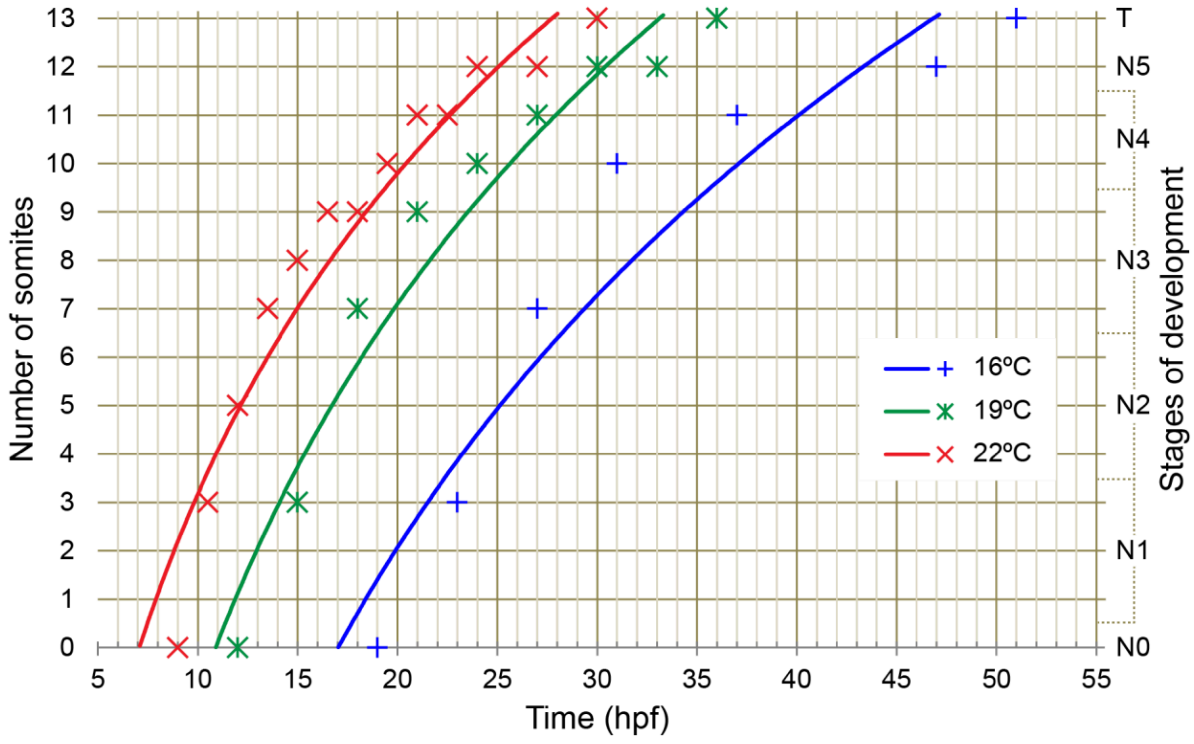


Fig. II.5 – Relationship between neurula/transition stages and number of somites in amphioxus. B. lanceolatum *relationship between number of somites, stage of development and time (in hours post-fertilization - hpf) in embryos developing at three different temperatures: 16ºC, 19ºC and 22ºC.*

The archenteron is no longer contacting with the exterior, but it communicates with the forming neural tube. Moreover, the blastopore has now become the neurenteric canal, which connects the CNS with the presumptive gastric cavity[114]. The embryos keep elongating by the addition of new somites having now 4 to 6 pairs (N2 stage) and burst through the fertilization envelopes (i.e. hatching), thus freely and random swimming by means of ciliary movements. Individual structures that were before indistinguishable, are now becoming individualized, namely the neural fold, mesoderm and notochord. Furthermore, the neural plate shrunk ventrally and has, at this point, a V-shape and the primordium of the notochord is a round mass of cells that expand dorsally under the neural plate. Clear differences between mesodermal structures and the neural tube are observable, although notochord and somites boundaries are rather faint[162].

At the mid-neurula stage (N3 stage), the embryo is characterized now by 7-9 somites pairs and it is observable the closure of the neural tube, although the structure will only later become completely circular when seen in transverse section. Similarly the notochord is now fully individualized from the somites, except at the most anterior ones[161,163]. At this point, ventral extensions of the somites start to generate the lateral and ventral coeloms as well as the musculature on the floor of the atrium[119]. The archenteron region situated just anterior of the first pair of somites starts expanding by the means of a dorsal lobe that forms a slight constriction in both sides. At the following stage (N4 stage), somitogenesis is clear occurring asymmetrically in both sides of the embryo and, contrarily to what happened until this stage, they originate from the tailbud. Thus, while early amphioxus somitogenesis occurs by means of enterocoely, later somites are formed schizocoelically[96]. Furthermore at this stage characterized by 10-11 somites, anterior constrictions of the archenteron (i.e. dorsal diverticulum) are now two independent sacs, the left and right head cavities or diverticula[114,134].

As development proceeds the rate of somite formation slows down and at the stage characterized by 12 pairs of somites (i.e. N5), left and right diverticula are now asymmetrically organized: while the left diverticulum maintains roughly the original form and dimensions, the right one move anteriorly and become larger and his epithelial walls are now more flat[114]. At this late neurula stage, the primordium of a very particular structure, the club shaped gland, is identifiable in the ventral side of the anterior endoderm towards the left side of the embryo. Furthermore, Hatschek's nephridium early marker genes start to be expressed in the corresponding region, a thickening of the mesothelial wall of the first somite on the left side[164,165]. Globally, proliferative activity becomes patchy in the somites and is limited to the

ends of the notochord, but continues in the endoderm, in the anterior neural plate as well as in the tailbud[166]. In a later embryonic stage, with 13 pairs of somites, the embryo exhibits now a transitional morphology between neurula and larva (T stage). A quite marked A-P differentiation is recognizable, as the anterior region is enlarged at the level of pharynx. At this point of development, embryos start to twitch rather than moving by ciliary movements, despite their long cilia surrounding the entire embryo[162]. The first pigment in the nerve tube appears at this stage and it is located at the level of the fifth somite pair. It is a black pigment-spot located in a cell at the ventral wall of the neural tube, followed by another smaller pigment posteriorly located in relation to the first one[114]. Furthermore, the caudal ectodermal tail fin starts to be apparent[162].

The larva

At the onset of the larval stage (L0 stage), the entire embryo is longer, although this difference is not necessarily associated with the acquisition of new somites, but with the expansion of existing ones. Generally, the body becomes slender ending with a tail. In the anterior region, the rostrum elongates originating the snout and the right diverticulum kept expanding giving rise to the cavity of the snout, below the notochord. Differently, the left diverticulum, soon to be the pre-oral pit fuses with the left sided ectoderm no opening exists yet, also the mouth anlage is present at this stage[167]. In amphioxus, the larva (L1 stage) is rather asymmetric with very specialized structure in both left and right side of it. Thus, the mouth of amphioxus larva is considered to be a modified gill slit that opens on the left side of the developing animal and originates by fusion of the ectoderm and endoderm. Still on the left side, the pre-oral pit, located just above the ciliary tuft, is now fully formed. This very particular structure, after metamorphosis will become mostly the Hatschek's pit (i.e. the homologous of the adenohypophyses)[168] and its functions are still a matter of a debate, between a sensory organ and a secretory gland[169]. Furthermore, the larval kidney (Hatschek's nephridium) is also located on the left side of the larva, between the ectoderm and the anterior-most somite[161]. In contrast, the club-shaped gland that is initially a thickening of the endoderm in the right wall of the archenteron forms[170]. It is localized almost opposite but slightly posterior to the mouth and extends till an ectodermal opening on the left side, just anterior to the mouth[171]. At the ectodermal exit of the duct the cells are characterized by tall cilia which beat with undulating movements and create a water current from the exterior into the organ[172]. The club-shaped gland is an organ strictly larval and it has been suggested to

function in feeding, since it secretes mucoproteins[96]. Also present on the right side of the pharynx of amphioxus larva is the endostyle. This organ arises as a hexagon like thickening of the endodermal wall just anterior to the club-shaped gland that finishes as a bean-shaped structure. The function of the endostyle is associated with the secretion of mucous that traps food particles and it has been proposed to be an homologous structure to the thyroid gland in vertebrates[84,97,135]. Furthermore, the lower edge of the pharynx wall coalesces with that of the body to form the first gill-slit. If initially this structure is formed at the ventral midline of the larva, it shortly migrates to the opposite side to the opening of the mouth[96]. At this stage of development the anus arises at the posterior end, just anterior to the ectodermal caudal fin, and similarly to the primary gill slits, it originates at the ventral midline and migrates towards the left side[171]. With the establishment of all the structures referred above the one gill-slit larva (L1 stage) is now able to feed. After the L1 stage several changes continue to happen, firstly the mouth opening keeps expanding towards the posterior side, giving rise to an elliptical form and later on lens shaped. New slits are then added one behind the other until the metamorphosis: L2 stage, for 2 gill-slits, L3 stage for 3 gill-slits and so on. Although, metamorphosis occurs at different larval stages (different number of primary gill slits) in different amphioxus species[173–175].

Amphioxus developmental staging

Since the 19th century amphioxus embryology has been characterized and initial studies were performed using developing *B. lanceolatum* embryos obtained in Naples, Italy[176]. This study was then followed by a series of other very complete and detailed embryological studies, using the same species[161,163,177]. More recently, *B. japonicum* early development has been characterized and described by means of electron microscopy[152,154,162]. The same kind of approach has been used to characterize neurulae, larva and even post-metamorphic larva of *B. floridae*[178]. Furthermore, the embryonic development of *A. lucayanum* until the larval stage has been described using classical Nomarski microscopy[179,180]. Globally, until the neurula stage no striking differences are observed between *Asymmetron* and *Branchiostoma* development, although at the mid neurula stage (N3 stage) and the following stages no anterior left or right diverticula are formed in *Asymmetron*[179]. Differently, *Branchiostoma* species have a very similar early embryonic development, with only minor differences between each other.

			A. lucayanum	*B. floridae*		*B. japonicum*	*B. lanceolatum*		
Stage	**Key feature**	**Other names**	***at 27ºC***	***at 24ºC***	***at 30ºC***	***at 25ºC***	***at 16ºC***	***at 19ºC***	***at 22ºC***
1-cell stage	fertilized egg	-							
2-cell stage	2 cells	-	1h-1h30min*	45min	30min	50min-1h	1h30min	1h	
4-cell stage	4 cells	-	2h-2h30min*	1	50min	1h10min	2h30min		
8-cell stage	8 cells	-	2h-2h30min*	1h30min	1h	1h35min	3h		
16-cell stage	16 cells	-	3h-3h30min*	2h	1h15min	1h55min			
32-cell stage	32 cells	-	3h-3h30min*	2h15min	1h30min	2h15min	3h45min		
64-cell stage	64 cells	-	4h-4h30min*	2h30min	1h45min	2h35mpf			
128-cell stage	128 cells	-	4h-4h30min*	3h	2h	3h10min			
B stage	initiation of asynchronous cell division	blastula	5h	4h	2h30min	3h20min		5h	
G1 stage	flaten vegetal pole	onset of gastrulation	nd	4h30min	3h30min	3h40min		6h	
G2 stage	invaginated vegetal pole	early gastrula	nd	5h	nd	4h10min		7h	
G3 stage	cap-shaped	early-mid gastrula	nd	6h	4h	5h35min		8h	
G4 stage	cup-shaped	mid gastrula	9h	7h	4h30min	6h-6h20min	15h	9h	
G5 stage	vase-shaped	mid-late gastrula	nd	nd	nd	7h40min		10h	
G6 stage	bottle-shaped	late gastrula	nd	8h	5h	8h50min		11h	
N0 stage	0 somites, neural plate	onset of neurulation	12h	8h30min	6h	nd	19h	12h	9h
N1 stage	1-3 somites	early neurula	15h	9h30min	6h30min	10h30min	23h	15h	10h30min
N2 stage	4-6 somites	early-mid neurula	19h	10h30min	7h30min	13h			12h
N3 stage	7-9 somites	mid-neurula	nd	12h	nd	nd	27h	18h-21h	13h30min - 18h
N4 stage	10-11 somites	mid-late neurula	nd	15h	nd	18h	31h-37h	24h-27h	19h30min-22h30min
N5 stage	12 somites	late neurula	32h ?	nd	nd	nd	47h	30h-33h	24h-27h
T stage	13 somites, mid-body pigment spot	late embryonic/ transition stage	50h	20h-24h	17h	24h	51h	36h	30h
L0 stage	pre-oral pit and mouth enlagen		nd	30h	21h-24h	nd		42h	
L1 stage	1 gill slit, mouth and anus opened	ealy larva	72h	36h	28h	36h		48h	
L2 stage	2 gill slits	2 gill slits larva	nd	42h-72h	36h	48h			
Ln stage	n gill slits	-							

*Table II.1 – Development stages and correspondent time post-fertilization in 4 amphioxus species. Developmental times obtained from the literature (*A. lucayanum[410], B. floridae[405,409,411,412] *and* B. japonicum[383,385,393,412]*) and from lab observations (*B. lanceolatum*, Fig. III.5).*

Nowadays, though, multiple species of amphioxus are used in different studies that address developmental processes from a molecular and genetic point of view[97]. Notwithstanding the amount of data collected in amphioxus species development, different labs culture embryos at different temperatures and use either an "in-house" defined nomenclature or simply mention time post fertilization to designate a given stage. For instance, some authors suggest that the blastula stage is established as early as the entire blastocoel in enclosed by cells (64-cell stage)[174], while a different view defines the blastula stage only after the 8th round of synchronous cell divisions, i.e. following the 128-cell stage[152]. Another controversial nomenclature is associated with the transition between the gastrula stage and the neurula stage, this point of development is sometimes referred to as a very late gastrula[154] or a very early neurula[160,181]. Moreover, some authors claim that the larval stage starts when "tissues and cells prepare for performing their own function" in young individuals, which actually corresponds to the newly-defined T stage[162]. Alternatively, some other authors suggest that the beginning of the larval stage is marked by the opening of the mouth (L1 stage)[96]. This fact artificially complicates the amphioxus literature and renders it cryptic, in particular when trying to compare results obtained from different species or sometimes even using the same amphioxus species and comparing results from different labs. Therefore, establishing an easy and systematic classification for the embryonic development that could be applied to the different species of amphioxus seems essential.

To achieve such a task, it is crucial to review the entire embryological literature of amphioxus along with recent molecular studies and define stereotypical stages of development, such as the ones mentioned while describing amphioxus development above. Using as base the detailed studies made by Hirakow and Kajita[152,154,162], as well as their original nomenclature on *B. belcheri*, it is important to expand, for instance their classification in terms of neurulae stages. In Table III.1 are summarized the development stages defined above, including 10 stages addressing cleavage and gastrulation, 6 time points during the gastrulation process, 7 stages that define the development of neurulae and pre-larva amphioxus, as well as 3 early stages in larval development. In order to define a systematic classification, the correspondence between developmental stages and the time post-fertilization, based in the literature, at a given temperature is shown for 4 different amphioxus species. One of the major improvements of the following classification is: (1) a more detailed description of the neurula stages, here being divided in 6 well defined time points, contrasting with the initial 3 stages proposed by Hirakow and Kajita[162] and the 4 different neurula stages defined by Lu and collegues[160]; (2) the inclusion of a description of a transition stage between

neurula and larva, called T stage and (3) the definition of L0 stage as the onset of the amphioxus larva life.

Chapter III: Bringing amphioxus to the bench-side

Cephalochordates, commonly named amphioxus or lancelets, are small fish-like invertebrate marine animals that are widely distributed in tropical and temperate seas. Almost all the described cephalochordate species live in shallow coastal waters with sandy and shell-sand habitats. Amphioxus spend most of their lives buried in the sand, filter-feeding through jawless, ciliated mouths[96]. The sexes are generally separate, and they reproduce periodically during the summer, coming up out of the sand they swim in the water column, after sundown, to shed their gametes. Amphioxus embryos and larvae are planktonic in both inshore and offshore areas, and development is considered to be direct, with a gradual metamorphosis[84,122,182].

Cephalochordates are one of the three subphyla of the chordates, along with tunicates (also known as urochordates) and vertebrates. They were conventionally believed to be the closest living relatives of vertebrates, however recent molecular phylogenetic analyses have placed tunicates as the sister taxon of vertebrates and cephalochordates as the most basal chordate group[87]. The cephalochordate subphylum comprises about 30 species that are subdivided in three different genera: *Branchiostoma*, *Epigonichthys* and *Asymmetron*[130,179,183–186]. Furthermore, whole mitochondrial genome analysis indicate that within the cephalochordates, *Asymmetron* has a basal position with a divergence time of 162 Mya from the *Epigonichthys* and *Branchiostoma* clade, and that *Epigonichthys* diverged from *Branchiostoma* at 120 Mya[183], whereas within the genus *Branchiostoma*, Pacific and Atlantic species have a divergence time of 112 Mya[187].

Cephalochordates are characterized by the typical chordate features, such as the dorsal hollow nerve cord and notochord, the perforated pharynx with pharyngeal gill slits, the musculature of the body divided into V-shaped blocks, or myomeres, and the presence of a post-anal tail. Amphioxus also possesses a perinephric kidney, an endostyle (homolog of the vertebrate thyroid gland) and a pre-oral pit (equivalent to the vertebrate adenohypophysis)[98,99]. Since the 19th century, when these clear affinities between cephalochordates and vertebrates were first identified, amphioxus has been used as model to understand vertebrate evolution and development[122,161,188,189]. More recently, along with the advent of modern molecular biology and genomics, cephalochordates have become key model

organisms to clarify chordate, and even deuterostome, genome evolution, since its genome appears to be a good surrogate for the ancestral chordate genome with respect to gene content, exon–intron gene structure and even chromosomal organization[31,123,128,129]. Additionally, recent studies have placed amphioxus as an excellent animal model to study both innate and adaptive immunity[190], as well as organ regeneration[191].

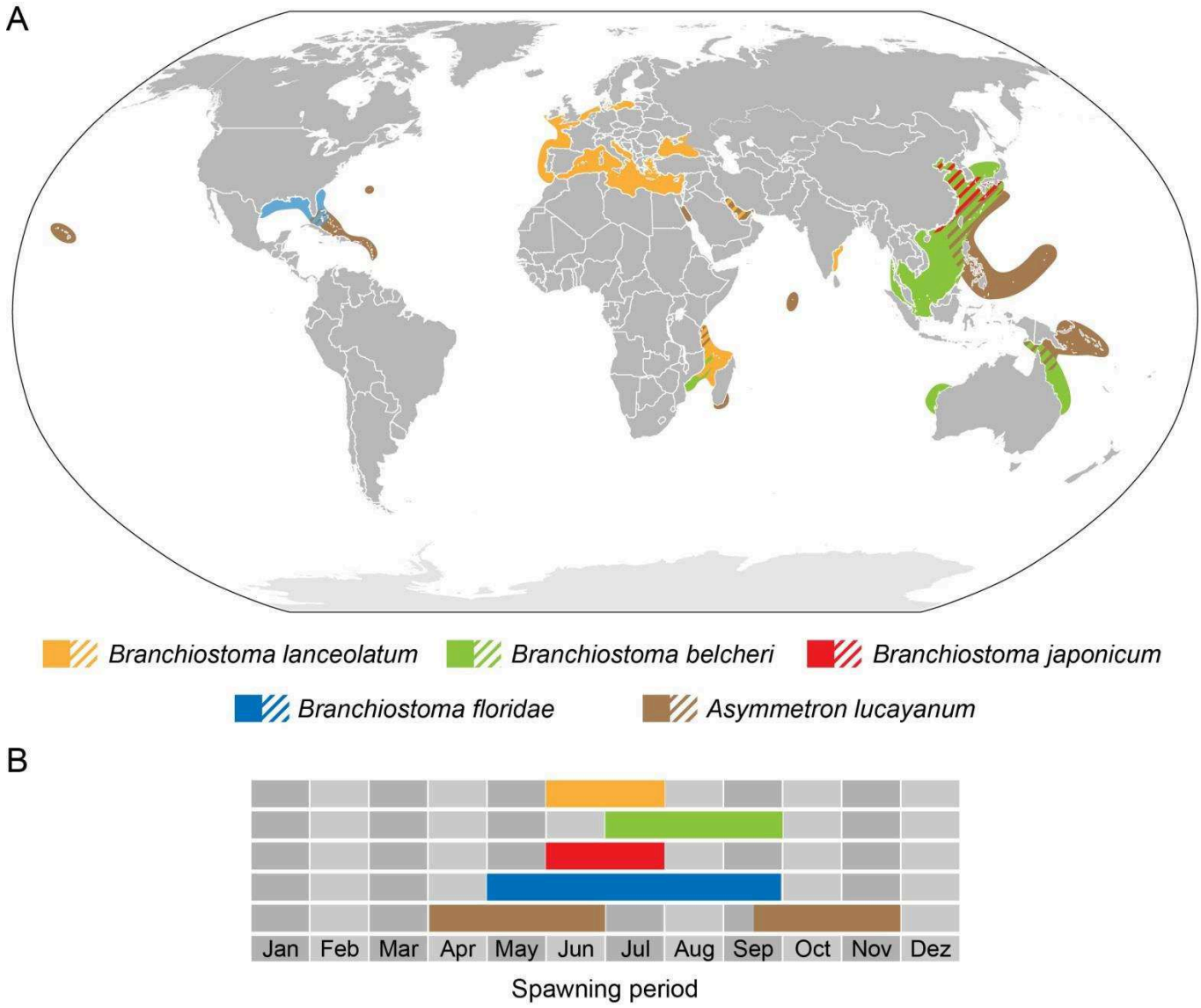


Fig. III.1 – Amphioxus model species distribution and spawning season. *(A) World distribution., where distribution areas correspond to a merely representation based on discrete point location[185,192] (B) Natural occurring spawning season.*

Amphioxus model species

Amphioxus, since it was first described in the 18th century[193], has become a very important animal model to understand the origin of vertebrates from an invertebrate perspective. While most of the early descriptive work done in amphioxus embryonic development and anatomy was performed using the European species (*Branchiostoma lanceolatum*), current studies also use other species of amphioxus, such as *Branchiostoma floridae*, *Branchiostoma belcheri*, *Branchiostoma japonicum* and *Asymmetron lucayanum*. In general, adult amphioxus from different species can be distinguished morphologically by their symmetry, number of muscle segments, dorsal and anal fin chambers, and pharyngeal slits[185].

Branchiostoma lanceolatum

Initially named *Limax lanceolatus*[193], was the first amphioxus species to be identified. Despite the fact that this species was erroneously assigned to the group of the mollusks, in the early 19th century, Costa and Yarrell identified its affinities with fishes and agnathans, placing it as a Chordate[194,195]. This species are mainly widely dispersed along the European coast and the Mediterranean Sea (Fig. III.1 A). The natural breeding season from the population obtained at Argelès-sur-Mer spawns from end of May till end of July[196] (Fig. III.1 B). The first embryological studies done in amphioxus, were performed using *B. lanceolatum* developing embryos obtained in Naples, Italy[176], followed by a series of others, very complete and detailed, embryological studies[153,161,163,177]. Multiple genomic and transcriptomic tools will soon be available[129]. *B. lanceolatum* is currently used as a model organism to study evo-devo and regeneration in several research laboratories around Europe.

Branchiostoma belcheri and Branchiostoma japonicum

First described by the end of the 19th century[197] and called *Amphioxus belcheri*, this species initially collected in Borneo has a large dispersion mostly in the coastal area of Asia (Fig. III.1 A). Recent studies identified that this initial description correspond to two distinct species, *B. belcheri* and *B. japonicum*, and that they share partially the same habitat space[186] (Fig. III.1 A). Although, the reproductive period is slightly different, since *B. belcheri* collected at Xiamen spawns in the wild from July to September and *B. japonicum*, collected in the coast of Japan, during the months of June and July[198] (Fig. III.1 B). These species were used in extensive developmental studies addressing the organization of the egg, the

presumptive fate of early cleaving blastomeres and the induction potential of different[155,199–203] and later on, also evo-devo studies[204]. Currently, several ESTs, transcriptomic and genomic resources are available[128,205]. Both, *B. belcheri* and *B. japonicum*, are used as a model organism to study evo-devo, innate and adaptive immunity.

Branchiostoma floridae

The earliest described evidence of this species was collected in Tampa Bay and several other locations in the Gulf of Mexico[206] (Fig. III.1 A). The natural spawning season of animals collected in Tampa Bay is during the months of May till September[174] (Fig. III.1 B). The first studies done using *B. floridae* focused in the karyotype characterization and in adult physiology, such as the circulatory and the excretory system[207–209]. By the late 1980's Nicholas and Linda Holland initiated work by characterizing the eggs of *B. floridae*[150] and since then they have pushed the model to be a reference in the field of evo-devo. *B. floridae* was the first amphioxus species to have its genome sequenced[31,123], highlighting even more the importance of cephalochordates to understand the evolution of chordates, now from a genomic perspective.

Asymmetron lucayanum

Also known as the Bahamas lancelet, where it was first described[210], is a rather cosmopolitan distributed species of amphioxus. *A. lucayanum* is distributed from the western Indian Ocean to the central Pacific Ocean, and the western Atlantic Ocean and it can be subdivided into several cryptic species[192] (Fig. III.1 A). The reproductive season for this species is April to June and mid-September till end of November, at least for the population found in the Bahamas[211] (Fig. III.1 B). Morphologically, *Asymmetron* and *Branchiostoma* species have several differences, with the most striking being the fact that *Asymmetron* has gonads only on the right side. In addition, the larvae of *A. lucayanum* have a pigmented tail fin, comparable to what is observed in *B. lanceolatum*, but not in *B. floridae*[179]. Surprisingly, it has been reported that *A. lucayanum* and *B. floridae* when crossed can produce viable hybrids, albeit they probably diverged in the Mesozoic Era, about 162 Mya[180]. Moreover, recently complete transcriptomes of the phylotypic stage of development and the adult form have been made available[130].

Capture and transport

Amphioxus can be captured in the wild by collecting sea floor substratum and followed carefully sieving to obtain adults. Several techniques are used to obtain substratum containing amphioxus, this variability depends mostly on the depth at which a given species is found. In general, three different methods are used to collect sandy substratum. The first one is mostly used when amphioxus are found at low depth, and it consists of shoveling sand and passing it through a thin mesh sieve (i.e. 1.25mm mesh), alternatively a geology sieve (i.e. 20cm diameter, 710µm mesh opening) can be used directly to collect and sieve the animals[174,179,212]. A second method consist of point specific bottom grabbing using a Van Venn-type (i.e. 300cm^2), Smith-McIntyre or a cylindrical dredge (i.e. 20 or 60cm in diameter and 40 or 90 cm in length) attached to a hydraulic winch able to lower and pull the device, usually in a boat[213–217]. The third method is more adapted to collecting the larger, rarer or more motile species. This method is based on benthonic trawling (dragging) and uses specifically adapted Charcot-Picard, Sanders or naturalist sediment dredges (i.e. 60cm x 20cm, mesh size 50µm). In these dredges the mouth of the net is held open by a solid metal beam, attached to two "shoes", which are solid metal plates, welded to the ends of the beam, which slide over and disturb the seabed when dragged by a boat[213,215,218]. For the two latter methods, the benthonic substratum obtained is then either sieved through or carefully verified by hand in order to isolate adult amphioxus[213,215].

Once isolated adult amphioxus can be transported small and long distances by simply paying attention to a few important points: temperature, ratio air/water and luminosity conditions. It is preferable though to ship or transport amphioxus adults with sand, since individuals will be less disturbed while burrowed in the sand. To limit stress induced by transportation it is very important not to overcrowd the animals and to keep the temperature stable and cool, by using a temperature-controlled cool box or by controlling externally the shipping temperature[148]. Moreover, it is important to have a balanced ratio air/water ratio and in the case of ripe animals, the light conditions are fundamental to avoid spontaneous spawning during the shipment, the use of transparent containers or artificial illumination is recommended[148,174]. For instance, for small distances simple transparent water bottles of 1,5L containing 3cm of sand and 750mL of sea water can be used to transport up to 10 individuals. Differently, for long distances a transparent plastic carboy of 20L capacity with 5-10cm high of sand in the bottom and with 10-12L of sea water under A/C controlled temperature can be used to transport up to 150 individuals. Or alternatively, up to 20 amphioxus can be

transported in 3L plastic sleeves, containing enough sand and at least 1L of sea water, well conditioned into a big Styrofoam box or temperature-controlled cool box[148,219].

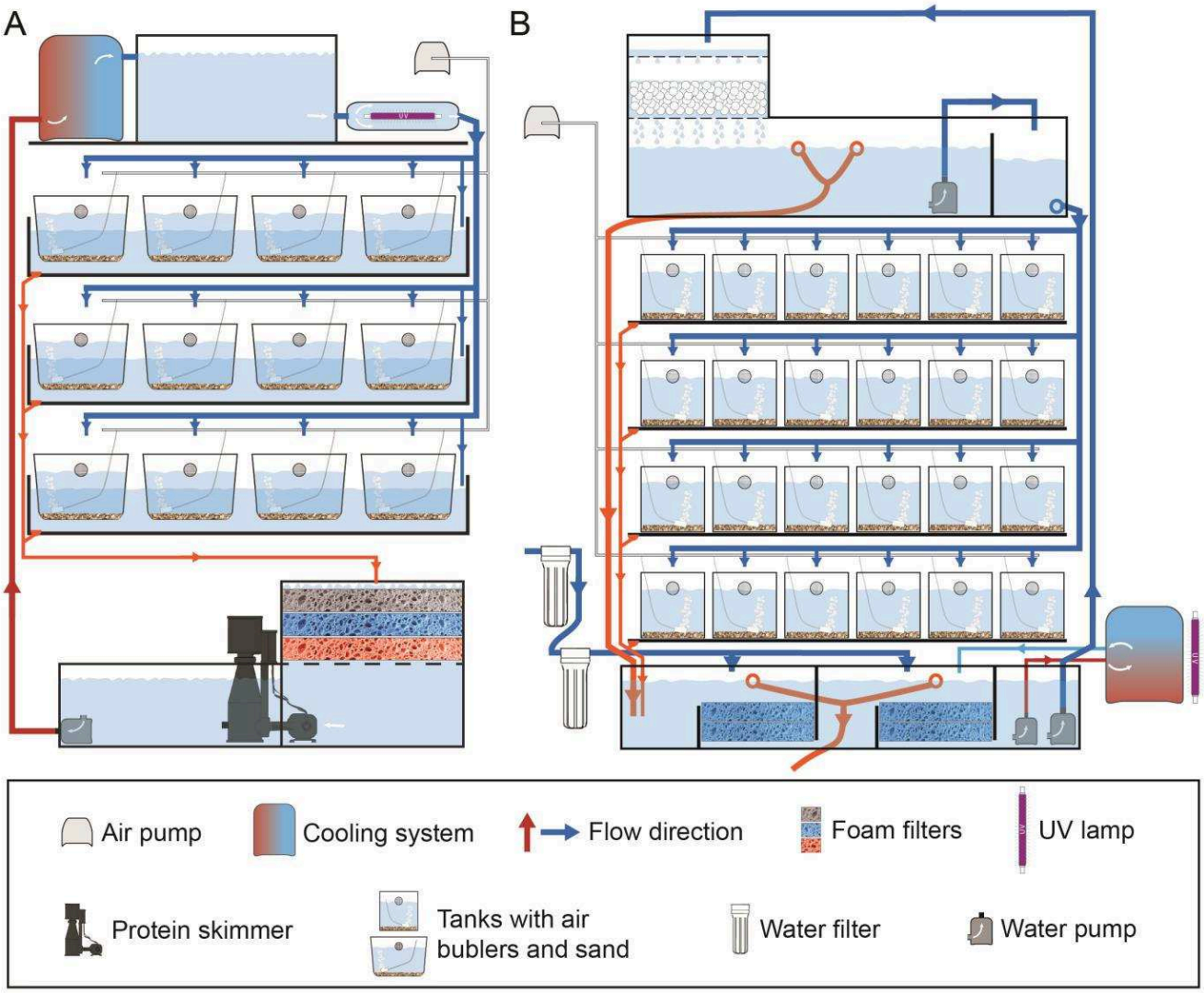


Fig. III.2 – Amphioxus main types of aquaculture systems. *(A) Schematics based on description by Benito-Gutiérrez ad collegues[219]. (B) Schematics based on our own amphioxus facility and described by Theodosiou and collaborators[148].*

Aquaculture

In order to implement amphioxus as an exploitable model of study, during the past few years a great effort has been made to generate tools to reproduce the complete life cycle in laboratory controlled conditions. The prospect of obtaining embryos on daily basis and avoiding the seasonal component of the natural reproduction of amphioxus impelled several labs to establish amphioxus aquaculture facilities. Aquaculture techniques to maintain adult livestock have been developed for the five species of amphioxus, described as model systems.

Many different modalities of amphioxus maintaining facilities, more or less advanced, using natural and artificial sea water both in marine stations and in inland laboratories have been described[148,179,196,198,213,219–225], with *B. lanceolatum* and *B. japonicum* being temperate species maintained in captivity at a range of temperature from 14ºC to 21ºC and with *B. floridae*, *B. belcheri* and *A. lucayanum* being sub-tropical species cultivated in the laboratory at temperatures above 24ºC[179,220,226]. Essentiality, two models of system to maintain amphioxus livestock have been described (Fig. III.2). One uses individualized tanks that are alimented also by a reservoir, but these individual tanks are placed into a water bath (similar to a *Xenopus* aquarium system) that is used to avoid temperature fluctuations induced by the room temperature changes[219,221] (Fig. III.2 A). The second one is based on the classical zebrafish aquarium systems, where a reservoir provides sea water to individualized tanks distributed in racks[148,222] (Fig. III.2 B).

At our lab, in Villefranche-sur-Mer, in 2013 we have built a custom made (Müller + Pfleger GmbH & Co. KG) improved version of the adapted zebrafish-like system previous described[148]. Our two aquarium system facility (maximum capacity of 48 aquariums of 5L) is enclosed in a room that has no source of natural light and where the room temperature is completely controlled (ranging from 10ºC to 20ºC). Furthermore, both systems can run under totally independent water temperature, since each system is attached to an independent water temperature regulator (TECO - TR30). To increase the modularity of the present system, each of the 4 rows of each system is equipped with fully controlled and independent sources of light (ProfiLux 3.1T attached to Profilux extension box 2, from GHL) and with adapted blinds to avoid light “contamination” into neighboring rows. The aquarium system itself consists of an upper, a lower reservoir and four rows that can host tanks (Fig. III.2 B). We use a semi-open configuration with constant natural, pre-filtered by two 25µm filters, water influx in the lower reservoir. Fresh sea water constantly arriving in the lower basin subsequently passes through four foam filters before being pumped up to the top reservoir which contains a trickle

filter system. Simultaneously, water in the lower basin is pumped through a temperature controller equipped with an internal UV light for sterilization. From the top reservoir, sea water is constantly distributed to the individual tanks, except during the "feeding" period (twice a day, during a period of 3h each) when the water flow is stopped. Each tank has air bubblers and a thin layer of sterilized sand. Sand is changed roughly once a month.

Gonad development and spawning

One of the striking problems of long term cultures of amphioxus lays in the ability to develop and maturate gonads in newly metamorphosed juveniles as well as in spawned adults. Due to the lack of ecological information about how this process happens in wild populations of amphioxus, a limited set of conditions can be tested in lab controlled conditions. Nevertheless, a few of them seem to be fundamental for this process. Accordingly to what was described for *B. lanceolatum*, by testing several types of alimentary regimes, only a very rich and constant feeding process can account for a complete gonad development that abuts into viable spawning[213,219]. Furthermore, studies using *B. belcheri* highlighted other factors that were proven critical for this process: the tidiness of the sea water and settlement substratum, as well as the low density of animals in each tank[220,225].

Another shortcoming of the reproduction of amphioxus is associated with the on-demand induction of spawning that relies in two key factors: periodicity of spawning and capacity to artificially induce spawning. Currently, *B. lanceolatum* can be induced to spawn on any given day during the breeding season, by simply applying a 36h long heat shock at 23ºC[148,196,213]. This technique can also be applied to *B. lanceolatum* that have been induced to mature out of phase with the natural occurring breeding season[219]. Nevertheless, continuous and repetitive spawning in *B. lanceolatum* has not been achieved under aquaculture conditions, meaning that each individual only spawns once or twice during their reproductive period[213]. Differently, for the other amphioxus species it seems to be more difficult to induce spawning in a given day. For instance, *B. belcheri* and *B. japonicum* can also be induced to spawn via a thermal shock, although the efficiency obtained is much lower than what is verified for *B. lanceolatum*[198]. *B. japonicum* seems to behave a lot like *B. lanceolatum*, since each individual spawns only once or twice and only during their reproductive season[198,223]. However, *B. belcheri* is able to spawn consecutively several times in year-round basis, although due to high genetic diversity of the population the time between individual spawnings can very from two to six months[198,220]. *B. floridae* was initially described as induced to spawn by an electrical shock of

10msec of 50V[174]. Technique that was abandoned and spawning induction is now done by elevating the temperature for 5ºC during 24-36h. This heat shock method seems to work, but is not highly efficient[180]. As well as in *B. belcheri*, *B. floridae* also spawns at intervals throughout the year, especially around the times of the new and full moon[180]. So far no data has been conclusive in reporting the capacity to develop gonads in the Bahamas lancelet, *A. lucayanum*[211]. Nevertheless, both field and captivity observations suggest that when kept at regular warm temperature these animals can spawn in regular basis throughout the year, having a strong correlation with the lunar cycle[180,211]

Observations produced from our animal facility during the last 4 years (2013, 2014, 2015 and 2016) are resumed in Table III.1. Of note, all the animals in our facility were collected every year in a single day by dredging in Argelès-sur-Mer, France, and retrieved from the sand by sieving, and thus each year's population has its own specificities. After collection and transport, the newly arrived animals were accommodated to the facility conditions during two weeks. During these two weeks the animals were kept into 45X28X18cm tanks with sand, aeration and with constant running water where they were fed daily. The summer-like light cycle of 14/10h (light/dark) was inverted upon arrival in the facility, i.e. nightfall at 8 A.M. Light cycle was slightly adjusted each spawning year, accordingly to the experimental times needed. After this period animals were checked individually, scored and deposit in 24x16x14,5 cm tanks with sand and aeration (Fig. III.2 B), between 10 to 20 animas per tanks. In total, during 79 spawning days we thermal shocked 694 individuals and we obtained spawning from 423 of them, thus we obtained a global spawning efficiency of about 60%. The distribution of the spawning efficiency in a given day for the 4 past *B. lanceolatum* spawning seasons is shown in detail in Fig. III.3. The peak of spawning efficiency was obtained in mid-June during the season of 2013, mid-June to beginning of July during 2014 and from mid-June to mid-August in 2015 (Fig. III.3). Differently, in 2016 the spawning efficiency was rather variable and mostly low when compared to the results obtained in other seasons (Fig. III.3). For the seasons from 2013 to 2015 we estimate that the progressive increase in the capacity and span to obtain spawning (from 58,29% to 72,10% and a single day with 100% efficiency to 9days) is associated with the gradual and fine-tune adjustment of the facility conditions over the years (light, temperature and feeding conditions), since 2013 was the first year we were actually using this kind of aquarium system. Surprisingly, the 2016 results were far from ideal and expected in the light of the experience that was acquired in the previous years and considering that the photoperiod and temperature regimes were similar

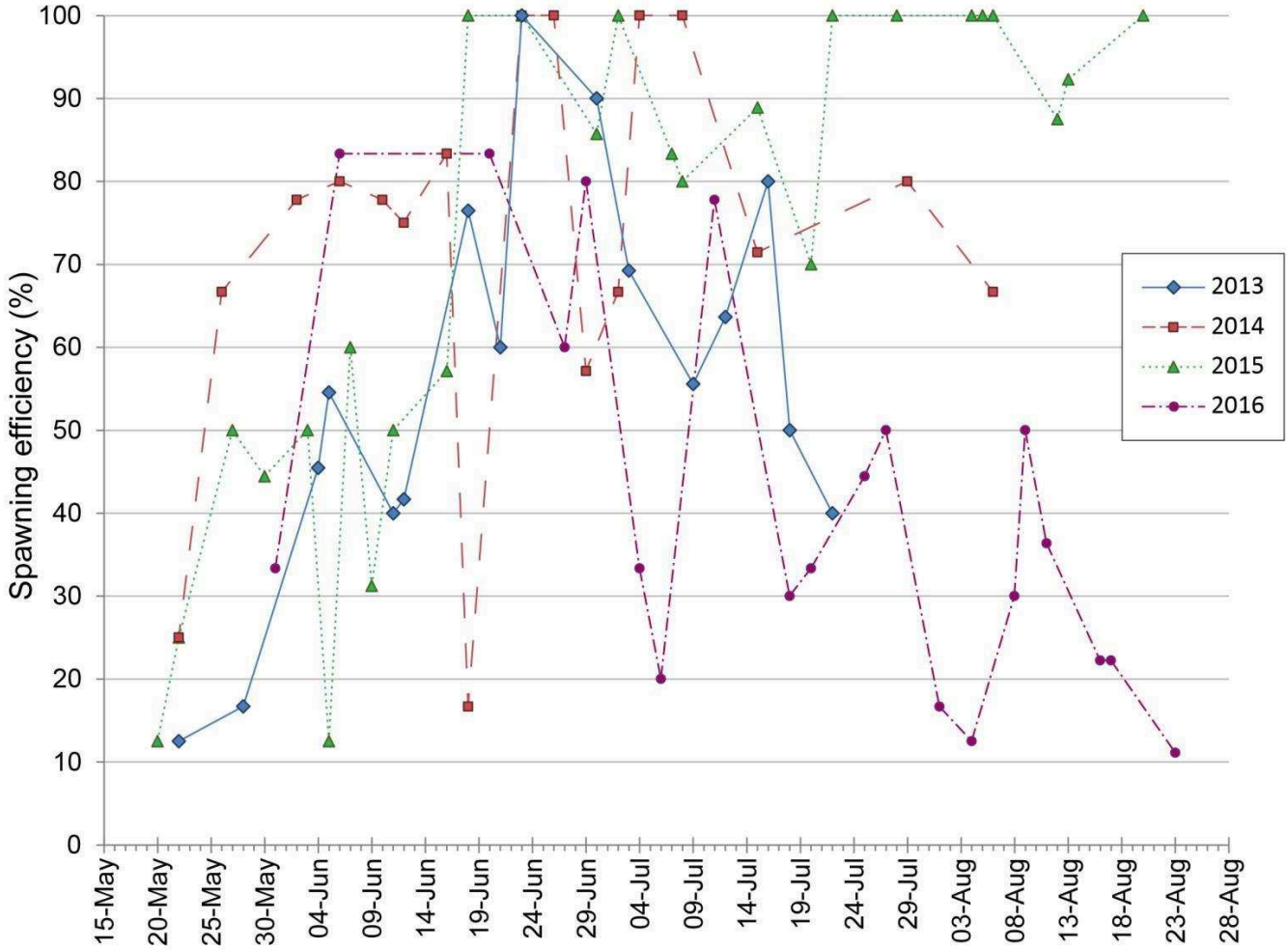


Fig. III.3 – Spawning efficiency in the different spawning days of the 2013, 2014, 2015 and 2016 seasons.

(Fig. III.4, May-July 2015 compared to May-July 2016). For instance, in August 2016, to try to increase the spawning capacity and induce a late maturation of the animals we specifically increased the overall temperature of about one degree when compared to what have been measured during the previous season (Fig. III.4). We, thus, interpreted our disappointing results of the 2016 season (38,96% of spawning efficiency and not a single day where all the selected animals spawned) as the result of a population collected that was mostly constituted by very small and immature individuals. Nevertheless, this is a speculative idea and the message that should be retained is that despite our ability in controlling amphioxus maintenance and reproduction, we are still far from fully understand the ecological cues that define the amphioxus life cycle.

Season:	2013	2014	2015	2016
Number of spawning days	16	17	26	20
Number of shocked animals	175	132	233	154
Number of animals that spawned	102	93	168	60
Global efficiency of spawning (%)	58,29	70,45	72,10	38,96
Ripe animals from previous years (%)	N/A	N/A	27,00	30,37
Efficiency of spawning of ripe animals from previous years (%)	N/A	N/A	32,43	23,33

Table III.1 – Overall results for the amphioxus season of 2013, 2014, 2015 and 2016 in our lab at Villefranche-sur-Mer.

In order to have our stabilized adult cultures developing gonads and being able to spawn, we developed a temperature based method where we simulated environmental conditions experienced in the wild (Fig. III.4). Starting form September 2014, we applied a slowly decreasing temperature until the month of December (to simulate a winter period) and we gradually increased the temperature until it reaches a summer like temperature by the end of April. The temperature changes were induced by modifications at the same time of the air temperature and by regulating water temperature at the exit of the cooling systems. Along with the temperature changes we also slowly shifted the photoperiod, form a regime with 14/10h of light/dark during the spawning season to 10/14h of light/dark during the month of December and return to the initial light cycle by April (as represented for an homologous period in Fig. III.4). During this period we fed our animals with a mixture of brown and green algae supplemented with artificial complements once a day. With this approach we obtained a total of 71 animals from previous years (27% of the adults in captivity at that time) that were able to recover gonads, nevertheless the efficiency of spawning of these animals only reached about 30% (Table III.1). Following this initial test, we applied more or less the same temperature conditions from September 2015 to April 2016 (Fig. III.4) but this time we fed the animals with the same diet twice a day. These resulted in the gonadal development of 116 animals in our existing population (~30% of the animals) and a spawning efficiency of about 22% (Table III.1). These results suggest that *B. lanceolatum* is able to develop viable gonads in captivity[219], but for the moment we do not completely master the ideal cocktail of conditions to make this process happens efficiently.

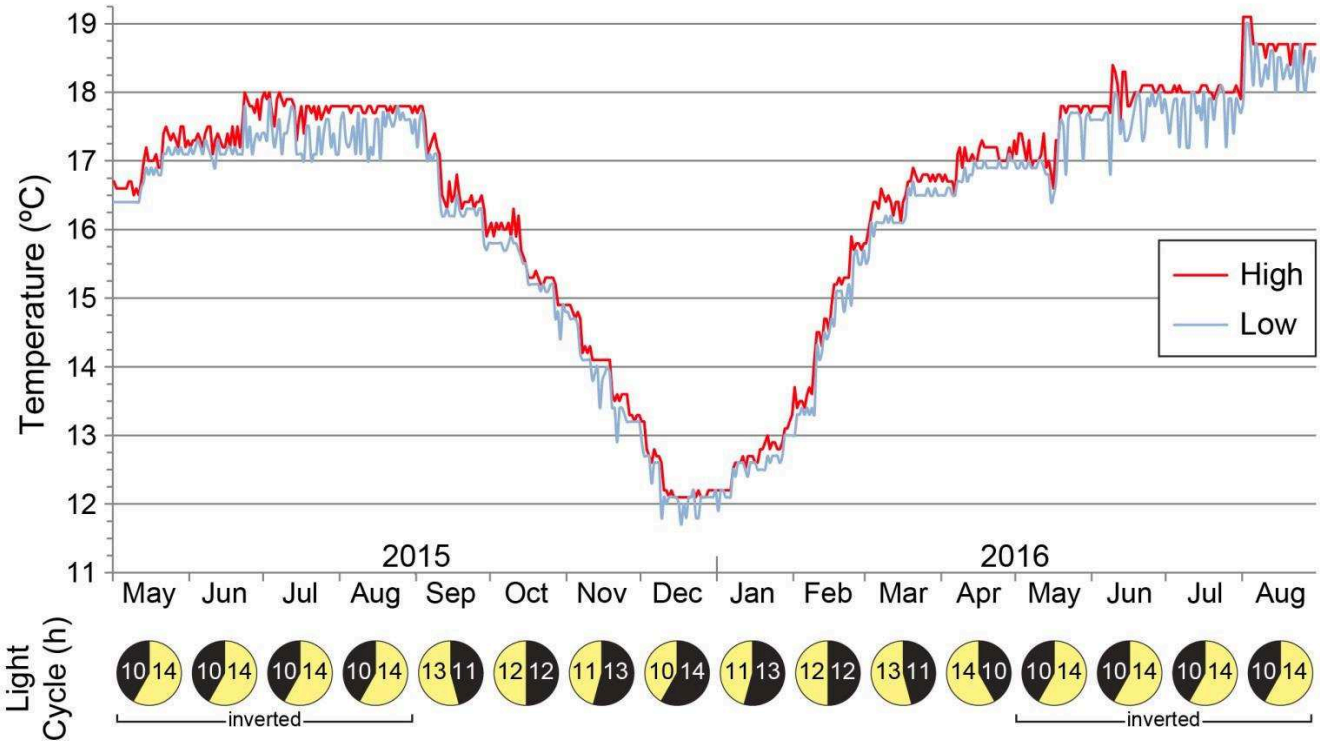


Fig. III.4 – Temperature measurements and artificial light cycle in the amphioxus facility. *Three measurements were made every day from the beginning of the 2015 spawning season (May 2015) till the end of the 2016 spawning season (August 2016), representing a direct comparison of the conditions between two distinct spawning seasons and an example of the year round conditions in our aquaculture system. Higher and lower temperatures for a given day are plotted. Light cycle for a given month is represented by a circle with yellow (14-10 h of light) and black (10-14h of dark) sides. Period when the facility runs with an inverted light cycle are indicated.*

Embryo and larval cultures

Fertilization can be achieved once eggs and sperm are obtained either after spawning induction or by spontaneous spawning. By simply adding a few drops of concentrated sperm to several thousands eggs in petri dishes. Within the first minutes after adding the sperm, the fertilization envelope starts rising[150]. Embryos can be raised easily in petri dishes (i.e. 15cm diameter) without specific aeration conditions. It is essential to avoid overcrowding the embryos at all times, thus is important to consider that as development proceed the size of the embryos tends to increase. The development continues with several rounds of cleavages that originate a blastula that develops into a gastrula[152,154,176]. At the onset of the neurulation

process the embryos start spinning inside the fertilization envelope by the action of uniformly distributed cilia. After hatching, embryos swim up to the surface of the water[174]. Hatched neurulae keep elongating and the pharynx regions slowly become enlarged. The next phase in development is the opening of the larval mouth and 1st gill slit, followed by the formation of the anus[162,212]. At this stage the larva should start to be fed. The larval development proceeds by posterior elongation of the larva and the addition of gill slits posteriorly to the 1st one[162,212]. The onset of metamorphosis is variable depending on the amphioxus species, for instance *B. belcheri* start metamorphosis when larvae exhibit 16-18 primary gill slits[173]. Differently, in *B. floridae* the second row of gill slits (indicator of the metamorphic process) starts to appear when larvae exhibit 9 to 10 gill slits[174] and at least 14 gill slits for *B. lanceolatum*[175].

The bottleneck for obtaining filial generations of amphioxus is the limited capacity of maintaining larvae that go through metamorphosis and of juvenile cultures able to reach sexual maturity. For instance, in *B. lanceolatum*, although embryos can be cultured through metamorphosis and until the adult stage in captivity[84,227], a complete life cycle has not been achieved in the laboratory. By contrast, long-term cultures of several generations of *B. belcheri* and more recently of *B. japonicum* have been developed[222,223,225]. Moreover, when maintained in natural sea water with sand, *B. belcheri* cultures have been successfully raised in phase with the natural life cycle observed in the wild through two generations in the laboratory[225].

To date three different approaches to obtain amphioxus juveniles were described in detail: larvae were raised either in large petri-dishes[174,212] (Fig. III.5 A), or in large tanks with[228] (Fig. III.5 B) or without sand at the bottom[173,222,225] (Fig. III.5 C-D).

After starting to feed *B. floridae* larvae ware grown in Petri dishes (14cm diameter) and the sea water was partially changed every day by transferring the embryos into a new clean dish. This method is ideally used to raise a few dozens of larvae per dish and juveniles obtained were not developed until reaching sexual maturity[174,212].

Massive cultures of 3 days-old *B. japonicum* larvae were maintained and fed in 94x94x70cm tanks with a 5cm layer of sand at the bottom and half of the total water was changed daily by the action of semi-automatic flow. The survival rate of this method before metamorphosis was of 5,5% and only 0,7% of the initial larvae reached sexual maturity[228]. Of note, amphioxus larvae that stopped developing at the 3 gill slits stage died within few days, suggesting this stage as a critical importance into achieving a metamorphic state[228].

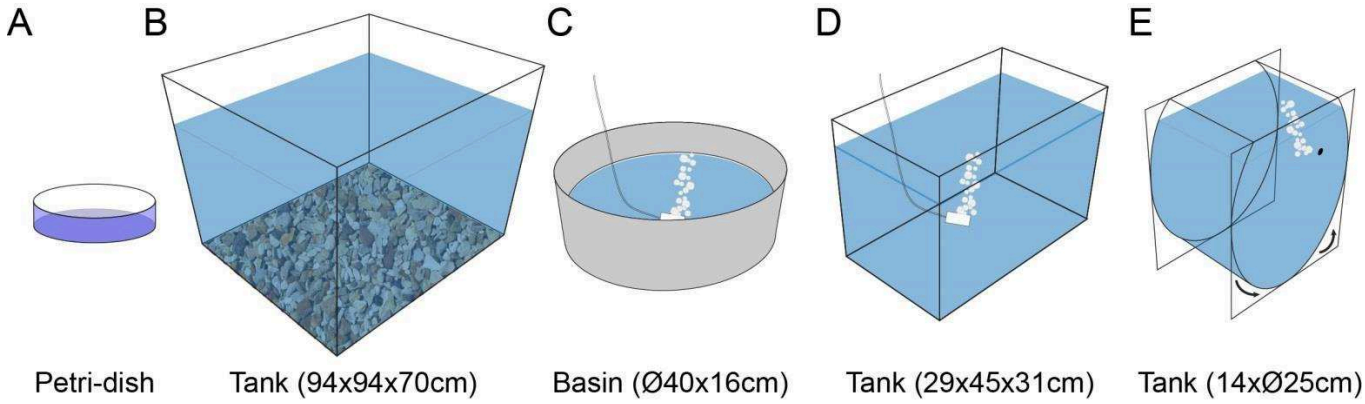


Fig. III.5 – Amphioxus larvae culture systems. *(A) in petri-dished, water changed every day; (B) tank with sand in the bottom, half of the total water changed daily; (C) basin with continuous aeration, water changed every day, after 30 days of culture 1cm thick layer of sand was deposit; (D) tank with continuous aeration, water changed weekly; and (E) cylinder tank with continuous aeration that creates an internal circular flux (indicated by the curved arrows), 2/3 of water changed daily.*

Differently, *B. belcheri* and *B. japonicum* hatched neurulae ware transferred into round opaque basins (40cm in diameter and 16cm in depth) with continuous aeration without sand. Just before feeding the seawater was partly replaced with fresh seawater by filtration. Dead algae and feces accumulated at the bottom were purged weekly, by pouring the upper clean water together with most larvae into another basin and carefully recovering the reaming larvae in the bottom water. After 30 days of culture, a 1cm layer of sand was deposited to allow larval settlement and thus metamorphosis. After metamorphosis the juveniles were cultivated using the same conditions as adults thus achieving with this method a survival rate of about 3% for both amphioxus species[225].

Alternatively, hatched *B. belcheri* embryos were raised in aerated 33L glass tanks (29x45x31cm containing 30L of seawater) without sand in the bottom. Cultures were fed daily and water was changed weekly with a silicon-tube syphon passing through an adequate nylon mesh. Fecal pellets, fouling algae, and other debris were removed at this stage by pipetting under a dissecting scope. Although the authors do not mention exactly the survival rate achieved by this method, they mention that it is "not high"[173,222].

In our lab we implemented an amphioxus larvae culture module that we believe will increase significatively the overall survival rate with minimal time investment. By re-using a *Clytia hemisphaerica* (a cnidarian) nursery tank (https://clytia.wikispaces.com/Nursery+tank),

we were able to maintain a large population of amphioxus larvae during 40 days, unfortunately the culture died due to a fungal contamination present in our algae supply. The system in cause is an acrylic cylinder (25cm of diameter and 14cm height) cut on top of the rounded side and closed on both sides by two squared acrylic plates that function as support base. The aquarium has a maximal capacity of approximately 6L (5,8L of sea water while running). The cylinder wall is perforated by a little tube that serves as air supplier/bubbler when connected to an air pump (Fig. III.5 E). When set at the right pace the air bubbles create a circular current that allow active larval movements and, avoids larval settlement and air-water interphase blockage[179]. In our preliminary test, we fed the larvae with live algae twice a day and replaced 2/3 of the total volume with fresh sea water every day. Old water was removed by aspiration using a 50µm nylon mesh as filter. Once a week the aquarium walls were cleaned with running fresh water and major algae debris were removed.

Feeding

In general amphioxus has a huge capacity to filter and ingest a large spectrum of particles sizes (from 0,062 to 100 µm), suggesting that their diet can include microbial as well as planktonic food. Due to this ability amphioxus may play a fundamental trophic role in coastal food webs[229]. So far, only a single study clearly aimed to characterize the natural diet of amphioxus and it suggests that amphioxus can selectively consume a variety of prey and be regarded as a typical omnivore. Its diet can range from microbes, bacteria, microalgae, marine fungi, diatoms, dinoflagellates, zooplankton to detritus. In Hong Kong waters, the amphioxus diet is based mostly in dinoflagellates, zooplankton, detritus and bacteria, with minor contribution from diatoms and fungi[230]. The analysis of the amphioxus feces further support the importance of the bacteria in the diet of these animals[231]. Although, in most labs cultivating amphioxus the feeding regimes for larvae, juveniles or adult are based in a mix of brown (i.e. *Isochrysis* sp. and *Pavlova* sp.) and green algae (i.e. *Dunaliella* sp. and *Tetraselmis* sp.) and eventually supplemented by some diatoms species (i.e. *Thalassiosira* sp. and *Chaetoceros* sp.)[122,180,220,222,223,227]. To complement the lack of certain essential nutrients in diets based in phytoplankton some labs are also using artificial supplements[148,213,219].

Adult amphioxus can be maintained alive even without being fed, nevertheless that might compromise their capacity of becoming ripe and spawn[213,220,221]. In general, amphioxus adults in aquaculture should be supplied with food once or twice a day in large quantity that they develop gonads and become fully ripe[148,198,219]. In the case of amphioxus larvae virtually

nothing is known about their dietary needs and behavior in the wild. Even though, it is possible to maintain and grow amphioxus larvae into sexual mature juveniles, with very low survival rates, by abundantly feeding them with a mix of diatoms, green and/or brown algae [173,225,228]. Nevertheless, this kind of diet is not always effective, like in the case *A. lucayanum* larvae development[179]. It would certainly be very interesting to carry out additional studies to understand the dietary and ecological needs of both developing larvae and adults.

Conclusion and perspectives in amphioxus as a model

Amphioxus represents one of the key models to understand evolutionary diversification at the base of the chordates and the establishment of conditions to obtain ripe animals in captivity during and outside the regular spawning season, the capacity to induce spawning and the ability to replicate the whole life cycle under aquaculture conditions thus represent major improvements for amphioxus as a laboratory animal system. Although some of the described methods still show some limitatations, they have nonetheless already allowed a significant extension of the experimental availability of amphioxus, compared to what would have been possible without amphioxus animal husbandry protocols. Furthermore, husbandry methods are currently being developed for different amphioxus species, including members of both the genus *Branchiostoma* and *Asymmetron*, and will thus ultimately allow access to a wide variety of different cephalochordate species, hence allowing comparative studies within the cephalochordates to reveal similiarities and differences of the members of this phylum. Nevertheless, to further improve amphioxus as a model system, a better understanding will be required of the dietary and environmental factors that control, in the wild, larval development, pre-metamorphic settlement, juvenile maturation, spawning and gonad development. This calls for a renewed interest in the ecology of cephalochordates in their natural habitats.

Chapter IV: Retinoic acid signaling[1]

A large amount of data collected during the past 100 years has established that vitamin A (also known as retinol) and its derivatives, the retinoids (Fig. IV.1) play crucial roles during vertebrate development. The retinoids are fat-soluble morphogens that are involved in a variety of processes, such as pattern formation during early development as well as in organogenesis, cell proliferation, cell differentiation, apoptosis, vision, immune responses and tissue homeostasis[233–242]. Initial insights on retinol-dependent developmental processes were obtained from nutritional excess and deficiency studies carried out in different animal models including pigs and rats[243,244]. In particular, the so-called vitamin A deficiency (VAD) models produced valuable information about the biological roles of retinol in different organs during embryonic development[245]. The subsequent emergence of genetic and molecular tools allowed a much more detailed dissection of the developmental roles of retinol and its derivatives including the identification of the molecular machinery underlying retinoid-dependent signaling[237,246,247].

It is now generally accepted that all-*trans* retinoic acid (RA) is the main biologically active form of retinol, although a biological role for other retinoids, such as 9-*cis* RA, is still a matter of debate[239,248,249] (Fig. IV.1). Moreover, although alternative mechanisms for triggering retinoid-dependent signaling have recently been proposed[239], in vertebrates the main cellular mechanism of RA action occurs by binding of RA to heterodimers of two nuclear receptors: the retinoic acid receptor (RAR) and the retinoid X receptor (RXR)[237]. This classical mode of retinoid action involves fixation of RA by the DNA-bound RAR/RXR heterodimer, followed by the recruitment of co-activators and ultimately the transcriptional activation of target genes[250] (Fig. IV.2)

[1] Adapted and updated from the book chapter published in 2013[232]. Authorship: João E. Carvalho and Michael Schubert.

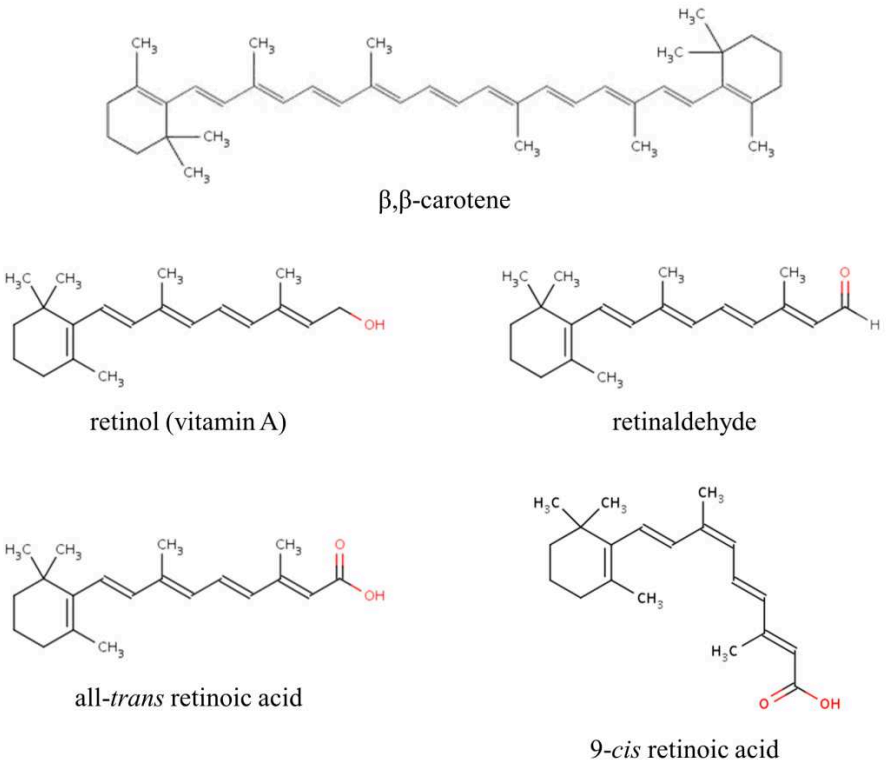


Fig. IV.1 – Chemical structures of different retinoids.

Retinoid metabolism

Due to their inability to synthesize retinol *de novo*, animals need to obtain retinoids through the diet, mainly in the form of carotenoids, such as *β,β*-carotene (Fig. IV.1). Carotenoids are isoprenoids that are widespread in nature and are typically associated with the yellow, orange, red or purple colors found in vegetables, fruits, flowers, birds and crustacean[251]. Alternatively, retinol can also be taken up directly in the form of pro-vitamin A carotenoids[235]. The conversion of ingested carotenoids, such as *β,β*-carotene, into retinaldehyde is the first step of the classical retinoid metabolism (Fig. IV.2). This process is driven by the enzyme *β,β*-carotene-15,15′-monooxygenase 1 (BCO-I)[252]. This enzyme is crucial for proper embryonic development, at least on diets with *β,β*-carotene as the only source of retinol. For example, *BCO-I* knockout mice are accumulate high levels of *β,β*-carotene[253], a condition that is similar in zebrafish embryos[254]. Additional to BCO-I, *β,β*-carotene-9,10-dioxygenase 2 (BCO-II) has been described as a second carotenoid oxygenase

in vertebrates[255]. BCO-II mainly functions in regulating cell survival, proliferation and preventing oxidative stress in accordance with its mitochondrial location[256]. Additionally, BCO-II might also function as the key enzyme for a non-canonical pathway of RA production[257], but this involvement of BCO-II in retinol metabolism remains to be confirmed *in vivo*[258,259].

Retinaldehyde is subsequently hydrolyzed to retinol, which can be used for two different purposes: esterification and tissue storage or oxidative metabolism leading to synthesis of RA, the main biologically active retinoid. The process of esterification is conducted by the enzyme lecithin:retinol acyltransferase (LRAT), which produces retinyl esters (Fig. IV.2). Under retinoid oversupply, LRAT has been proposed to play a crucial role in maintaining retinol homeostasis[260]. For RA synthesis, retinol is transported by retinol binding protein (RBP) to target tissues[261,262]. RBP is a specific transport protein circulating in the blood and associated one-to-one with another serum protein, transthyretin (TTR)[262–264] (Fig. IV.2).

In mice, RBP, that cannot cross the placenta[265], has been detected at the junction of the uterine wall and the placenta and at E13.5 (embryonic day 13.5) and during subsequent development in the visceral endoderm of the yolk sac[258]. Similarly, RBP has been detected in developing chick and zebrafish embryos[266,267]. Given that RBP is the only known specific carrier protein of retinol, it is rather surprising that mice lacking RBP ($RBP^{-/-}$) are viable showing no major malformations when exposed to a retinol-sufficient diet[268]. Moreover, using an elegant genetic approach, it was shown that, in a $BCO\text{-}I^{-/-}/RBP^{-/-}$ background, $BCO\text{-}I^{+/-}/RBP^{-/-}$ offspring can develop normally and maintain retinoid homeostasis, when $BCO\text{-}I^{-/-}/RBP^{-/-}$ pregnant mice are maintained under a VAD diet and *β,β*-carotene being the only source of retinoids, is administrated via intra peritoneal injection[269]. These results also suggest that *β,β*-carotene alone can be an adequate source of maternal retinoids[269].

The cellular uptake of retinol is carried out by a RBP membrane receptor, called stimulated by retinoic acid gene 6 (STRA6)[262,270] (Fig. IV.2). STRA6 works as bidirectional retinol transporter and its activity is controlled by intracellular retinoid levels[261,271]. Interestingly, STRA6 does not seem to function like any other known membrane receptors, transporters or channels[272]. Experimental evidence pointed to the existence of a pore or "passageway", made by STRA6's transmembrane domains, through which vitamin A is transported[273]. The mechanism used by STRA6 ensures that target cells only take up vitamin A when they have means to store it, and it also prevents excessive uptake and random diffusion of vitamin A[271].

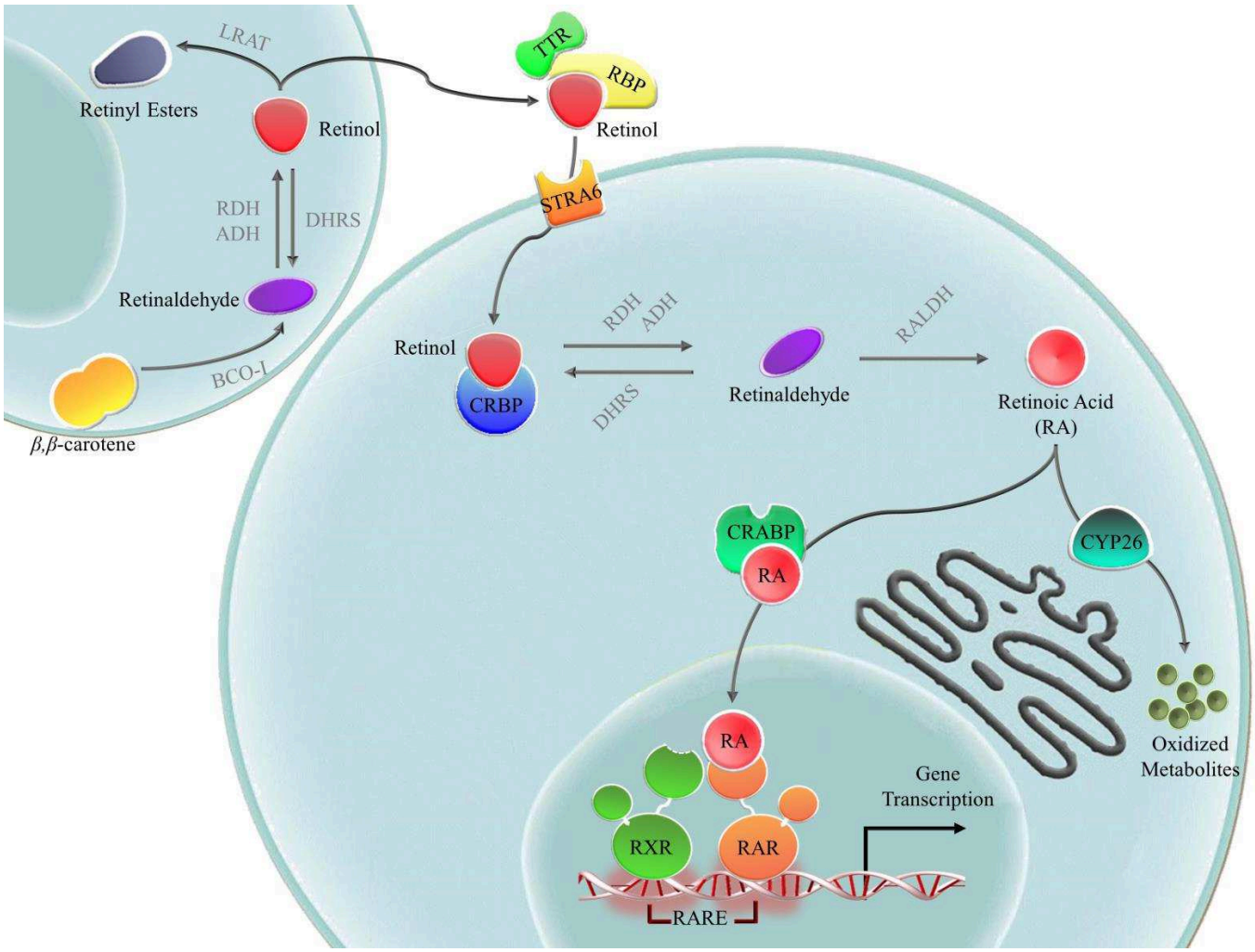


Fig. IV.2 – Retinoid metabolism. *Conversion of* β,β*-carotene to retinol (vitamin A), synthesis of retinoic acid, the major biologically active metabolite, and activation of retinoid-dependent signaling. ADH, alcohol dehydrogenase; BCO-I,* β,β*-carotene-15,15'-monooxygenase 1; CRABP, cellular retinoic acid binding protein; CRBP, cellular retinol binding protein; CYP26, cytochrome P450 family 26; DHRS, dehydrogenase/reductase SDR family member; LRAT, lecithin:retinol acyltransferase; RA, retinoic acid; RALDH, retinaldehyde dehydrogenase; RAR, retinoic acid receptor; RARE, retinoic acid response element; RBP, retinol binding protein; RDH, retinol dehydrogenase; RXR, retinoid X receptor; STRA6, stimulated by retinoic acid gene 6; TTR, transthyretin.*

Once inside the cell, RA is synthesized from retinol in two oxidation steps. The first step consists of the reversible oxidation of retinol into retinaldehyde, which is carried out by enzymes of two different classes: the cytosolic alcohol dehydrogenases (ADHs), belonging to the medium-chain dehydrogenase/reductase family, and the microsomal retinol dehydrogenases (RDHs), which are members of the short-chain dehydrogenase/reductase family[274] (Fig. IV.2). Several members of the ADH family, including ADH1, ADH3 and ADH4, have been shown to catalyze the oxidation of retinol into retinaldehyde *in vitro*[238,274]. Mice carrying null mutations for ADH1, ADH3 and ADH4 have revealed partially redundant functions of these three ADHs[274]. Moreover, while both ADH1 and ADH4 seem to be required specifically in conditions of retinoid deficiency or excess, ADH3 can probably be regarded as a ubiquitous ADH[239]. Like for the ADHs, multiple RDHs, such as RDH10, have been shown to be capable of synthesizing retinaldehyde from retinol *in vitro*[274]. However, recent experimental evidence suggests that the most important player in retinaldehyde synthesis during embryonic development is RDH10[275]. Loss of RDH10 functions in mice is lethal between E10.5 and E14.5 with embryos showing severe defects in embryonic patterning and morphogenesis[276]. It has recently been suggested that the regulation of endogenous retinaldehyde levels represents an additional mechanism for controlling RA signaling activity in developing embryos. Within a given cell, RA production is thus dependent on the balance between retinaldehyde synthesis from retinol and retinaldehyde reduction to retinol. While in developing embryos the former reaction, as mentioned above, is catalyzed by ADH or RDH10, the latter reaction can be carried out, for example, by DHRS3a (dehydrogenase/reductase SDR family member 3a), which, like RDH10, is a member of the short-chain dehydrogenase/reductase family[277].

The second oxidation step is the conversion of retinaldehyde into RA, an irreversible reaction carried out by retinaldehyde dehydrogenases (RALDHs) (Fig. IV.2). Three RALDHs of the ALDH1 class (RALDH1, 2 and 3 also called ALDH1A1, ALDH1A2 and ALDH1A3, respectively) and one RALDH of the ALDH8 class (RALDH4) have been implicated in endogenous RA production with each enzyme presenting distinct tissue-specific distribution[239]. RALDH2 is the earliest RALDH enzyme to be expressed during development and its expression is crucial for early embryogenesis, since mice lacking RALDH2 exhibit severe abnormalities and die before mid-gestation[278]. These phenotypes can be almost completely rescued by maternal administration of RA, suggesting that RALDH2 is indeed required for endogenous RA synthesis during development[278,279]. RALDH2 homologs in the frog *Xenopus laevis* and in zebrafish have been shown to have similar roles during early

embryonic development[280,281]. RALDH1, RALDH3 and RALDH4 are expressed at later developmental stages, with RALDH1 and RALDH3 contributing to the patterning of the respiratory and visual systems[282–284] and RALDH4 being detectable in the fetal liver[284,285].

Some reports have suggested that RA synthesis during embryonic development may also occur independently of RALDHs, possibly through the action of members of the Cytochrome P450 family of mono-oxygenases[234]. Functional studies point to CYP1B1 as a potential candidate[286,287], because this enzyme can efficiently oxidize retinol to retinaldehyde and subsequently to RA and exhibits an expression pattern consistent with RA synthesis patterns in the developing embryo[286].

Catabolism of endogenous RA is driven by enzymes of the Cytochrome P450 subfamily 26 (CYP26A1, CYP26B1 and CYP26C1) that are responsible for the oxidation of RA into metabolites (Fig. IV.2), such as 4-hydroxy-RA, 4-oxo-RA or 18-hydroxy-RA[288]. All three CYP26 enzymes are involved in embryonic development exhibiting dynamic expression patterns both in time and space[289,290]. *CYP26A1* expression starts during the gastrula stage in mouse endoderm, in chick ectoderm and mesoderm (close to Hensen's node) and in zebrafish anterior ectoderm. In frogs, expression of *CYP26A1* is located in anterior ectoderm during early gastrulation[291]. Mice devoid of CYP26A1 die before the first postnatal day presenting a posteriorized hindbrain and vertebral column as well as severe caudal truncations[292]. In contrast, CYP26B1 seems to be fundamentally involved in limb patterning and outgrowth[293] and required for maintenance of germ cells during testis development[294]. Like *CYP26A1*, *CYP26C1* is expressed in the hindbrain, but mice deficient for *CYP26C1* are viable and do not show severe anatomical abnormalities, suggesting some degree of functional redundancy between CYP26A1 and CYP26C1[295]. Double mutants for *CYP26A1* and *CYP26C1* have a much stronger phenotype, when compared to both *CYP26A1* and *CYP26C1* single mutants. Simultaneous loss of RALDH2 partially rescues these phenotypes, confirming that the malformations observed after loss of both CYP26A1 and CYP26C1 are due to elevated levels of RA[295].

Retinoids present within a given cell are bound by intracellular binding proteins, such as cellular retinol binding proteins (CRBPs) or cellular retinoic acid binding proteins (CRABPs) (Fig. IV.2), all of which exhibit extremely high affinities for their substrates[271,296]. CRBP-I, for example, protects retinol from oxidation and isomerization by limiting the interaction with enzymes capable of recognizing retinol[297]. During development, *CRBP-I* is expressed in several tissues, including motor neurons, spinal cord, lung, liver and placenta[298].

Mice lacking CRBP-I do not exhibit any morphological abnormalities, suggesting that RA metabolism is not dependent on CRBP-I function[298]. Nonetheless, in the absence of CRBP-I retinol stocks are severely depleted[299]. Compared to *CRBP-I*, *CRBP-II* has a more restricted expression pattern, being mainly localized in fetal intestine and liver, but also in the yolk sac from E10.5 till 15.5[296]. Its expression suggests a role as binding protein for newly absorbed retinol[300]. Interestingly, developmental expression of *CRBP-I* overlaps significantly with that of *CRABP-II*, while the patterns of *CRBP-I* and *CRABP-I* are mutually exclusive[301]. Several functions have been proposed for the CRABP proteins, but their antagonistic effect on RA transport seems to be the most important: while CRABP-I seems to mediate CYP26-dependent RA degradation[301,302], CRABP-II is probably responsible for the transport of RA to its receptors in the nucleus[303]. A fact supported by the increase of RA signaling when CRABP-II is overexpressed in cell lines and *Xenopus* embryos[296,304]. Unfortunately, *CRABP-I*$^{-/-}$ and *CRABP-II*$^{-/-}$ single mutant as well as *CRABP-I*$^{-/-}$/*CRABP-II*$^{-/-}$ double mutant mice failed to reveal the biological functions of CRABPs[297]. Although, a recent analysis in zebrafish has highlighted the importance of CRABPs in stabilizing the RA morphogen gradient in the developing hindbrain by delivering RA both to its receptors and to CYP26 for degradation[305].

Classical retinoid signal transduction: the RAR/RXR heterodimer

Until very recently, the main dogma of retinoid research stated that RA functions are mediated by heterodimers of two nuclear receptors: RAR and RXR[250,306]. There are usually three RAR genes encoded in mammalian genomes (*RARα*, *RARβ* and *RARγ*)[307], while zebrafish, for example, have lost *RARβ*, but duplicated both *RARα* (yielding *RARαa* and *RARαb*) and *RARγ* (yielding *RARγa* and *RARγb*)[308]. Congruent with the overall RAR toolkit, mammalian genomes generally also possess three RXR genes (*RXRα*, *RXRβ* and *RXRγ*). In contrast, the zebrafish genome, for instance, encodes six RXR genes (*RXRαa*, *RXRαb*, *RXRβa*, *RXRβb*, *RXRγa* and *RXRγb*)[308]. The RAR/RXR heterodimer acts as ligand-dependent transcription factor through binding of RA response elements (RAREs) located within the regulatory DNA region of RA target genes. *In vitro* studies have shown that RAR is able to bind to and to be activated by both all-*trans* RA and 9-*cis* RA, whereas RXR binds and is activated only by 9-*cis* RA. However, it remains to be determined, whether 9-*cis* RA is

present in vertebrate embryos and if its interaction with RXR effectively triggers a biological process[309].

Regarding the DNA binding properties of the heterodimer, RAR/RXR typically associates with two DNA sequence stretches composed of direct repeats (DRs) with the conserved nucleotide sequence (A/G)G(G/T)(G/T)(G/C)A. The spacing between the two DRs is typically one, two, five or eight nucleotides long[250,310–312]. According to the canonical model, RAR/RXR heterodimers associate with DNA in the absence of ligand, thereby recruiting a co-repressor complex that mediates chromatin compaction and target gene repression[313]. In the presence of ligand, due to a conformational change of the receptor induced by the ligand, there is dissociation of the co-repressor complex and recruitment of a co-activator complex, which mediates chromatin decondensation and assembly of the transcriptional pre-initiation complex ultimately leading to target gene transcription[283].

Using genetic studies, it has been established that the three mammalian RAR paralogs have both specific and partially redundant functions during embryogenesis[314]. Interestingly, *RARα*$^{-/-}$, *RARβ*$^{-/-}$ and *RARγ*$^{-/-}$ single mutant mice are able to reach adulthood, presenting only minor morphological phenotypes indicative of VAD syndromes[315]. In contrast, *RARα*$^{-/-}$*/RARβ*$^{-/-}$, *RARα*$^{-/-}$*/RARγ*$^{-/-}$ and *RARβ*$^{-/-}$*/RARγ*$^{-/-}$ double mutant mice die during embryonic development exhibiting severe malformations typical of VAD[315]. *RXRα*$^{-/-}$ mice present several severe abnormalities and die during development, with most of the abnormalities equally found in the different *RAR*$^{-/-}$ compound mutants[315]. In contrast, *RXRβ*$^{-/-}$*/RXRγ*$^{-/-}$ double mutant mice do not show any evident morphogenetic defects, indicating that RXRα is the most important RXR implicated in morphogenetic patterning[316]. *RAR*$^{-/-}$ and *RXRα*$^{-/-}$ mutations seem to act in synergy, since null mutations for individual RAR paralogs combined with null mutations of RXRα reveal potentially redundant functions of the different RAR paralogs. Moreover, these compound mutants also provide very strong support for the fundamental functions of the RARα/RXRα, RARβ/RXRα and RARγ/RXRα heterodimers in transducing the RA signal during embryogenesis[317,318].

Non-classical retinoid signaling transduction

In addition to this classical mode of retinoid action, mediated by RAR/RXR heterodimers, several alternative mechanisms for triggering retinoid-dependent signaling cascades in vertebrates have recently been proposed. For example, RA has been shown to

specifically activate the orphan receptor peroxisome proliferator-activated receptor β/δ (PPARβ/δ), but neither one of its paralogs, PPARα and PPARγ[319]. Interestingly, CRABP is not responsible for RA delivery to this receptor, as this role is played by FABP5[320,321]. Despite the evident existence of this alternative pathway, the affinity of RA for the association with FABP5/PPARβ/δ is much lower than the affinity for CRABP/RAR[321]. Importantly, the ratio of the proteins involved in intracellular RA delivery (FABP5 and CRABP) has been shown to be a crucial component for triggering alternative versus classical RA responses[321]. Given that the activation of the FABP5/PPARβ/δ pathway leads to a very different biological output both in terms of cell survival[321] and differentiation[322], this alternative pathway thus significantly increases the complexity of the RA signaling response.

A number of other nuclear receptors have also been shown to bind and be activated by retinoids. For example, testicular receptor 4 (TR4) has recently been suggested to be a retinoid receptor, because the binding of both retinol and RA can induce conformational changes of TR4 leading to the activation of this receptor[323]. TR4 can either function as homodimer or heterodimerize with testicular receptor 2 (TR2) to function in spermatogenesis, lipid and lipoprotein regulation and central nervous system development[323] and additional experiments are required to reveal the physiological and developmental implications of the activation of TR4 by retinoids, such as retinol and RA[323].

Like TR4, COUP-TFII is capable of recruiting co-activators and activates reporter construct expression in the presence of retinoids[324]. In the case of COUP-TFII, both all-*trans* RA and 9-*cis* RA can associate with and activate the receptor. However, the biological relevance of this interaction is questionable, because the retinoid concentration needed to trigger a response might be significantly higher than endogenous retinoid levels[324]. Future work will nonetheless have to assess the potential role of retinoids in mediating COUP-TFII-specific functions, for example, in neuronal development, cell fate determination or circadian rhythm[324].

RA receptor-related orphan receptor β (RORβ), also functions in the regulation of the circadian rhythm and a recent study has shown that RA can bind RORβ *in vitro* and regulate RORβ activity *in vivo*[325]. Given that RAR has been shown to inhibit a fundamental protein of the circadian clock, CLOCK/BMAL1, at least *in vitro*[326,327], the modification of RORβ functions by RA might represent an alternative pathway for integrating the circadian clock, although further studies are evidently required to understand how RORβ acts in association with RA.

Members of completely different protein families have also been suggested to convey retinoid signals. For instance, the association of retinol-bound RBP with STRA6 has been shown to trigger a signaling cascade mediated by the Janus kinase JAK2 and its associated transcription factors, the signal transducers and activators of transcription (STATs)[322]. This activation of the JAK/STAT cascade by retinol-bound RBP also leads to an increase of expression of STAT target genes, such as inhibitors of insulin signaling and regulators of lipid homeostasis[322].

Moreover, RA has been shown to modulate the activity of protein kinase C α (PKCα),. Thus, all-*trans* RA and acidic phospholipids compete for the binding to PKCα and this competitive binding is responsible for controlling PKCα activity[328]. RA can also activate cAMP response element-binding protein (CREB) in cells, where RXR/RAR heterodimers have been silenced[329]. RA rapidly activates CREB, which leads to a concomitant increase of DNA binding of activated CREB[329]. In parallel, RA is also responsible for an activity increase of both extracellular-signal-regulated kinases 1/2 (ERK1/2) and ribosomal s6 kinase (RSK), a major upstream kinase of CREB[329]. It has also been suggested that PKC is involved in this RA-dependent activation[329]. Another study has shown that RA activates ERK1/2 through induction of mitogen-activated protein kinase kinase 1/2 (MEK1/2) and this specific activation leads to a caspase-3-dependent apoptosis[330].

The phosphatidylinositol 3-kinase (PI3K)/Akt signaling pathway has also been shown to be activated by RA in a very rapid process essential for neural differentiation[331]. Interestingly, an atypical orphan nuclear receptor, DAX1 (dosage-sensitive sex reversal-adrenal hypoplasia critical region on the chromosome X protein 1), has been implicated in this process[332]: RA acting through PI3K/Akt signaling increases DAX1 expression, which leads to modification of nitric oxide signaling levels hence modulating neural differentiation[332].

An example of a non-genomic signaling response to RA is the RAR-dependent activation of p38MAPK and of a downstream mitogen and stress-activated kinase 1 (MSK1)[333,334]. This non-genomic signaling mechanism that ultimately leads to the activation of cdk7 is mediated by a fraction of cellular RAR anchored in lipid rafts at the cell membrane and forming a complex with G protein alpha Q (Gαq)[333]. A fraction of RAR proteins thus function in a RA-dependent, non-genomic manner to integrate processes occurring at the cell membrane. These roles of RA and RAR might serve as fine-tune mechanisms for the rapid

integration of extracellular cues. Future studies will certainly address the biological roles of these non-genomic signaling cascades during development.

Retinoid functions during development

During vertebrate development, RA regulates a significant number of pleiotropic effects, most of which, however, are not exclusively dependent on RA[335]. These processes also rely on signals from other intercellular signaling pathways, such as the fibroblast growth factor (FGF)[336], sonic hedgehog (SHH)[337], WNT, NODAL[338] and bone morphogenetic protein (BMP)[339] cascades. This combined action of RA and other signaling pathways can have very different architectures. Interactions can be synergetic or antagonistic, they may act directly or indirectly on a given target gene, they may also regulate upstream or downstream effectors of a particular cascade or be involved in one or several regulatory loops[335]. In the following section, I will discuss various tissue-specific roles of RA signaling during development (Fig. IV.3) and highlight examples of some of the interactions of the RA pathway with other signaling cascades.

Central nervous system patterning driven by retinoic acid

There is an extensive amount of experimental evidence describing roles of RA signaling in the developing central nervous system (CNS). For example, it has been known for a long time that treatment of vertebrate embryos with exogenous RA results in the loss of forebrain/midbrain territories and in a significant expansion of the hindbrain[340–343]. Thus, the precise regulation of the function of RA during development is fundamental for the proper establishment of the anteroposterior (AP) polarity of the vertebrate CNS[342–344].

The activity of RA in the hindbrain is regulated by dynamic expression of *RAR* and *RXR* (i.e. of the RA receptors), of *RALDH2* (i.e. the RA source) and of the *CYP26*s (i.e. the RA sinks)[291,345,346]. In particular, the expression of RAR paralogs and the activity of different CYP26 enzymes regulate the RA signaling in the hindbrain leading to rhombomere-specific activation of *Hox* genes[291,347]. Furthermore, co-expression of RA synthetic and catabolic enzymes maintains retinoid homeostasis within neurons in the hippocampal region[348].

The AP regionalization of the hindbrain is thus dependent on RA signaling and this function of RA is directly mediated by *Hox* genes[236,347]: within the hindbrain, the direct regulation by RA of the overlapping expression of different *Hox* genes, the so-called *Hox* code, ensures the proper patterning of this CNS region[347,349]. Interestingly, in addition to its

role in transcriptional activation of *Hox* genes, RAR and RXR may also promote *Hox* gene expression through interactions of with chromatin-remodeling enzymes[350].

The anterior limit of *Hox* expression in the CNS is at least in part defined by diffusible signals emanating from the midbrain-hindbrain boundary region, such as FGF8, working antagonistically to the RA signals[351]. RA/FGF antagonism seems to be a recurrent scheme for the regulation of *Hox* genes as such a crosstalk has also been described for the regulation of *Hox* genes along the AP axis of the vertebrate CNS[291]. RA thus preferentially activates anteriorly expressed *Hox* genes, while FGF induces progressively more posteriorly expressed *Hox* genes[351,352].

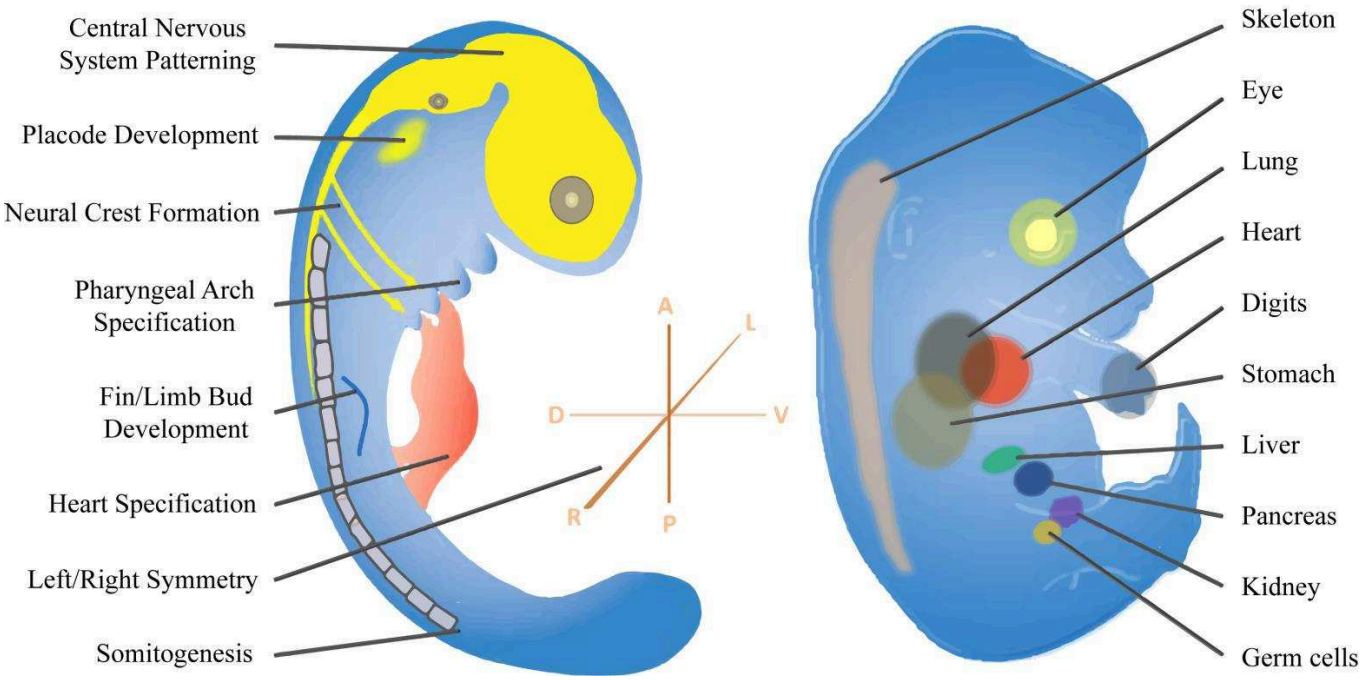


Fig. IV.3 – Summary of retinoic acid (RA) signaling functions during development. *The schematic drawings of two vertebrate embryos chiefly correspond to early somitogenesis (left) and organogenesis (right) stages. RA signaling is involved in a multitude of developmental processes[238,283,353] and the illustration highlights the roles of this signaling cascade detailed in the text.*

In addition to AP regionalization of the CNS, RA signaling also controls neuronal differentiation and specification, and is also involved in the patterning of the neural tube along the dorsoventral (DV) axis[354]. In the spinal cord, RA interacts with SHH to establish ventral determinants that later will give rise to motor neurons. This specification consists of the induction of genes, such as *Pax6*, *Nkx6.1* or *Olig2*, that are responsible for mediating DV patterning of the spinal cord[355]. When RA signaling is not present, the ventral neurons are not induced to differentiate from the spinal cord neuroectodermal progenitor cells[355,356]. Finally, neurite outgrowth during development is also under the control of RA signaling, since a lack of RA signaling is responsible for abnormal axonal projections[343,357].

Retinoic acid signaling in neural crest cells (NCCs) and placode development

NCCs are a transient population of cells in vertebrates that originate at the dorsal neural tube from where they delaminate and subsequently migrate through the embryo[358,359]. NCCs ultimately differentiate into a wide variety of cell types, such as sensory neurons, Schwann cells, myoblasts, chondrocytes, osteocytes or melanocytes, among others[360]. Migrating NCCs are already specified to a particular cell lineage according to their place of origin along the AP axis of the neural tube and hence actively participate in the regional patterning of the embryo[361,362]. For example, it has been shown that exogenous RA triggers abnormal migration of NCCs from rhombomere 4 into the first pharyngeal arch, as opposed to their normal migration to the second pharyngeal arch[361,363].

RA signaling has also been implicated in controlling differentiation, survival and precise migration of NCCs at different developmental stages[280,358,364,365]. One of the most important processes NCCs need to undergo to become migratory is an epithelial-mesenchymal transition, a process that is regulated through an interaction between RA and FGF signaling[358,366]. Indeed, alterations of RA or FGF signaling activity disrupts the timing of epithelial-mesenchymal transitions of NCCs along the AP axis of the CNS, which ultimately leads to changes in cell fate of the NCCs[359].

Placodes are discrete, specialized ectodermal areas located at stereotyped positions in the embryonic head[367]. They give rise to a wide range of cell types, their most important derivatives being sensory neurons that contribute to cranial ganglia[367]. RA signaling has been implicated in the development of the otic[368], lens[299], olfactory[369] and lateral line placodes[370], but the precise functions of RA in the different placodes remain elusive. Eye morphogenesis

and retina differentiation are also controlled by RA, although the RA-dependent signal seems to act primarily on the NCC-derived periocular mesenchyme[299,371].

Retinoic acid signaling and left/right (LR) axial patterning

In addition to patterning along the AP and DV body axes, RA signaling is also implicated in the early determination of LR asymmetry during vertebrate development. Indeed, RA seems to be involved in the side-specific activation of genes involved in LR specification, such as *Lefty*, *Pitx* and *Nodal*[372]. It has initially been suggested that the RA-dependent establishment of LR asymmetry is controlled by vesicles loaded with RA and SHH that are delivered to the future left side of the embryo by rotating cilia in a FGF-dependent process[373]. This hypothesis was subsequently challenged by the finding that an antagonism between FGF8 and RA in the ectoderm is required for generating symmetry. Failure of this antagonism to be established generates excessive FGF8 signaling in adjacent mesoderm that leads to reduction of somite size and LR asymmetry[374]. In addition, it has been shown that RA action during somitogenesis is associated with the maintenance of symmetric somite formation along the LR axis[375]. RA thus works as a buffer to stabilize and balance the LR signaling cues in the presomitic mesoderm (PSM)[375].

Retinoic acid signaling in somitogenesis

Somites are metameric structures of mesodermal origin that give rise to vertebrae and their associated muscles and tendons, the skeletal muscles of the body wall and limbs and to the dermis of the back[376,377]. The biological process of forming new somites appears to be highly conserved during vertebrate embryogenesis and is controlled by rhythmic gene transcription within the PSM, involving mainly NOTCH and WNT signaling components that establish the so-called segmentation clock[340,378]. In addition to its role in the maintenance of LR symmetry, RA produced by RALDH2 in anterior somites is involved in repressing FGF8 signals emanating from the posterior ectoderm[379]. In this way, reciprocal inhibition of FGF8 signals by RA and of RA signals by FGF8 is crucial for positioning segment boundaries in the PSM[380,381]. This interaction is responsible for the translation of pulsations of the segmentation clock into the periodic establishment of segment boundaries, with RA canceling the effects of FGF8 and hence triggering the expression of segmentation genes responsible for somite formation[379,380]. Although, existing experimental data questions this paradigm of interaction between RA and FGF8 and until what extent it affects somitogenesis[378]. Furthermore, through

the control of *hox* gene expression, the RA/FGF antagonism also has a very important function in defining AP positional information in structures derived from the PSM[350].

Retinoic acid in heart field specification

RA signaling also plays important roles in patterning the developing heart, in particular during cardiac field specification during AP regionalization of the heart tube and in the establishment of LR asymmetry during heat looping[278,382–388]. RA signaling is initially required to limit the size of the cardiac progenitor field[389] and later to limit the specification of both atrial and ventricular cells[390].

Interestingly, at least the role of RA in AP patterning of the heart is well conserved within vertebrates, given that exogenous RA leads to similar heart regionalization defects in mice[391], chicken[392] and zebrafish[393]. In vertebrates, RALDH2 is expressed early in the heart mesoderm in an AP wave hence defining cardiac precursors and delimitating the heart field along the AP axis[392]. RA signaling, also probably exerts its effect on AP regionalization of the heart at least in part by downregulating *Isl1* expression, potentially by reducing FGF signaling levels[394].

Retinoic acid signaling and kidney development

The kidney is another target of RA signaling during development. The kidney field is derived from intermediate mesoderm and studies have shown that RA signaling is required for the early specification of this kidney field[395–397]. In the frog *Xenopus laevis*, for example, early determinants of a pronephric fate, such as *Pax8* and *Lhx1*, are controlled by RA signaling[396] and, in zebrafish, RA is required for positioning and segmentation of the pronephric kidney[397]. Moreover, frog *Pteg* and zebrafish *Wt1*, both of which play important roles during kidney development in the respective animal, are direct RA signaling targets[398,399]. In mice, it has further been shown that RA signaling, by controlling *Ret* expression in ureteric bud cells, is crucial for ureteric bud formation and for branching morphogenesis within the developing collecting duct system[400]. Interestingly, the RA signaling activity in ureteric bud cells depends mainly on RA generated by RALDH2 in adjacent stromal mesenchyme cells[400].

Relevance of retinoic acid signaling for body appendage development

Pharmacology-based experiments had shown, for example, that exogenous RA induces mirror-image duplications in chick wing buds[401], it had initially been suggested that RA signaling plays crucial roles in limb bud patterning of tetrapod vertebrates[402]. However, more recent work suggests that RA signaling might be less important for tetrapod limb development than initially expected[403]. Albeit required during forelimb induction, RA might not be necessary at all for hindlimb budding and patterning[404]. Moreover, the involvement of RA in forelimb formation seems to be indirect, creating a permissive environment for FGF signaling to develop a limb bud[404]. Conversely, in zebrafish RA is indispensable for pectoral fin induction[405,406], which is suggestive of a divergence in RA signaling functions between the fin buds of fish and the limb buds of tetrapods[402]. At later developmental stages, RA signaling controls tissue remodeling and apoptosis at the digit-interdigit junction, with *RALDH2* being expressed in interdigital mesenchyme and *CYP26B1* in digits, hence restricting RA action to the interdigital zones[407]. This activity is necessary for interdigital tissue loss, as shown by $RAR\beta^{-/-}RAR\gamma^{-/-}$ double mutants, as well as by $RALDH2^{-/-}$ and *RDH10* mutants[408,409].

Endoderm specification and retinoic acid

RA signaling is also essential for patterning the developing endoderm. In particular, RA is required for the determination along the AP axis of endodermal fields, each of which will give rise to a particular organ[410]. For example, exposure to exogenous RA prevents the expression of genes normally present in the most anterior endoderm and, at the same time, activates and shifts anteriorly the expression of genes normally expressed at more posterior endoderm levels[410].

In the anterior endoderm, RA signaling is required for the regional patterning of the pharyngeal arches. The pharyngeal arches are structures that develop along the embryonic foregut endoderm in a very complex process that involves the interaction between cells from all germ layers (ectoderm, mesoderm, endoderm and the neuroectoderm-derived NCCs). RA is produced locally in the pharyngeal arches by RALDH2 in the mesoderm adjacent to the endoderm[410,411]. When RA signaling is disrupted, the first and second pharyngeal arches develop normally, while severe defects are observed in structures derived from more posterior pharyngeal arches[411]. Interestingly, RA seems dispensable for the specification of pharyngeal

endoderm, but required for morphogenesis and segmentation of the posterior pharyngeal arches[412].

Some organs derived from the endoderm in the posterior foregut also require RA for proper development[413]. These organs include the lungs, the stomach, the liver and the pancreas[413]. In the lung, for example, loss of RA signaling causes a disruption of lung bud outgrowth[413]. This defect is due to a reduction of FGF10 levels, which, together with BMP4, is required for mediating proper branching morphogenesis[413]. Activation of FGF signaling by RA is also required for early stomach and liver development[413], while, during pancreas development, mesodermal RALDH2 is required for specification of the dorsal, but not the ventral pancreatic bud[414,415].

Retinoic acid signaling and evolution

Retinoid signaling was long thought to be vertebrate-specific, but recent studies have revealed major roles for RA signaling that are conserved in all chordates (Fig. IV.4), which, in addition to the vertebrates, include two invertebrate groups: the cephalochordates and the urochordates (also called tunicates)[402,416,417]. While the tunicates constitute the sister group of vertebrates, the cephalochordates are located at the base of the chordate phylum[87,99]. Nonetheless, of all invertebrates the cephalochordates, such as the lancelet or amphioxus, are characterized by the most vertebrate-like retinoid signaling system, both in terms of molecular composition and biological functions[88,99,239,402].

Inside chordates

When considering the key players of RA metabolism some conserved and divergent features are observable when comparing vertebrates with invertebrate chordates (Fig. IV.4). For example, while vertebrate genomes encode multiple RAR and RXR paralogs, amphioxus has only one RAR and one RXR, forming a single heterodimer[418]. Like its vertebrate counterparts, this amphioxus RAR/RXR heterodimer binds to and is activated by RA[418]. Using ligands specific for the three mammalian RAR paralogs (RARα, RARβ and RARγ), it was shown that the ligand binding capacities of amphioxus RAR resemble those of mammalian RARα and RARβ. Although, only RARβ-selective retinoid compound activates transcription of a reporter suggesting that the ancestral RAR of vertebrates was of the RARβ type[419,420]. Comparison of developmental gene expression patterns between amphioxus RAR and the three vertebrate RAR paralogs further supported this notion and suggested that, while

RARβ kept the ancestral features in terms of ligand specificity and development expression after duplication, RARα and RARγ might have acquired novel functions in the course of vertebrate diversification[419].

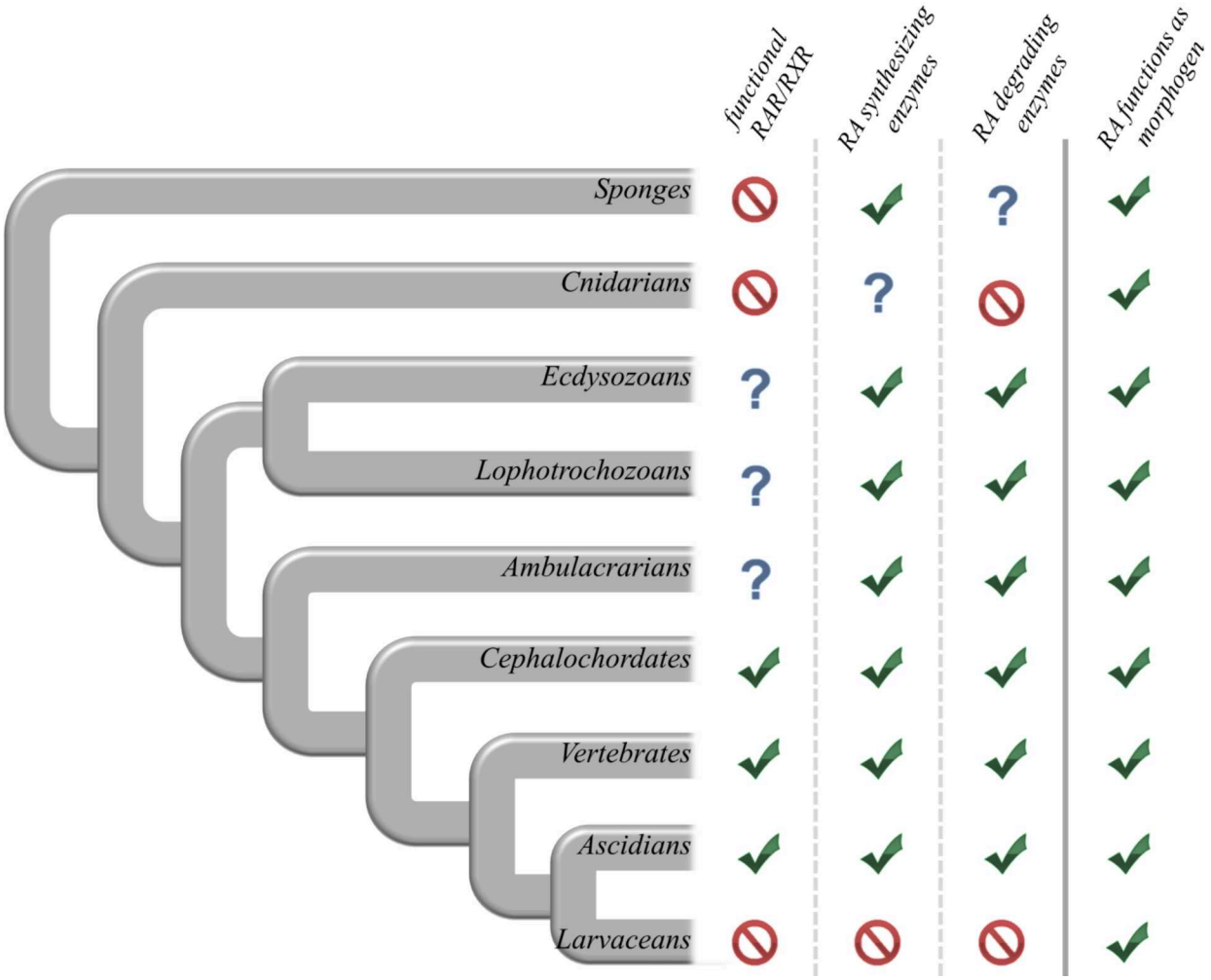


Fig. IV.4 – Evolution of RA signaling components and functions. *A simplified phylogenetic tree of metazoan animals is shown and the presence (✓) or absence (⊘) of functional RAR/RXR heterodimers, of RA synthesizing (RALDH) and of RA degrading (CYP26) enzymes is highlighted. Uncertainty over presence or absence is illustrated by (?). The phylogenetic tree also indicates the distribution of intercellular signaling functions of RA (as morphogen) in the different lineages. RA, retinoic acid; RAR, retinoic acid receptor; RXR, retinoid X receptor.*

Although the amphioxus genome encodes orthologs of the enzymes required in vertebrates for RA synthesis (RALDH) and degradation (CYP26), it remains to be determined, whether these cephalochordate RALDH and CYP26 are indeed capable of, respectively, RA synthesis and degradation[127,421]. Regarding the endogenous production of retinaldehyde from retinol, a *bona fide* ADH has previously been characterized biochemically in amphioxus[422–424], but the functional capacities of the RDH orthologs that have been identified in cephalochordate genomes still need to be assessed[123,127]. Interestingly, although a candidate for an intracellular retinol binding protein (CRBP) has previously been described in amphioxus, there is no convincing evidence for the existence of an amphioxus intracellular RA binding protein (CRABP)[123,127,425], and the concerted absence in amphioxus of the main components for retinol storage, transport and cellular uptake further suggest that these systems might be functional innovations of vertebrates[127].

Several studies have assessed the roles of RA signaling during amphioxus development [88,402,418,426]. For example, in the amphioxus CNS it was shown that RA regulates both AP regional patterning and neuronal specification in a *Hox*-dependent manner[427]. Likewise, the amphioxus ectoderm is also patterned by RA: disruption of endogenous RA signaling levels lead to a alterations of an ectodermal *Hox* code, which in turn leads to AP regionalization defects and changes in the combinatorial code defining specific populations of ectodermal sensory neurons[428]. Recently, it has also been shown that, at later developmental stages, RA signaling is involved in the formation of the amphioxus tail fin, which arises from the posterior ectoderm[429]. In the endoderm, RA signaling via *Hox1* has been shown to be required for the definition of the posterior limit of the amphioxus pharynx by limiting expression of pro-pharyngeal genes to the most anterior (thus pharyngeal) endoderm[430]. Moreover, in the posterior endoderm, RA signaling controls the expression of two *Parahox* genes, *Xlox* and *Cdx*, hence mediating AP patterning of the developing amphioxus hindgut[431]. While it is very likely that RA signaling directly activates expression of amphioxus *Hox* genes[432], further work is needed to assess in more detail the regulation of amphioxus *Parahox* genes by RA[431].

Contrasting the situation in the cephalochordate amphioxus, RA signaling in tunicates has secondarily been modified in different tunicate lineages[88,402,416,433]. For example, although the genome of the ascidian tunicate *Ciona intestinalis* contains the basic molecular components of the RA signaling cascade, that is a single receptor heterodimer (RAR/RXR) and enzymes potentially capable of RA synthesis (RALDH) and degradation (CYP26), the regulation of at least some these components in this tunicate has been modified[88,402,434]. Thus,

albeit directly regulated by RA in vertebrates and amphioxus, transcription of RAR in this ascidian tunicate is not regulated by RA[435]. Moreover, the larvacean tunicate *Oikopleura dioica* has lost most of the molecular components defining the canonical RA pathway, including the receptor RAR as well as RA synthesizing (RALDH) and degradation (CYP26) enzymes, a scenario associated with a regressive evolution event[433,436].

Along these lines, the roles for RA signaling during tunicate development seem to be limited. For example, although RA signaling might function during development of the ascidian tunicate CNS[435,437], RA probably only has a limited role in the regulation of *Hox* gene expression[438–442]. Moreover, given that tunicates in general have lost several *Hox* genes and that both cluster organization and collinear expression have at least partially been lost[443,444], the contribution of a RA-regulated *Hox* code to tunicate CNS development is very limited[443]. Interestingly, exogenous RA has been shown to activate and shift anteriorly the expression of ascidian *Hox1* in both the CNS and ectoderm[438,441,442,445,446] and, at least in the ectoderm, this activation is directly mediated by the ascidian RAR/RAR heterodimer, which represents a very rare example for RA-dependent gene expression in ascidians[441,442]. This RA-*Hox1* network in the ectoderm is required for formation of the ascidian atrial siphon placode, which is considered homologous to the vertebrate otic placode[447]. The *Hox1* gene is also expressed in the ascidian endoderm[443], where exogenous RA leads to a loss of expression of pharyngeal markers and to a posteriorization of the endoderm leading to a complete loss of the branchial basket[448–450]. Intriguingly, these effects of RA are reminiscent of the phenotypes observed in amphioxus and vertebrates following treatment with exogenous RA[402].

Taken together, at the base of the chordate lineage the basic molecular components constituting the RA signaling cascade were most likely already present. Moreover, the RA pathway was probably involved in the specification of neuronal cell populations in both CNS and ectoderm and RA signals also contributed to the regional patterning along the AP axis of the CNS, ectoderm and endoderm. In all three tissue layers, this RA-dependent AP regionalization process was probably directly mediated by *Hox* genes[402,417]. In the tunicate lineage, the functions of the RA signaling cascade have been secondarily modified leading to diverged functions, such as the partial loss of RA-dependent regulation of the *Hox* code in ascidians, and culminating in the loss of key components of the RA pathway, including RAR, RALDH and CYP26 in larvacaens[433,434].

Outside chordates

Outside the chordate lineage, evidence for functional roles of retinoids is much scarcer (Fig. IV.4). Although genes encoding orthologs of the vertebrate receptors RAR and RXR as well as orthologs of the basic vertebrate components for synthesis (RALDH) and degradation (CYP26) of endogenous RA are present in ambulacrarians (such as hemichordates and echinoderms), lophotrochozoans (such as annelids and mollusks)[239,402,451] and also in ecdisozoans (i.e. at least in priapulids also known as penis worms where CYP26 genes can be identified). Extensive experimental evidence for retinoid and/or RAR/RXR heterodimer functions in these lineages is unfortunately still lacking. It has been suggested that RA treatments results in a delay in embryonic development of echinoderms and, in mollusks, retinoids (in particular 9-*cis* RA) disrupt embryogenesis and are involved in neuronal differentiation, outgrowth, and growth cone guidance[452–455]. This role of retinoids in mollusk neurogenesis is apparently independent of a functional RAR/RXR heterodimer, but instead requires a cytoplasmic localization of RXR[456]. This is also supported by recent reports that show that mollusk RAR is unable to bind all-*trans*-RA as well as other retinoids and to activate transcription or recruit coactivators[249,457]. Also, in mollusks RA affects eye formation and leads to arrested development at the trochophore larval stage[458].

Intriguingly, the currently available *in silico* data suggest that some ecdysozoans (which include insects, crustaceans and nematodes) might have secondarily lost both the canonical retinoid receptor RAR and a CYP26-based RA degrading machinery[239,402,451], although CYP26s genes where recently identified in priapulids (Carvalho et al. in press). Nevertheless, retinoid functions have been described in tissue regeneration of both insects and crustaceans[459–461]. Other functions of retinoids in animals, whose genomes do, *a priori*, not encode the RAR receptor include, for example, hydroid specification, cell proliferation and neuronal differentiation in cnidarians[462,463] and tissue regression and spicule formation in sponges[462,464]. Collectively, these data suggest that in the course of metazoan evolution the functions for retinoids might have originated before a functional RAR/RXR heterodimer and that the presence of both RAR and RXR in a genome does not automatically imply a functional RAR/RXR heterodimer capable of binding and being activated by retinoids. In this context, it is interesting to note that retinoids have very ancient roles in light reception and the visual system, which represents a pool of retinoids that, in the course of evolution, might have been co-opted for roles in morphogen-dependent signaling and that has ultimately been elaborated into one of the most important intercellular signaling cascades in development.

Chapter V: Amphioxus and the study of retinoic acid signaling

In amphioxus, RA signaling manipulations cause a very severe morphogenetic phenotype. Constant exposure of the amphioxus developing embryos to RA, starting from the early gastrula (G1 stage), gives rise to a malformed larva lacking or presenting reduced anterior structures, such as the pre-oral pit, mouth and gill slits, as well as it shows some important defects at the posterior end of the larva, with for instance a reduced or missing anus[426,465]. Phylogenetic, genomic, and structural analyses show that amphioxus genome might lack some components of vertebrate retinoid signaling, such as the four main components involved in storage, transport, and cellular uptake of retinoids. Although, its genome encodes all the basic components involved in retinol oxidation and retinal reduction (i.e. SDR-retinol dehydrogenases), RA synthesis and degradation (i.e. ALDH1a, ALDH8 and CYP26) as well as single copies of RA and retinoid X receptors (i.e. RAR and RXR)[127]. For the genes involved in the retinol oxidation, retinol reduction, RA synthesis and RA degradation amphioxus exhibits a large expansion of these gene families, suggesting a very intricate mechanistic control of the retinoid metabolism throughout the amphioxus life cycle[127].

The expression of the single amphioxus retinoic acid receptor (RAR) starts during the gastrula stage (G3 stage) showing an ubiquitous expression, although ectodermal expression seems stronger than the presumptive endomesoderm domain of expression[418]. This suggests that RA-dependent patterning in amphioxus initiates during gastrulation. Later, during the N2 stage RAR expression is down-regulated anteriorly in the neural plate and in the endoderm, as well as globally in the general ectoderm[418]. In the following developmental stages (i.e. N3 and N4 stage) RAR gene expression domain is further reduced and is now restricted to the middle third of the embryo, with some weak expression towards the posterior end[418]. By the T stage, the expression of this gene is restricted to the middle region of the endoderm, in a very small domain that is almost not identifiable later during the larval development. Furthermore, at the T stage, the expression of RAR in the amphioxus neural tube is present from the anterior border of the 1st somite till the posterior side and it exhibits a domain of expression in the

somites, more precisely in the middle third of the embryo[418]. Globally, in the presence of exogenous RA, RAR expression is upregulated anteriorly in a small domain of the pharynx, in all the neural tube and anteriorly in the region of the tail bud, endoderm and general ectoderm. Differently, in the presence of a RAR antagonist the expression of this gene at the T stage seems to be mostly abolished[418].

In amphioxus, the RA synthesizing enzymes repertoire is potentially constituted by six genes of the ALDH1a subfamily (i.e. *ALDH1a*, *ALDH1b*, *ALDH1c*, *ALDH1d*, *ALDH1e* and *ALDH1f*), originated by lineage-specific duplication, five of which are located on two distinct genomic scaffolds, and by one *ALDH8a* encoding gene[127,416,421]. This represents a large expansion of the genes encoding for RA synthesizing enzymes, since vertebrates possess, in general, four retinoic acid dehydrogenases, *ALDH1a1*, *ALDH1a2*, *ALDH1a3* and *ALDH8a*[416]. Even though, the six *ALDH1a* subfamily genes of amphioxus are expressed in very similar domains, at least during the neurula stages, being located mainly in posterior dorsal domains of the embryo with some expression in the endoderm[421]. These data suggests that RA signaling is very likely dynamic during amphioxus embryogenesis and potentially involves a posterior source, from where the morphogen diffuses throughout the embryo, thus creating a posterior-high signaling gradient. In addition, amphioxus has three *CYP26* genes encoding RA degrading enzymes, CYP26-1, CYP26-2, and CYP26-3, which are clustered on a single scaffold, suggesting their origin by lineage specific duplication[127]. The expression, functional implication and evolutionary diversification of the three amphioxus RA degrading enzymes will be further addressed in the results section.

It has been shown in amphioxus that Wnt/β-catenin signaling, in the region of the blastopore, specifies the posterior end of the embryo, while RA signals confer different positional identities along the A-P axis, mainly by modulating the collinear expression of *Hox* genes[427,432,466]. For instance, the RA-*Hox* hierarchy is responsible for both patterning of the developing epidermal sensory neurons and CNS as well as specification of neuronal identity[427,428]. Specifically in the anterior neuroectoderm, where *Hox1* expression is absent, RA signaling is inhibited partially by the competitive RAR-inhibitor orphan receptor TR2/4[418]. Furthermore, *Hox1* expression driven directly by RA signaling[440] is responsible by pharyngeal patterning, since it establishes the posterior boundary of expression of pharyngeal genes (e.g. *Otx*, *Pax1/9*, *Nodal*, *Notch* and *Pitx*), thus setting the posterior limit of the pharynx[430]. In the gut endoderm *Xlox* expression is expanded posteriorly in response to exogenous RA, whereas the *Cdx* domain in the gut is reduced anteriorly, thus the entire A-P

axis of the amphioxus endoderm seems to be patterned through RA-dependent mechanisms[431]. RA also functions in pharyngeal segmentation in a regulatory network involving the mutual inhibition of RA- and Tbx1/10-dependent signaling[467]. It seems that globally, RA signaling functions posterior to the cerebral vesicle but anterior to the tail bud of amphioxus, to modulate the expression of *Hox* genes and several other transcription factors involved in A-P patterning.

Gene candidate approaches have been widely used in amphioxus to identify both direct and indirect targets of the RA signaling pathway. So far, gene expression changes induced by RA signaling at the gastrula stage have been reported for only a very reduced set of genes that includes *RAR*, *Hox1*, *Hox3*, *Hox4*, *Hox6*, *HNF3-1* (*FoxAb*) and *Wnt3*. While for most of these genes RA signaling, through the action of RAR binding to RAREs, might directly regulate expression levels[432], *Wnt3* expression seems to be regulated indirectly[466]. In striking contrast to what is observed in vertebrates, the expression patterns of seven other amphioxus Wnt genes (*Wnt1*, *Wnt4*, *Wnt5*, *Wnt6*, *Wnt7*, *Wnt8* and *Wnt11*), as well as those of several Wnt antagonists (*Dkk1/2/4*, *Dkk3*, *sFRP2-like*, *sFRP3/4*) seem not to be affected by changing RA levels[432]. It seems that RA patterns the A-P axis during gastrulation by directly regulating expression of *Hox* genes and of *FoxAb*. Later in development, these genes are in turn involved in the definition of positional identity during the neurula stage. Hox genes regulate the expression of gene involved in the patterning of the CNS, like *ERR*, *Otx* and *Pax2/5/8*[427], as well as *Coe*, *ERR* and *Islet* implicated in the establishment of epidermal sensory neurons in the developing embryo[428]. Surprisingly, it has been shown that RA signaling pathway is not implicated in somitogenesis, contrasting with the situation verified in vertebrates[379,468]. RA signaling do not functions in FGF dependent anterior somite formation or in posterior elongation, since activating or inhibiting the RA pathway at early stages or during posterior elongation leads to no alteration in number and shape of the newly formed somites[468].

All in all, even though the amphioxus RA signaling pathway started to be studied about 20 years ago, not much is know about the bioavailability of RA during development. Moreover, the target genes of the RA signaling pathway and their hierarchical relationship during amphioxus development represent an interesting open question. Therefore, this work aims at providing a detailed description of two fundamental aspects of the RA signaling pathway during amphioxus development: (1) the regulation of the bioavailability of RA in the developing embryo and (2) the target genes under the control of the RA signaling pathway, as

well as their hierarchical regulatory relationship. To address these questions, the European amphioxus, *Branchiostoma lanceolatum*, was used as a model system.

Not only these questions were fundamental during my research project, but also the implementation of amphioxus as a reliable model system and thus the establishment of multiple aquaculture improvements (already discussed in Chapter III) as well as *in vivo* techniques, such as the microinjection of mRNAs in amphioxus eggs.

To obtain insights into the bioavailability of RA during development of amphioxus, I focussed on the study of the enzymes that mediate endogenous RA catabolism, the CYP26 subfamily. CYP26 enzymes are known to oxidize RA into other retinoid compounds that do not have a biological function. Here, I characterized not only the evolutionary diversification of CYP26 genes in deuterostomes, but also the expression of these genes, their function as well as the mechanism that governs the feedback loop controlled directly by RA during amphioxus development.

Differently, to characterize the target genes under the control of the RA signaling pathway during development of amphioxus, I combined pharmacology treatments using retinoid specific drugs with two different techniques of high throughput sequencing: RNAseq that reveals the entire RNA profile, and thus the genes being expressed at a given moment in time, and ATACseq (assay for transposase-accessible chromatin) that provides a global overview of accessible regions of the chromatin (i.e. open chromatin regions). By combining the data obtained by these techniques I identified genes that are under the control of RA signaling and, in parallel, established a framework for characterizing gene hierarchies during development.

Results

Overview:

- **R1:** ***Expression of fluorescent proteins in*** **Branchiostoma lanceolatum** ***by mRNA injection into unfertilized oocytes***
- **R2:** ***Lineage-specific duplication of amphioxus retinoic acid degrading enzymes (CYP26) resulted in sub-functionalization of patterning and homeostatic roles***
- **R3:** ***RNAseq and ATACseq data reveals new targets and mechanisms of action of the RA-signaling pathway in amphioxus development***

R1: Expression of fluorescent proteins in *Branchiostoma lanceolatum* by mRNA injection into unfertilized oocytes[2]

Short abstract

We report here the robust and efficient expression of fluorescent proteins after mRNA injection into unfertilized oocytes of *Branchiostoma lanceolatum*. The development of the microinjection technique in this basal chordate will pave the way for far-reaching technical innovations in this emerging model system, including *in vivo* imaging and gene-specific manipulations.

Long abstract

We report here a robust and efficient protocol for the expression of fluorescent proteins after mRNA injection into unfertilized oocytes of the cephalochordate amphioxus, *Branchiostoma lanceolatum*. We use constructs for membrane and nuclear targeted mCherry and eGFP that have been modified to accommodate amphioxus codon usage and Kozak consensus sequences. We describe the type of injection needles to be used, the immobilization protocol for the unfertilized oocytes, and the overall injection set-up. This technique generates fluorescently labeled embryos, in which the dynamics of cell behaviors during early development can be analyzed using the latest *in vivo* imaging strategies. The development of a microinjection technique in this amphioxus species will allow live imaging analyses of cell behaviors in the embryo as well as gene-specific manipulations, including gene overexpression and knockdown. Altogether, this protocol will further consolidate the basal chordate amphioxus as an animal model for addressing questions related to the mechanisms of embryonic development and, more importantly, to their evolution.

[2] article published in the Journal of Visualized Experiments in 2015, accompanied by a video protocol[469]. Authorship: Estelle Hirsinger, João E. Carvalho, Christine Chevalier, Georges Lutfalla, Jean-François Nicolas, Nadine Peyriéras and Michael Schubert. Full article can be found in Appendix 7. Video featuring João E. Carvalho. Link: https://goo.gl/1b24Ew

Introduction

During development, a single cell gives rise to an entire organism in a highly complex process that involves both cell divisions and movements. To better understand the biological principles underlying the dynamics of cell behavior, developmental biologists have started to use fluorescence-based *in vivo* imaging techniques. Specific compartments of cells, such as cell membranes, can either be labeled by treatments with fluorescent dyes, an approach hampered by a lack of specificity and of tissue penetration[470], or by the specific introduction into the embryo of exogenous mRNAs encoding fluorescent proteins[471]. Different techniques can be used for the efficient delivery of exogenous compounds, such as mRNAs. These include, but are not limited to, microinjection, electroporation, bombardment with microparticles, lipofection and transduction[472,473]. Although all of these approaches can be used to introduce exogenous compounds into a developing embryo, only microinjection allows the application of predefined and precise quantities into each cell[472]. Microinjection techniques have been described for all major developmental model systems[473] (e.g., fruit flies, nematode worms, zebrafish, frogs, mice) as well as for some alternative models[473], including those used for comparative studies aimed at understanding the evolution of developmental mechanisms (e.g., sea anemones, annelid worms, sea urchins, ascidian tunicates, the cephalochordate amphioxus).

Cephalochordates, which together with tunicates and vertebrates establish the chordate phylum, are particularly well-suited models to study the evolution of chordates and the diversification of vertebrates from an invertebrate ancestor[93,97,99,474]. The cephalochordate lineage diverged very early during chordate evolution; and extant cephalochordates, which are subdivided into three genera (*Branchiostoma*, *Asymmetron* and *Epigonichthys*), resemble vertebrates both in terms of overall anatomy and genome architecture[93,97,99,474]. Of the about 30 species of cephalochordates that have been described so far, five are available for embryological and developmental studies[97,122]. *Asymmetron lucayanum* (the Bahama lancelet), *Branchiostoma floridae* (the Florida amphioxus), *Branchiostoma lanceolatum* (the European amphioxus), *Branchiostoma belcheri* (the Chinese amphioxus) and *Branchiostoma japonicum* (the Japanese amphioxus). Ripe adults of three of these species (*B. lanceolatum*, *B. belcheri* and *B. japonicum*) can be induced to spawn on-demand during the breeding season[196,198]. In addition, at least for *B. lanceolatum*, efficient spawning can also be induced in artificial sea water[148], thereby making this particular cephalochordate species accessible for laboratories that do not have access to natural seawater. The combination, in *B. lanceolatum*,

of a convenient and reliable access to embryos with an efficient delivery method, such as microinjection, so far the only delivery technique developed in amphioxus (in both *B. floridae* and *B. belcheri*)[174,475,476], will enable the development of a novel suite of manipulative techniques, including lineage tracing- and dynamic cell behavior-based approaches.

A protocol for the efficient microinjection of mRNAs to express fluorescent proteins in the *B. lanceolatum* embryo was hence developed. Furthermore, to provide a basic toolkit for live imaging of *B. lanceolatum* embryos, vector systems were developed that allow membrane-associated and nuclear expression of fluorescent proteins. For membrane targeting, enhanced Green Fluorescent Protein (eGFP) was fused to the human HRAS CAAX box and nuclear localization of mCherry and eGFP was obtained by fusion to the zebrafish histone 2B (H2B) exon (Fig R1.1). Furthermore, with the goal to optimize protein translation, the Kozak sequences and codons of the constructs have been modified and adapted to usage in *B. lanceolatum*. Taken together, the injection method and expression vectors presented here will serve as a basis for the generation of new experimental approaches for cephalochordates, notably analyses using the latest fluorescence-based *in vivo* imaging techniques.

Protocol

1. Preparation of instruments and reagents

1.1 Transfer Pasteur Pipettes

1.1.1 Generate a series of transfer Pasteur pipettes with different tip diameters by pulling 230 mm long Pasteur pipettes above a flame at different speeds. Ensure that the taper is as long as possible for smooth and fine control of aspiration.

1.1.2 With a diamond scribe, scratch the pipette along a line perpendicular to the length of the pipette. With both hands, pull the pipette parallel to its length to generate a blunt cut. Swiftly flame-polish the pipette without sealing the tip.

1.1.3 Vary the diameter of the tip with the stage of the oocytes/embryos to be pipetted: 300-400 μm for unfertilized oocytes (around 150 μm in diameter), 600 μm for fertilized eggs (around 500 μm in diameter), 200-400 μm for hatched neurulae. Use pipettes with a mouth-monitored aspiration tube to transfer oocytes/embryos from one dish to another.

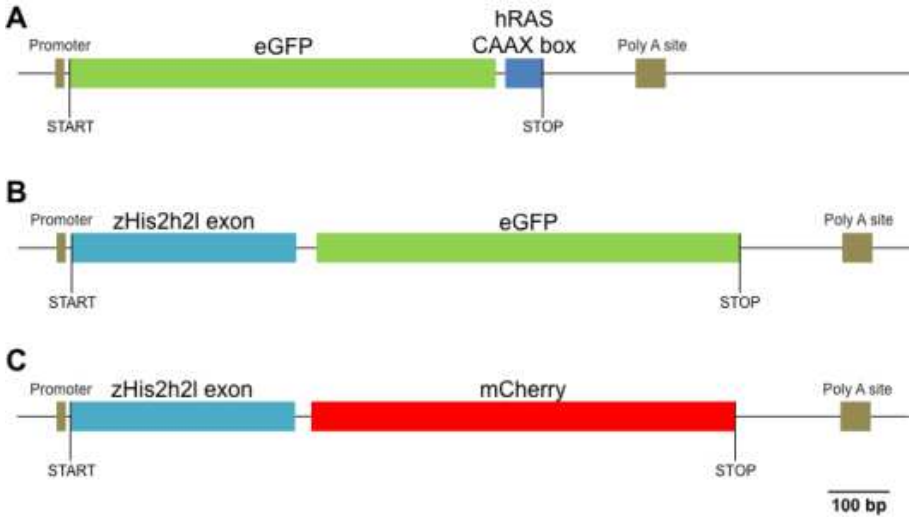


Fig. R1.1 – Construct details. *Sketches of (A) the membrane-targeted eGFP (eGFP:CAAX box), (B) the nuclear-targeted eGFP (H2B:eGFP) and (C) the nuclear-targeted mCherry (H2B:mCherry) constructs are shown.*

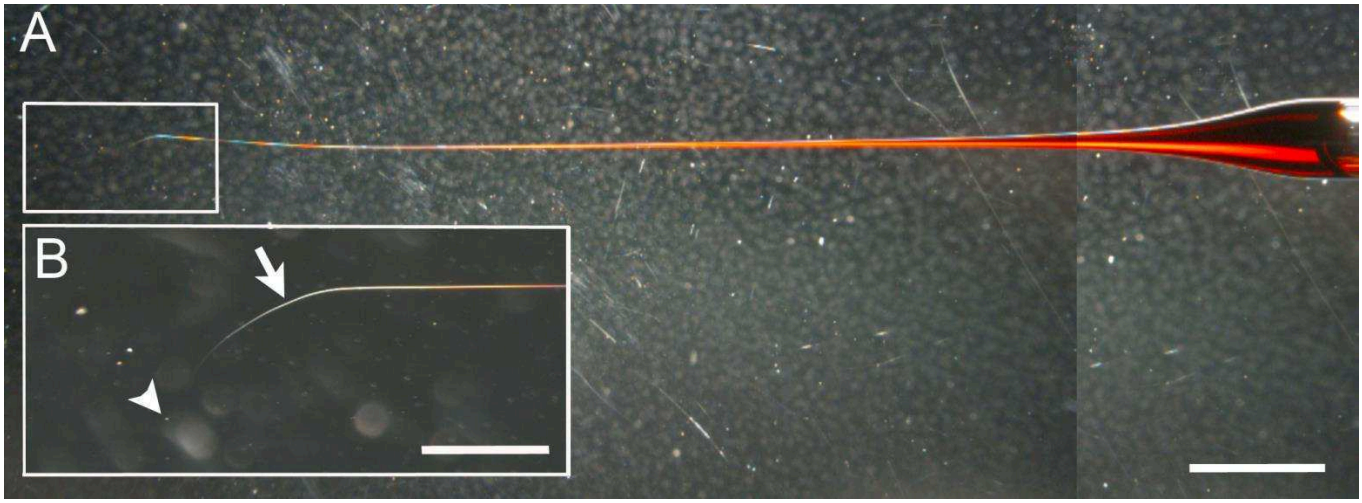


Fig R1.2 – Shape of a representative injection needle. *(A) The tamper is about 2 cm in length and the needle curves at the tip. Scale bar = 2 mm. (B) Magnification of the boxed area in (A). The position for clipping the needle (arrow) is in the curve of needle. The tip of the needle is indicated by the arrowhead. Scale bar = 500 µm.*

1.2 Injection needles

1.2.1 Manipulate the capillaries with gloves to ensure RNase-free conditions. Use borosilicate glass with filament capillaries with dimensions of OD 1.20 mm, ID 0.94 mm, length 10 mm.

1.2.2. If capillaries are pulled on the type of heating-filament needle puller described in the Table of Equipment and Reagents, use the following settings: Heat 600, Pull 50, Velocity 80, Time 60, Pressure 200 or 300. Otherwise, ensure that the shape, which is crucial for successful injections, is as shown in Fig. R1.2 with the following properties: 4-8 µm outer tip diameter, 2 cm taper length. Note: Needles can be pulled before the spawning season and used throughout the season.

1.3 5% Phenol Red stock solution (4x)

1.3.1 Prepare fresh before each spawning season.

1.3.2 In a 0.22 µm-filtration tube, weigh out 25 mg of Phenol Red powder.

1.3.3 Add 0.5 ml of DNase- and RNase-free water to the tube containing the powder.

1.3.4 Spin for 3-5 min at 18000 x g at room temperature to filter-sterilize the solution and remove crystals that could clog the injection needle.

1.3.5 Store at 4 °C or store 250 µl aliquots at -20 °C.

1.4 0.25 mg/ml Poly-Lysine Solution

1.4.1 Prepare fresh before each spawning season.

1.4.2 Dissolve 5 mg of poly-L-lysine in 20 ml distilled water. Store 5 ml aliquots at -20 °C.

1.4.3 Use a defrosted aliquot immediately and only once to ensure reproducible and robust adhesion of the oocytes to the poly-lysine-coated dish.

1.5 mRNA synthesis

1.5.1 Prepare fresh before each spawning season.

1.5.2 Linearize 5 µg of DNA of interest with the adequate enzyme (usually for 2 hrs, at 37 °C).

To check the completeness of the digestion, run 2% in volume of the digestion mix on a 1% agarose-TBE gel in TBE buffer at 150 W for 20 min.

1.5.3 Extract the linearized DNA with 25:24:1 phenol (pH 8.0):chloroform:isoamyl alcohol. Vortex for 20 sec, centrifuge 10 min at 18000 x g and collect the aqueous (upper) phase.

1.5.4 Extract the aqueous phase again with 24:1 chloroform:isoamyl alcohol. Vortex for 20 sec, centrifuge 10 min at 18000 x g and collect the aqueous phase.

1.5.5 Precipitate the linearized DNA with 100:10:300 linearized DNA: 3 M sodium acetate (pH 5.2): 100% ethanol overnight at -20 °C.

1.5.6 Centrifuge 20 min at 18000 x g at 4 °C. Rinse in 70% ethanol.

1.5.7 Centrifuge 10 min at 18000 x g at 4 °C. Let dry and resuspend in RNase-free water at a final concentration of 0.5 µg/µl.

1.5.8 Transcribe 1 µg of linearized DNA using an mRNA synthesis kit with the appropriate polymerase, according to manufacturer's instructions.

1.5.9 Extract the mRNA with 5:1 phenol (pH 4.7): chloroform and ammonium acetate stop solution provided with the kit.

1.5.10 Vortex for 20 sec, centrifuge 10 min at 18000 x g and collect the aqueous phase.

1.5.11 Extract the aqueous phase again with 24:1chloroform: isoamyl alcohol. Vortex for 20 sec, centrifuge 10 min at 18000 x g and collect the aqueous phase.

1.5.12 Precipitate the mRNA with 100% isopropanol overnight at -20 °C.

1.5.13 Centrifuge 20 min at 18000 x g at 4 °C. Rinse in 80% ethanol.

1.5.14 Centrifuge 10 min at 18000 x g at 4 °C. Let the pellet dry at room temperature for no longer than 5-10 min as it will then be difficult to resuspend. Resuspend in DNase- and RNase-free water to a final concentration of at least 2 µg/µl to ensure a decent final mRNA concentration in the injection mix.

1.5.15 To check the quality and size of the transcription product, run 0.5 µl of the mRNA on a RNase-free 1% agarose-TBE gel in TBE buffer at 150 W for 20 min. Store 2 µl aliquots at -80 °C.

1.6 Poly-Lysine-coated dishes

Note: Poly-Lysine coated dishes are used to immobilize the oocytes during injection.

1.6.1 For each 35 mm cell-culture Petri dish (5 in total), cover the bottom of the Petri dish with 1 ml of the thawed 0.25 mg/ml poly-lysine solution. Incubate at room temperature for 5 min.

1.6.2 For each 35 mm cell-culture Petri dish (5 in total), transfer the 0.25 mg/ml poly-lysine solution into another 35 mm cell-culture Petri dish. Incubate at room temperature for 5 min.

1.6.3 Discard the 0.25 mg/ml poly-lysine solution.

1.6.4 Let the Petri dishes dry, upside-down at room temperature for 2 hr.

1.6.5 Store the poly-lysine-coated dishes wrapped in a plastic wrap at 4 °C to avoid contamination for one week maximum.

1.7 Agarose-coated dishes

Note: They are used to culture injected embryos. The agarose provides a cushion for the injected embryos and prevents them from sticking to the bottom of the dish.

1.7.1 Make artificial seawater (ASW) using 37-38 g/l commercial salts + 0.25 mM $NaHCO_3$ in reverse osmosis water.

1.7.2 Dissolve agarose to a 1% concentration in 0.22 µm-filtered ASW by heating the solution in a microwave.

1.7.3 Swiftly pour the warm agarose solution from one 35 mm Petri dish into another one in order to ensure a very thin agarose coating of the dish.

1.7.4 Store the agarose-coated dishes wrapped in Saran wrap at 4 °C to avoid contamination for one week maximum.

1.8 Injection mix and loading of the injection needles

1.8.1. About 2 hr before starting the injections, make a 2 µl injection mix in RNase- and DNase-free water with final concentrations of 1-1.8 µg/µl of mRNA, 15% glycerol, 1.25% Phenol Red.

Note: The Phenol Red colors the solution, which allows monitoring of the injection efficiency and the identification of successfully injected embryos. Glycerol favors mRNA diffusion within the oocyte.

1.8.2 Centrifuge 4 min at 18000 x g to pellet crystals. Keep on ice until use.

1.8.3 With a 10 µl pipette, collect 0.5 µl of the injection mix, avoiding the bottom of the tube, where the crystals have been pelleted.

1.8.4 Backfill at least two injection needles (in case one breaks during the injection) by pipetting the 0.5 µl drop of injection mix at the large opening of the needle.

1.8.5 Install the needles in a storage jar with liquid at the bottom at 4 °C to prevent evaporation of the injection mix. Let the injection mix slowly travel to the tip of the injection needle for at least 1 hr to minimize the creation of bubbles.

1.8.6 Store additional injection mix at -80 °C for maximum of three additional uses, after which the mRNA quality deteriorates (data not shown).

2. Collection of biological material, microinjection and embryo culture

2.1 Oocyte and sperm collection

Note: See Theodosiou[148] for a detailed protocol for inducing spawning and for gamete collection.

2.1.1 Shock males and females in ASW at 23 °C for 24 hr.

2.1.2 One to two hours before sunset, transfer the adults into individual cups in ASW at 19 °C because most adults will spawn 1-2 hr after sunset.

2.1.3 Rinse 35 mm Petri dishes in filtered ASW and let them dry upside-down to prevent the oocytes from sticking to the bottom of the dish.

2.1.4 Upon spawning, immediately collect sperm and oocytes with a 1000 µl pipette.

2.1.4.1 Keep sperm and oocytes apart from adult amphioxus because the contact is detrimental for gamete health (data not shown).

2.1.4.2 Furthermore, avoid startling the adults, which leads to movements that dissipate, and hence dilute, both sperm and oocytes. To keep the sperm active as long as possible and to optimize fertilization rate, collect the sperm as concentrated as possible.

2.1.5 Keep sperm on ice in a 1.5 ml tube.

2.1.6 Transfer oocytes in filtered ASW into the pre-rinsed 35 mm Petri dishes.

2.1.7 Transfer 100-500 oocytes with a previously pulled 300-400 µm transfer Pasteur pipette to another 35 mm Petri dish to perform the injections.

2.1.8 Fertilize the remainder of the clutch as a control for sperm and oocyte quality or for other experiments.

2.2 Oocyte injection

2.2.1 Install the injection needle on the micromanipulator at a 50° angle relative to the horizontal plane.

Note: Angles of less than 50° will push the oocytes around on the dish, while angles of more than 50° will not allow an appropriate monitoring of the needle position relative to the oocyte.

2.2.2 Under a fluorescent dissecting scope with 25x oculars, transfer 30 oocytes with the 300-400 μm transfer Pasteur pipette on a poly-lysine-coated dish containing filtered ASW.

2.2.3 Deposit the oocytes along a line to carry out injections in an ordered way and to distinguish injected from non-injected oocytes. Inject small numbers (30 oocytes) to minimize the exposure time of oocytes to poly-lysine, which tends to deform developing embryos (data not shown).

2.2.4 Use the dark field illumination to render the oocytes as translucent as possible.

2.2.5 With the coarse movement knob of the micromanipulator, bring the injection needle close to an oocyte.

2.2.6 With fine forceps, cut open the needle at the level where the tip starts to be curved. By pulsing with the injector, verify that red injection mix is actually flowing out of the needle.

2.2.7 At 200x magnification and with the fine movement knob of the micromanipulator, gently move the injection needle inside the core of the oocyte.

Note: If inserted too superficially, the injected solution will not remain inside the oocyte. If inserted too far, the oocyte will be destroyed.

2.2.8 Inject with 1-3 pulses of 120 msec duration and 1-10 psi pressure. If the needle is fine enough, inject with continuous flow at constant pressure. Ensure that the injection volume corresponds to 1/5 to 1/3 of the volume of a single oocyte.

2.2.9 Following injection, pull the needle out swiftly to avoid leakage of the oocyte.

2.2.10 Verify that the injected solution remains within the oocyte and that after a few seconds, the injected solution spreads throughout the oocyte.

2.2.11 Move on to the next oocyte in line.

2.2.12 Keep some uninjected embryos of each series as negative control to estimate the background fluorescence when scanning for injected embryos.

2.3 Fertilization, selection of injected embryos and embryo culture

2.3.1 Fertilize the oocytes as soon as a series has been injected. As oocyte quality declines with time, inject and fertilize oocytes within 1 hr after spawning[174].

2.3.2 Depending on the sperm concentration, add 1-5 drops of sperm to the oocytes and swirl the dish.

Note: The fertilization envelope should become apparent on the embryos after about 1 min.

2.3.3 Allow the embryos to detach from the poly-lysine-coated dish, while injecting another series of oocytes.

2.3.4 Transfer the embryos with the 600 µm transfer Pasteur pipette into an agarose-coated Petri dish. Remove the embryos from the poly-lysine-coated dish as soon as possible, if at all possible before the 2-cell stage.

Note: In case of prolonged exposure to poly-lysine, the embryos tend to become densely-packed blastulae, flattened on the side touching the bottom of the dish.

2.3.5 At the 2-cell to 4-cell stage, select with a fluorescent dissecting scope with DSR filter the successfully-injected embryos, *i.e.,* those with a normal morphology that exhibit a Phenol Red-derived red fluorescent signal.

2.3.6 Keep the embryos in culture in filtered ASW in agarose-coated Petri dishes at 19 °C until the desired stage for *in vivo* imaging.

Representative results

The protocol detailed above provides the basis for the microinjection of *B. lanceolatum* oocytes and hence for the introduction into developing *B. lanceolatum* embryos of mRNA encoding fluorescent proteins for *in vivo* imaging. Although the technique is certainly robust and reliable, the rate of successful injections using this protocol remains variable (R1 – Table 1). The very likely explanation for this intriguing fact is the extreme variability of oocyte clutches: different egg batches do indeed behave very differently, when subjected to the injection pressure. Some oocytes are rather elastic and tend to flatten at the bottom of the dish, while others are quite stiff and remain round upon injection. Furthermore, some oocyte batches tend to inflate their chorion membranes upon injection, while others simply lyse (data not shown). No obvious correlation could be established between oocyte behavior upon injection and embryonic development after fertilization. It is thus difficult to anticipate which category of oocytes is best suited for microinjection. However after fertilization, embryos undergoing normal development can be identified as early as cleavage stages: 2-cell to 4-cell stage embryos can be considered normal when the contact surface between blastomeres is small, while a compacted cleavage-stage embryo, where individual cells are difficult to discern, is definitely abnormal.

In our hands, about half of the oocytes do not survive the trauma caused by the injection and about half of the embryos that do survive the injection exhibit a specific fluorescent label. Thus, we estimate that with this injection protocol, between 50 and 120 embryos can be successfully injected on a given spawning day yielding between 15 and 60 labeled embryos.

Experiment number	Injected construct mRNA	Number of injected embryos	Number of labeled embryos
1	H2B:eGFP	53	23
2	H2B:eGFP	50	31
3	H2B:eGFP	48	20
4	H2B:eGFP	54	24
5	H2B:eGFP	114	47
6	H2B:eGFP	80	46
7	H2B:eGFP or H2B:mCherry + eGFP:CAAX box	62	24
8	H2B:eGFP	87	60
9	H2B:eGFP	77	45
10	eGFP:CAAX box or H2B:mCherry + GFP:CAAX box	110	51
11	H2B:mCherry + eGFP:CAAX box	45	26
12	H2B:mCherry + eGFP:CAAX box	52	16
13	eGFP:CAAX box or H2B:mCherry + GFP:CAAX box	74	35
	Total	**906**	**448**
		Success rate	**49%**

R1 – Table 1: Injection success rates*. The injected constructs, the number of injected embryos that survived the injection and the number of labeled (or successfully injected) embryos are indicated, as is the overall percentage of successful injections.*

The red fluorescence of the Phenol Red in 2-cell to 4-cell stage embryos very reliably marks embryos that have been properly injected, as 100% of embryos selected in this manner subsequently produce the fluorescent proteins encoded by the injected mRNA. Additionally, there is a very clear positive correlation between the intensity of Phenol Red fluorescence at

the 2-cell to 4-cell stage and the time of onset of the fluorescence produced by the proteins encoded by the injected mRNA. This correlation thus allows the identification of those embryos that have received the highest quantity of mRNA during the injection process.

If the signal of the fluorescent protein encoded by the injected mRNA is not detected after selection of Phenol Red-positive embryos, we recommend to run the injection mix on an RNAse-free gel in order to check for mRNA degradation or trapping (Fig. R1.3). In lane 1, an intact mRNA band is detected. Injection of this mix led to strong fluorescent signal in the embryo. On the contrary, in lane 2 which corresponds to another experiment where Phenol Red was replaced with Dextran dye in the mix (see discussion for further details) the mRNA seems to have been trapped by the Dextran dye and injection of this mix never led to fluorescent signal in the embryo.

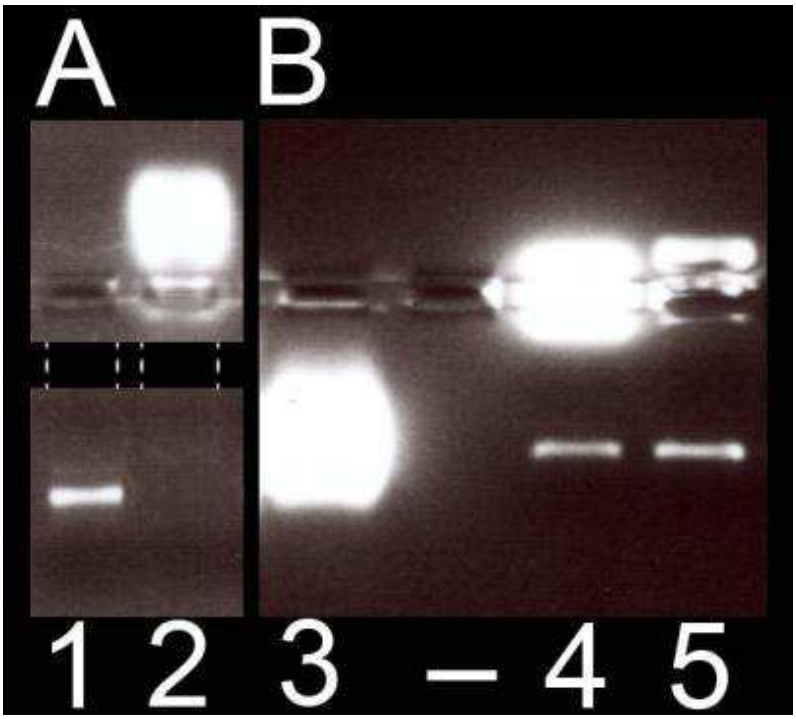


Fig. R1.3 – Interactions of mRNA and selected fluorescent dyes. *(A) Result of the migration of a mix containing mRNA and Phenol Red in lane 1 and of mRNA and Texas Red dextran (at a final concentration of 3.3 mg/ml) in lane 2. (B) Result of the migration of a mix containing mRNA and Oregon Green dextran (at a final concentration of 3.3 mg/ml) in lane 3, of mRNA and FITC dextran (at a final concentration of 3.3 mg/ml) in lane 4 and of mRNA and Rhodamine dextran (at a final concentration of 3.3 mg/ml) in lane 5. Gel migration time in (A) and (B) was different. For the sake of clarity, the black rectangle in (A) masks a light reflection and the dashed lines delineate the migration lanes.*

Using this microinjection technique and our constructs (R1 – Tables 2, 3) (see discussion), we can thus reproducibly produce homogenous fluorescent labeling throughout the embryo with mCherry or eGFP in the nucleus and with eGFP at the membrane (Fig. R1.4). The imaging protocol used to generate Figure 4 will be described elsewhere. Depending on the amount of mRNA injected, fluorescent protein expression typically becomes detectable between 16-cell (Fig. R1.4 A) and 64-cell stages and stays detectable at least up to the late neurula stage (data not shown). Later stages have not been tested. Nuclear signal typically appears earlier than membrane signal, potentially because of the intrinsically diffuse nature of the membrane compared to the compact nature of the nucleus (data not shown). Although it has not been monitored systematically, embryos injected with both a nuclear and a membrane label appear to be less healthy than embryos injected exclusively with a nuclear eGFP label. Intriguingly, this effect seems to be independent of the total amount of mRNA injected as the total amount of injected mRNA is the same, whether one or two mRNA species are included to the mix (data not shown).

Construct	Position in the pCS2+ vector	Original sequence	Mutated sequence
pCS2+ eGFP:CAAX box	79..87	GGA TCC ACC	ACC GTC AAC
pCS2+ H2B:eGFP	82..90	TCC GAC ACG	ACC GTC AAC
pCS2+ H2B:mCherry	82..90	TCC GAC ACG	ACC GTC AAC

R1 – Table 2: Tentative optimization of Kozak sequences. *Original and mutated Kozak sequences are indicated for each construct. The introduced mutations recover the Kozak sequence of the B. lanceolatum actin gene, which very closely resembles that of the inferred theoretical preferred Kozak sequence of amphioxus (A/CGA/C G/TTC A/GAC atg TG/CT).*

Discussion

In this article, we present, for the first time, a detailed and reproducible protocol for the injection of *B. lanceolatum* oocytes, which, after *B. floridae*[174,475] and *B. belcheri*[476], is thus the third amphioxus species, for which such a technique has been described. Importantly, the protocol described here also includes the description of vector systems suited for the production of fluorescent proteins in *B. lanceolatum* from injected mRNA produced *in vitro*

(described below). Together, these new tools allow *in vivo* imaging of the early development of amphioxus and the analysis of dynamic cell behaviors that underlie the earliest morphogenetic events in the embryo, such as cleavage and gastrulation.

pCS2+ eGFP:CAAX box

Original codon (usage level)	**Mutated codon (usage level)**	**Position in the pCS2+ vector**	
GTA (0.08)	GTG (0.43)	154..156	eGFP sequence
AGA (0.09)	AGG (0.27)	814..816	Linker

pCS2+ H2B:eGFP

Original codon (usage level)	**Mutated codon (usage level)**	**Position in the pCS2+ vector**	
GTA (0.08)	GTG (0.43)	223..225	H2B sequence
		289..291	H2B sequence
		469..471	Linker
		568..570	eGFP sequence
CTA (0.06)	CTG (0.51)	226..228	H2B sequence

pCS2+ H2B:mCherry

Original codon (usage level)	**Mutated codon (usage level)**	**Position in the pCS2+ vector**	
GTA (0.08)	GTG (0.43)	223..225	H2B sequence
		289..291	H2B sequence
		469..471	Linker
		913..915	mCherry sequence
CTA (0.06)	CTG (0.51)	226..228	H2B sequence

R1 – Table 3: Tentative optimization of codon usage. Original and mutated codons are indicated for each construct. While this has not been experimentally tested, the introduced mutations are meant to optimize mRNA translation in amphioxus.

In general, the microinjection technique has the benefit of allowing the delivery, into a target cell, of a pre-defined volume of a specific compound. With this being said, one major limitation of this technique is the limited number of cells that can be injected during a given experiment. For example, the protocol described here for *B. lanceolatum* allows the injection of 100 to 120 eggs per injection session, of which about half will be labeled (Table 3). For *B. floridae*, the numbers are very similar with 500 or more eggs that can be injected per day, of which more than 50% will survive the injection[174,475]. While the protocols for injecting *B. lanceolatum* and *B. floridae* eggs closely resemble each other, the *B. belcheri* technique is slightly different: the eggs are placed on poly-lysine-coated coverslips and injections are performed under an inverted microscope using a microinjector of a different brand. This approach apparently allows the injection of between 200 and 300 *B. belcheri* eggs per injection session with a survival rate, after hatching at the neurula stage, of 89.39% to 95.83%[476]. Considering the differences in the protocols, it is difficult to know whether the differences in success rates are due to species-related or protocol-related differences. One would have to test the different species in parallel with the different protocols to answer this question. Nonetheless, to further increase the number of oocytes targeted for the delivery of exogenous compounds, other protocols will have to be developed for amphioxus. Candidate techniques include electroporation, bombardment with microparticles, lipofection, and transduction[473]. Of note, electroporation has already been established as a standard technique for the introduction of exogenous material into fertilized ascidian tunicate eggs, another invertebrate chordate model used for studying the evolution of developmental mechanisms[93].

For successful microinjection and subsequent *in vivo* imaging of fluorescent proteins, suitable expression vectors are a mandatory prerequisite. Towards this end, three plasmids have been developed and experimentally validated (Fig. R1.1): for membrane targeting, the eGFP gene was fused at its 3' end to the human HRAS CAAX box (resulting in a eGFP:CAAX box construct)[477] and, for nuclear targeting, both the eGFP and the mCherry genes are fused at their 5' ends to the zebrafish histone 2B (H2B) exon (resulting, respectively, in a H2B:eGFP and a H2B:mCherry construct)[478]. Each one of these constructs has been cloned into the pCS2+ plasmid containing a SV40 late polyadenylation site and a SP6 promoter allowing *in vitro* capped mRNA synthesis[479,480]. Furthermore, with the goal to optimize translation of the constructs in *B. lanceolatum*, both the Kozak sequences and the codons have been adapted using single nucleotide mutagenesis (R1 – Tables 2, 3). Based on the approach by Nakagawa[481], a small pool of *B. lanceolatum* genes was analyzed to identify a preferred Kozak sequence in this species. The closest known naturally-occurring sequence

to this theoretical preferred sequence is that of the *B. lanceolatum actin* gene. The Kozak sequences of the three constructs was hence modified to match this *actin* gene sequence. Furthermore, the codon usage of *B. lanceolatum* was analyzed using the codon usage database[482] and codons in the constructs with a low usage probability in *B. lanceolatum* (< 10%) were replaced, wherever possible, by equivalent codons with a higher usage probability.

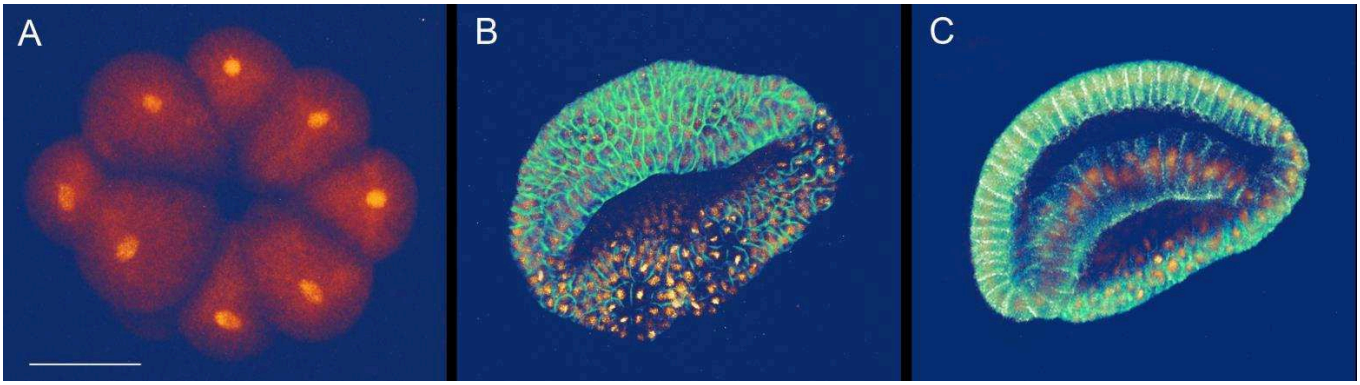


Fig. R1.4 – 3D rendering (Amira software) of amphioxus embryos injected with mRNAs coding for fluorescent proteins. *Images are extracted from 3D time-lapses acquired on an optimized Leica SP5 2-photon laser-scanning microscope. (A) 16-cell stage embryo expressing nuclear H2B: eGFP. (B) Gastrula stage embryo expressing nuclear H2B: mCherry and the membrane-targeted eGFP: CAAX box. (C) Gastrula stage embryo expressing nuclear H2B: mCherry and the membrane-targeted eGFP: CAAX box. The optical section shows internal cells with nuclear mCherry and membrane-associated eGFP signals. Scale bar = 50 µm.*

The results of the injections show that the level of protein expression from these vectors is suitable for *in vivo* imaging analyses. Given that the expression output of the modified vectors has not been compared with that of unmodified constructs, we can not conclude on the efficacy of the modifications that were introduced in the Kozak consensus and the coding sequences. It would certainly be interesting to test whether these modifications indeed have an impact on protein expression levels. Along these lines, a recent analysis based on microinjections into *B. belcheri* suggests that other, unmodified vectors can also be used for the synthesis of mRNA for fluorescent protein production in amphioxus[476].

Amphioxus contains endogenous green fluorescent proteins[483]. The embryos therefore exhibit low levels of green as well as red fluorescence (our unpublished observations). However, these endogenous levels are negligible with respect to the fluorescence levels

resulting from the injection of our mRNA constructs. Therefore, the endogenous fluorescence of amphioxus embryos does not disturb experimental observations using fluorescent binoculars or the multi-laser scanning microscopes. In addition, the injection experiments reveal that the intensity of the exogenous fluorescence generated by mRNA injection depends on the actual amount of injected material, which is directly correlated with the final concentration of mRNA used in the injection mix. *B. lanceolatum* embryos tend to tolerate a concentration of up to 1.8 µg/µl mRNA, although, independent of the actual mRNA concentration, some mRNA combinations seem to be less tolerated by *B. lanceolatum* embryos than others. Thus, the combination of nuclear mCherry and membrane eGFP mRNAs appeared to be more toxic than the injection of nuclear eGFP alone.

Texas Red dextran has previously been used as a tracer for successful injections into amphioxus oocytes[174,475,476]. Intriguingly, a fluorescent signal derived from the injected mRNA was never obtained after injection of solutions containing Texas Red dextran and this in both amphioxus oocytes (n=4 experiments) and zebrafish embryos (n=3 experiments). When an injection mix containing *in vitro* transcribed mRNA and Texas Red dextran is loaded on a RNase-free agarose gel (Fig. R1.3, lane 2), the band corresponding to the mRNA is absent. In contrast, *in vitro* transcribed mRNA is detectable on an RNase-free agarose gel in the presence of Phenol Red (Fig. R1.3, lane 1). Given that there is no evidence for mRNA degradation on the gel, these results suggest that Texas Red dextran tends to trap mRNA. When used in an injection solution, Texas Red dextran might thus prevent the translation of the injected mRNA. Following this observation, the interactions of other dyes (FITC dextran, Rhodamine dextran and Oregon Green dextran) with mRNA were tested, both on RNase-free agarose gels and by injection into *B. lanceolatum* or zebrafish (Fig. R1.3). The analyses indicate that FITC dextran does not trap mRNA on gel (Fig. R1.3, lane 4) and leads to fluorescent protein expression in zebrafish (n=1 experiment), but not in embryos of *B. lanceolatum* (n=1 experiment) (data not shown). Rhodamine dextran does not trap mRNA on gel (Fig. R1.3, lane 5) and leads to fluorescent protein expression in zebrafish embryos (n=1 experiment) (data not shown), but its activity could unfortunately not be tested in *B. lanceolatum* embryos. Finally, as the Oregon Green dextran migrates at the same level as the mRNA, mRNA trapping could not be assessed by agarose gel (Fig. R1.3, lane 3). This latter dye prevented fluorescent protein expression in amphioxus (n=1 experiment), but not zebrafish embryos (n=1 experiment) (data not shown). In sum, these data suggest that some dextrans can efficiently inhibit the translation of injected mRNA. Given that dose-response experiments were not carried out, these data are qualitative. Intriguingly, this inhibitory effect

seems to be dependent on the animal species (zebrafish vs amphioxus), which highlights the need for further studies to understand the mechanisms underlying this effect. Furthermore, these results call for preparatory analyses, if dextrans are to be used as color tracers for injections.

The development of the microinjection technique for a given model opens the door for gene-specific manipulations. In amphioxus, for example, the first description of microinjections (in *B. floridae*) was followed by the successful knockdown of a gene, *hox1*, using morpholino oligonucleotide injections[427,430]. The number of *B. lanceolatum* embryos that can successfully be injected every day using the protocol presented here is certainly compatible with this type of study. Furthermore, with these methods and the novel constructs at hand, one can now also design and perform overexpression and knockdown studies by injection of mRNAs of interest (native or dominant-negative forms) or use other strategies for disrupting gene expression (for example microRNA- or shRNA-based strategies). So far, amphioxus injections can only be performed at the oocyte stage, simply because this is the only stage that can efficiently be immobilized. After injection at the oocyte stage, the molecule of interest is ubiquitously expressed in the embryo. Thus, if gene expression needs to be altered or monitored only in a subset of cells, DNA constructs need to be injected, as these are expressed mosaically in the developing amphioxus embryo[123,476,484,485]. DNA construct injections can also be used to test the activity of regulatory regions during amphioxus development in transient transgenic assays, with the aforementioned drawback of mosaic expression[123,476,484,485].

Ultimately, the amphioxus microinjection technique also opens the door for stable transgenesis, including the targeted knock-out or knock-in of specific genetic loci. Systems for stable insertion of exogenous DNA already exist, for example with the fish transposon-based Tol2 system that seems to function in both insects and amniote vertebrates[486]. Furthermore, approaches based on modified zinc-finger nucleases (ZFNs) or transcription activator-like effector nucleases (TALENs) or RNA-guided genome editing tools (CRISPR/Cas system) could be used to generate point mutations and knock-in modifications in specific regions of the genome[487]. These efforts will require the year-round breeding of amphioxus in captivity, which has already been set up for *B. belcheri*[198,225] and *B. japonicum*[223,225] and is currently being developed for *A. lucayanum*, *B. floridae* and the amphioxus species featured here, *B. lanceolatum*[219].

R2: Lineage-specific duplication of amphioxus retinoic acid degrading enzymes (CYP26) resulted in sub-functionalization of patterning and homeostatic roles[3]

Abstract

During embryogenesis, tight regulation of retinoic acid (RA) availability is fundamental for normal development. In parallel to RA synthesis, a negative feedback loop controlled by RA catabolizing enzymes of the cytochrome P450 subfamily 26 (CYP26) is crucial. In vertebrates, the functions of the three CYP26 enzymes (CYP26A1, CYP26B1, and CYP26C1) have been well characterized. By contrast, outside vertebrates, little is known about CYP26 complements and their biological roles. In an effort to characterize the evolutionary diversification of RA catabolism, we studied the *CYP26* genes of the cephalochordate amphioxus (*Branchiostoma lanceolatum*), a basal chordate with a vertebrate-like genome that has not undergone the massive, large-scale duplications of vertebrates.

In the present study, we found that amphioxus also possess three *CYP26* genes (*CYP26-1*, *CYP26*-2, and *CYP26-3*) that are clustered in the genome and originated by lineage-specific duplication. The amphioxus *CYP26* cluster thus represents a useful model to assess adaptive evolutionary changes of the RA signaling system following gene duplication. The characterization of amphioxus *CYP26* expression, function, and regulation by RA signaling demonstrated that, despite the independent origins of *CYP26* duplicates in amphioxus and vertebrates, they convergently assume two main roles during development: RA-dependent patterning and protection against fluctuations of RA levels. In amphioxus, our analysis suggests that, while RA-dependent patterning is sustained by CYP26-2, RA homeostasis is mediated by CYP26-1 and CYP26-3. Furthermore, comparisons of the regulatory regions of *CYP26* genes of different bilaterian animals indicated that a CYP26-driven negative feedback system was present at least in the last common ancestor of deuterostomes, but not in that of bilaterians.

[3] Article in revision at BMC Evolutionary Biology. Authorship: João E. Carvalho, Maria Theodosiou, Jie Chen, Pascale Chevret, Susana Alvarez, Angel R. de Lera, Vincent Laudet, Jenifer C. Croce and Michael Schubert.

Altogether, this work reveals the evolutionary origins of the RA-dependent regulation of *CYP26* genes and highlights convergent functions for CYP26 enzymes that originated by independent duplication events, hence establishing a novel selective mechanism for the genomic retention of gene duplicates.

Background

In all animals, the vitamin A-derived morphogen retinoic acid (RA) is involved in a number of crucial developmental processes, including, for example, early embryonic patterning and organogenesis acting on different cellular processes ranging from proliferation to cell death[232,233,235,238,239,241,402]. In vertebrates, normal development requires a very tightly controlled balance of the total amount of available RA, which is maintained through positive and negative feedback loops associated, respectively, with RA production (chiefly by RALDH1, 2, and 3, for retinaldehyde dehydrogenase 1, 2, and 3) and RA degradation (chiefly by CYP26A1, B1, and C1, for cytochrome P450 subfamily 26A1, B1, and C1)[488–492]. The biological response to endogenous RA, in turn, is mediated by heterodimers of two nuclear receptors, the retinoic acid receptor (RAR) and the retinoid X receptor (RXR), with the expression levels of RAR in particular being tightly linked to the availability of RA[233,235,238]. RAR/RXR heterodimers directly exert their transcriptional function by binding to RA response elements (RAREs) in the regulatory region of DNA of RA target genes[493]. A typical RARE is composed of two direct repeats (DRs) corresponding to a conserved nucleotide sequence [(A/G)G(G/T)TCA] separated by a spacer composed of one, two or five nucleotides (corresponding to, respectively, DR1, DR2, and DR5 elements)[311,493,494]. Upon RA binding, RAR/RXR heterodimers generally function as ligand-activated transcription factors, but can also mediate RA-dependent repression of target genes in a context-specific manner, the exact molecular modalities of which still remain to be established[403].

During vertebrate development, the RA degrading enzymes of the CYP26 subfamily play critical roles in the formation of an anterior-posterior (A-P) RA gradient as well as in the compensation of RA level fluctuations by oxidizing RA into biologically inactive compounds[495]. They are thus characterized by dynamic, yet highly specific, developmental expression patterns in vertebrates[496], with *CYP26A1*, for example, being expressed in the anterior ectoderm in the early embryo and subsequently becoming localized, amongst other tissues, to the hindbrain, the pharyngeal arches, and the tail bud. Similarly, both *CYP26B1*

and *CYP26C1* are detectable in specific rhombomeres of the hindbrain and in pharyngeal arches as well as in fin and limb buds of the developing embryo[496]. Concomitantly, the loss of CYP26 function has been associated both with A-P patterning defects, most prominently in the developing central nervous system (CNS) and the mesoderm, and an increased sensitivity to RA teratogenicity[345]. For instance, *CYP26A1* knockout mice are characterized by a posteriorization of the hindbrain and the vertebral column, and *CYP26B1* genetic ablation leads to craniofacial and limb malformations[293]. Interestingly, while the loss of *CYP26C1* alone does not result in overt anatomical abnormalities[295], the combined removal of *CYP26C1* with either *CYP26A1* or *CYP26B1* induces phenotypes that are more severe than those aforementioned, thereby suggesting that *CYP26C1* plays an important cooperative role in the CYP26-mediated control of endogenous RA levels during vertebrate development[295,345].

In line with this cooperative action of CYP26 enzymes, the vertebrate RA signaling system in general is characterized by complex feedback mechanisms that are mediated, either indirectly or directly, by RAR/RXR-dependent signaling. As an example of an indirect regulation, it has been shown that, in the vertebrate trunk, RA, generated by RALDH activity, represses and confines *FGF8* expression to rostral and caudal domains (i.e. to the heart- and tail bud-associated progenitor fields)[403]. This action is mediated by RAR/RXR heterodimers binding to a repressive DR2 RARE located upstream of the *FGF8* gene[381,497] that, in turn, activates *CYP26* expression both anteriorly and posteriorly to limit the extent of RA activity[403,498,499]. In addition, the expression of *CYP26A1* and *CYP26B1* have been shown to be dependent on RA activity, thereby generating a *CYP26*-controlled negative feedback loop in RA sensitive tissues to reduce the overall amount of available RA[489,496,500,501]. For *CYP26A1*, this regulation is directly mediated by RAR/RXR heterodimers binding to DR5 RAREs in the promoter region, while for *CYP26B1* this control seems to be indirect[500,501]. Note further that the vertebrate *CYP26C1* gene is likely to contribute differently than its paralogs to this negative feedback system, as *CYP26C1* expression is actually downregulated following RA stimulation[489].

The intricate molecular mechanisms controlling the catabolism of endogenous RA during vertebrate development likely arose at the base of this lineage following the whole genome duplication (WGD) events that took place during early vertebrate diversification[127,419]. Therefore, the evolutionary elaboration of the RA signaling system in general seems to be tightly linked to the duplication of RA metabolism genes. The so-called DDC model (for Duplication-Degeneration-Complementation) predicts three possible outcomes following duplication of a gene: non-functionalization (i.e. the loss of one of the

duplicates), neo-functionalization (i.e. one of the copies retains the ancestral role, while the other duplicate assumes a novel functionality) or sub-functionalization (i.e. both duplicates assume a part of the function of the single ancestral gene)[35,502]. While the model predicts that the most likely outcome following duplication of a gene is the loss of one of the duplicates (i.e. non-functionalization), very clear examples for the neo-functionalization and the sub-functionalization of duplicated genes remain scarce[36,503].

In order to develop a credible scenario for the evolutionary diversification of the vertebrate RA system and investigate the implications of the DDC model in the duplication of RA metabolism genes, we decided to study the function and regulation of RA degradation during embryonic development of the cephalochordate amphioxus (*Branchiostoma lanceolatum*). Due to its phylogenetic position at the base of chordates, amphioxus is a very useful model to characterize chordate- and vertebrate-specific innovations, both on a morphological and a genomic level. For instance, at the morphological level, amphioxus and vertebrates share a dorsal CNS, a postanal tail as well as pharyngeal gill slits[97,504], while, conversely, amphioxus lacks some vertebrate-specific characters, such as definitive neural crest and placodes as well as a cartilaginous or bony skeleton[97,504]. Furthermore, amphioxus is a basal chordate that did not undergo WGD[31,123] and that possesses a vertebrate-like RA signaling pathway[123,127]. Thus, while RA signaling in vertebrates is generally controlled by three RARs (RARα, RARβ, RARγ) and three RXRs (RXRα, RXRβ, RXRγ) that form a multitude of different heterodimers, the amphioxus genome contains only one *RAR* and one *RXR* gene[418]. Nevertheless, administration of exogenous RA during amphioxus gastrulation leads, as observed in vertebrates, to the posteriorization of the amphioxus CNS and endoderm, hence preventing, for example, the formation of mouth and gill slits[418,426–428,430,467]. These regionalization defects are further associated with a deregulation of *RAR* and *Hox* gene expression, which have been shown to be direct targets of RA signaling in amphioxus, as they are in vertebrates[432].

In amphioxus, three *CYP26* genes (*CYP26-1*, *CYP26-2*, and *CYP26-3*) have been reported, which are clustered together in the genome and have possibly emerged from a lineage-specific duplication[127]. This *CYP26* locus offers a rare, if not unique, opportunity to investigate the adaptive changes following lineage-specific duplication that led to the retention of three *CYP26* genes in the genome. The results from our analyses thus show that the three amphioxus *CYP26* genes arose by lineage-specific tandem duplication of a single, ancestral *CYP26* gene. They further provide evidence that these three genes assume two main

functions during amphioxus development, as they do in vertebrates, i.e. patterning of the embryo and protection against RA level fluctuations. These two roles have been sub-functionalized in amphioxus with *CYP26-2* mediating RA-dependent developmental patterning and *CYP26-1* and *CYP26-3* assuming the protection of the embryo from RA teratogenesis. Moreover, the presence of functional RAREs in the amphioxus *CYP26* cluster indicates that RA degradation is regulated in cephalochordates like in vertebrates, i.e. directly by RAR/RXR heterodimers, hence establishing a negative RA feedback system. Comparative genomic analyses of *CYP26* regulatory regions from different bilaterian animals further revealed that this CYP26-dependent negative RA feedback system is likely not unique to chordates, but probably arose earlier in animal evolution and was already present in the last common ancestor of all deuterostomes, but not in that of all bilaterians. The adaptive advantages of an elaborate CYP26-driven RA degradation system are discussed. In sum, the evolutionary history of amphioxus *CYP26* genes provides an excellent example for the sub-functionalization of two distinct developmental functions and a paradigm for understanding the selective mechanisms acting on duplicated genes and leading to their retention in the genome.

Results

CYP26 genes were duplicated independently several times in bilaterian evolution

Previous analyses have reported three *CYP26* genes in the Florida amphioxus, *Branchiostoma floridae*, and have suggested that they likely originated by lineage-specific duplication from a single ancestral *CYP26* gene[123,127]. Here, we have identified and cloned three *CYP26* genes from the European amphioxus, *Branchiostoma lanceolatum*. To further assess the phylogenetic relationships of the amphioxus *CYP26* genes relative to each other and to other members of the *CYP26* subfamily, thereby distinguishing between orthologous and paralogous *CYP26* genes, we first carried out phylogenetic analyses using as outgroup the *CYP51* genes, which constitute the *CYP* subfamily that is most closely related to the *CYP26* genes[451]. For this phylogenetic tree reconstruction, we used all *CYP26* sequences from fifteen vertebrates and thirteen invertebrates, including three cephalochordate species (*B. lanceolatum*, *B. floridae*, and *B. belcheri*). Of note, while we successfully identified genes encoding CYP26 in the genomes of priapulids, brachiopods, mollusks, annelids, sea urchins,

hemichordates, cephalochordates, ascidian tunicates, and vertebrates, we were unable to do so in those of nematodes and arthropods (as previously reported[402,461]).

Within vertebrates, we found three *CYP26* paralogs in both cyclostomes (*CYP26A1*, *CYP26B1/C1a*, and *CYP26B1/C1b*) and gnathostomes (*CYP26A1*, *CYP26B1*, and *CYP26C1*). Multiple *CYP26* paralogs were also identified in most invertebrates species studied (two in *Capitella teleta*, *Ciona intestinalis*, *Lottia gigantea, Priapulus caudatus*, and *Saccoglossus kowalevskii*, three in *B. lanceolatum*, *B. floridae*, *B. belcheri*, *Crassostrea gigas*, and *Ptychodera flava*, and four in *Lingula anatina*), with the notable exceptions of the cephalopod *Octopus bimaculoides* and the sea urchin *Strongylocentrotus purpuratus*, each of which possesses only a single *CYP26* gene (Appendix 2).

The results of the phylogenetic analysis (Fig. R2.1 and Appendix 2), obtained with both the Bayesian Inference (BI) and the Maximum Likelihood (ML) methods, suggested an early phylogenetic separation of the vertebrate *CYP26A1* sequences from the vertebrate *CYP26B1/C1* sequences. Vertebrate *CYP26A1* and *CYP26B1/C1* thus formed two independent clades within the *CYP26* subfamily, both of which being strongly supported: 0,91/96 (posterior probability/bootstrap percentage) for the *CYP26A1* clade and 1/99 for the *CYP26B1/C1* clade. Within these two vertebrate *CYP26* clades, the cyclostome sequences consistently branched at the base: *Lethenteron japonicum CYP26A1* at the base of the vertebrate *CYP26A1* (0,96/98) and *CYP26B1/C1a* and *CYP26B1/C1b* from *L. japonicum* and *Petromyzon marinus* at the base of the vertebrate *CYP26B1/C1* (0,7/58). Within the vertebrate *CYP26B1/C1* clade, the association of the gnathostome *CYP26B1* sequences (0,73/58) was less robustly supported than that of the gnathostome *CYP26C1* sequences (1/95), which might be related to the presence of chondrichthyan-specific *CYP26B1* duplicates (*CYP26B1* and *CYP26B2* from *Callorhinchus milii* and *Leucoraja erinacea*) disrupting the base of the *CYP26B1* branch. Of note, while our analysis revealed the presence of *CYP26A1*, *CYP26B1*, and *CYP26C1* paralogs in chondrichthyans in general, we were unable to identify a *CYP26A1* gene in *C. milii* and a *CYP26C1* gene in *L. erinacea*. Altogether, these data suggest that the diversification of vertebrate *CYP26* genes was a highly complex process, involving WGD, lineage-specific duplications as well as secondary gene losses.

Outside the vertebrates, the *CYP26* sequences from individual ecdysozoan, ambulacrarian and cephalochordate species always grouped together in the tree, with very strong support values: the ecdysozoan *P. caudatus* (1/100), the ambulacrarians *P. flava* and *S. kowalevskii* (1/99), and the cephalochordates *B. lanceolatum*, *B. floridae* and *B. belcheri*

(1/99). Collectively, these results suggest that, within each of these invertebrate groups, the *CYP26* gene complement originated independently by linage-specific duplication. In contrast, the two sequences from the tunicate *C. intestinalis* did neither associate with each other, nor reliably with one of the major *CYP26* clades in the tree. It is therefore impossible to comment on the nature and origin of the *CYP26* duplication in this animal. Similarly, the reconstruction of the evolutionary history of lophotrochozoan *CYP26* genes is complicated by the lack of phylogenetic resolution between the sequences from the five analyzed lophotrochozoan species, which formed an unresolved polytomy in our analysis. Nonetheless, there is evidence for lineage-specific duplications of *CYP26* genes in the brachiopod *L. anatina*, which possesses four *CYP26* paralogs that established two distinct clades in the tree, one very strongly (1/100) and one very weakly (0,8/--) supported. Furthermore, two of the three *CYP26* sequences from the oyster *C. gigas* are grouped within in a single clade, but the support for this association is very weak (0,95/--). Future studies will thus have to address the evolutionary processes underlying the origin of the diversity of lophotrochozoan *CYP26* genes.

To gain further insights into the evolutionary diversification of *CYP26* genes in different animal lineages, we next conducted a phylogenetic dating analysis (Appendix 2). This survey indicated that the *CYP26* genes of the ecdysozan *P. caudatus* were likely duplicated independently at the end of the Ordovician (about 446 Mya). Similarly, within the ambulacrarians, we found evidence for linage-specific duplications in hemichordates, which likely occurred during two different periods: the early Carboniferous for *S. kowalevskii* (about 342 Mya) and the middle Triassic for *P. flava* (about 240 Mya). Finally, in chordate lineages, *CYP26* genes have also likely been duplicated independently in cephalochordates, tunicates, and vertebrates, with the three cephalochordate genes resulting from an initial duplication in the early Carboniferous (about 358 Mya) followed by a subsequent duplication during the middle Permian (about 295 Mya). In contrast, while it is difficult to conclude on the timing of the duplication giving rise to the two *CYP26* genes in the ascidian *C. intestinalis*, the evolution of the vertebrate *CYP26* complement is complex and implies a series of duplications, including an ancient split into *CYP26A1* and *CYP26B1/C1* and the subsequent diversification of *CYP26A1*, *CYP26B1*, and *CYP26C1* during the Cambrian period (about 548 to 510 Mya). It should be added that, consistent with the results of the phylogenetic tree, the dating analysis did not yield reliable information on the timing of the duplications of lophotrochozoan *CYP26* genes.

CYP26-2 expression is suggestive of a function in developmental patterning of the* Branchiostoma lanceolatum *embryo

In vertebrates, *CYP26A1, CYP26B1*, and *CYP26C1* have very distinct expression patterns with several key domains being conserved between different species[239,496]. Thus, we next assessed the temporal and spatial distribution of the three cephalochordate-specific duplicates in the amphioxus *B. lanceolatum* by *in situ* hybridization (ISH). The ISH results revealed that the expression profiles of *CYP26-1* and *CYP26-3* are generally quite similar (Fig. R2.2 A-E, T-X). For both genes, no signal was detectable by ISH from fertilization through mid gastrulation. Expression of both *CYP26-1* and *CYP26-3* is first identifiable at late gastrula stages as a weak signal in the lateral anterior mesoderm (Fig. R2.2 A, B, T, U). As development proceeds, this domain becomes associated with the most anterior somites at mid neurula stage (Fig. R2.2 C, D, V, W). At this stage, *CYP26-1* and *CYP26-3* are also discreetly and transiently expressed in the anterior central nervous system (CNS), at the level of the first somite (Fig. R2.2 C, V). Expression of both genes remains very weak and chiefly associated with mesodermal tissues during subsequent developmental stages (Fig. R2.2 E, X): while the *CYP26-1* signal is most evident in central and posterior regions of the larva (Fig. R2.2 E), *CYP26-3* is mainly detectable in central and more anterior larval territories (Fig. R2.2 X).

In contrast to *CYP26-1* and *CYP26-3*, *CYP26-2* has a much more complex developmental expression profile during *B. lanceolatum* embryonic and early larval development. Expression of *CYP26-2* is first detectable by ISH at the mid gastrula stage with the signal being localized globally around the blastopore (Fig. R2.2 F). The blastopore-associated signal subsequently weakens and, at the late gastrula stage an additional expression domain appears in a region corresponding to presumptive lateral mesoderm and anterior neuroectoderm (Fig. R2.2 G, H). By the mid neurula stage, the mesodermal signal has been expanded into the two anterior-most somite pairs (Fig. R2.2 I, J, L). At this stage, *CYP26-2* is further still detectable in the anterior neuroectodem (Fig. R2.2 I, J, L). Additionally, the gene is now expressed in the ectoderm, most conspicuously in the anterior and posterior tips of the embryo (Fig. R2.2 I, J, K-M), as well as in the anterior- and posterior-most endoderm (Fig. R2.2 I, J, M). At the mid neurula stage the blastopore-associated signal becomes perceivable in the newly formed tail bud (Fig. R2.2 F-J, M). In very early larvae, just before the opening of the larval mouth, expression of *CYP26-2* is detectable anteriorly in all germ layers, i.e. the ectoderm, the mesoderm, the endoderm as well as in the CNS, with the signal being least noticeable in the endoderm (Fig. R2.2 N-P). Furthermore, in both ectoderm and CNS,

individual cells are labeled along the A-P body axis (Fig. R2.2 N, Q) and, at the posterior end of the embryo, the tail ectoderm strongly expresses *CYP26-2* (Fig. R2.2 N, R). In the amphioxus larva, the overall domains of *CYP26-2* expression are maintained, with conspicuous labeling anteriorly and posteriorly and a weaker signal in the center (Fig. R2.2 S).

In sum, while *CYP26-1* and *CYP26-3* expression is very discreet and chiefly limited to the mesoderm, that of *CYP26-2* is detectable in all germ layers and dynamically changes in space and time throughout development.

CYP26 acts as a fine regulator of RA levels in the patterning of the B. lanceolatum larval tail fin

Disruption of CYP26 activity causes very severe defects during vertebrate development[496]. To determine the role of CYP26 enzymes in the amphioxus embryo, we subsequently disrupted endogenous RA degradation during amphioxus development by treatments with the CYP26-specific inhibitor R115866[505,506]. For comparisons, the R115866 treatments were carried out in parallel to treatments with RA or with two different RAR antagonists (BMS009 and BMS493). The capacity of R115866 to inhibit endogenous RA degradation was verified by double treatments of R115866 and BMS493. The results obtained from these different pharmacological treatments of *B. lanceolatum* embryos are in large agreement between each other and with previous studies that characterized the roles of RA signaling in amphioxus endoderm specification and pharyngeal patterning[418,426–428,430,467]. Thus, while the downregulation of RA signaling activity by RAR antagonists (1 µM of either BMS009 or BMS493) resulted in an enlarged pharynx and an expansion of pharyngeal structures (Fig. R2.3 E, F), the upregulation of endogenous RA signaling by 1 µM RA led to a shortening of the pharynx and the malformation of pharyngeal structures (such as the mouth and the gill slits) (Fig. R2.3 B). Consistently, the local upregulation of RA signaling by 0,5 µM R115866 yielded similar results (Fig. R2.3 C), and co-treatments of 0,5 µM R115866 and 1 µM BMS493 led to an attenuation of the severe phenotype induced by 0,5 µM R115866 alone with at least a partial recovery of pharynx formation and patterning (Fig. R2.3 D).

Importantly, these pharmacology-based experiments revealed a previously undescribed role of RA signaling in developmental patterning of the amphioxus larval tail fin, a finding that is consistent with localized expression of *CYP26-2* in the amphioxus tail fin ectoderm. At 60-hours of development, amphioxus larvae are characterized by an ectodermal

tail fin that is pointy in shape and on average 130,2 µm long (Fig. R2.3 A, G). When amphioxus embryos were treated with RA at the gastrula stage, the resulting tail fins of 60-hour larvae are round and significantly shorter. These effects were observable with exogenous treatments of both 0,1 µM and 1 µM RA (Fig. R2.3 B, G). Similar results were obtained with R115866 treatments at 0,1 µM and 0,5 µM (Fig. R2.3 C, G). Conversely, RAR antagonist treatments, with either BMS009 or BMS493, had the inverse effect: the tail fin becomes pointier in shape and is slightly elongated (Fig. R2.3 E-G). Co-treatment of 0,5 µM R115866 with 1 µM BMS493 led to a partial rescue of the tail fin phenotype with an almost normal shape and a slight reduction of the overall length (Fig. R2.3 D, G).

It has previously been shown that the amphioxus larval tail fin is composed of columnar epidermal cells that contain a large ciliary rootlet[507,508] and that RA signaling promotes tail regression in late, pre-metamorphic *B. floridae* larvae by downregulating the gene encoding the main component of the ciliary rootlet: the protein Rootletin[429]. However, the regulation of *Rootletin* expression by RA signaling in the tail fin of early *B. floridae* larvae has not yet been reported. Given the effects on early tail fin formation we observed in *B. lanceolatum* in response to the alteration of endogenous RA signaling levels, we decided to investigate the patterns of *Rootletin* expression in 60-hour *B. lanceolatum* larvae following the pharmacological treatment regimes detailed above. As previously described for late *B. floridae* larvae[429], *Rootletin* is expressed in the basal compartment of the columnar tail fin cells in 60-hour *B. lanceolatum* larvae (Fig. R2.3 H). At this developmental stage, the gene is further detectable in a small number of lateral ectodermal cells (Fig. R2.3 H).

While treatment with 0,1 µM RA resulted in a marked reduction of *Rootletin* expression concomitant with an apical compaction of the columnar tail fin cells (Fig. R2.3 I), 1 µM RA very strongly restricted the *Rootletin* expression domain (Fig. R2.3 J). The effect on tail fin development of either 0,1 µM or 0,5 µM of the CYP26 inhibitor R115866 was similar to that of exogenous RA and the R115866-treated larvae were thus generally characterized by a significant reduction of *Rootletin* expression (Fig. R2.3 K, L). The RAR antagonists BMS009 and BMS493 led to an apical expansion of *Rootletin* expression as well as to an increase of the overall length of the tail fin (Fig. R2.3 M, N). Furthermore, the RAR antagonist treatments induced *Rootletin* expression in additional lateral ectodermal cells not directly associated with the tail fin (Fig. R2.3 M, N). Intriguingly, while co-treatments of 0,5 µM R115866 and 1 µM BMS493 restore a shortened tail fin with almost normal shape, expression of *Rootletin* remains expanded into the apical territory of the tail fin and detectable

in lateral ectodermal cells (Fig. R2.3 O). Altogether, these observations suggest that the formation of the ectodermal tail fin in *B. lanceolatum* is dependent of RA signaling, with CYP26-dependent degradation playing an important role in fine tuning of endogenous RA signaling levels to ensure proper tail fin outgrowth.

CYP26-1 and CYP26-3 are highly responsive to RA and specifically upregulated to avoid teratogenic effects of RA

In vertebrates, CYP26 enzymes are known to function locally to reduce RA levels hence protecting target tissues from RA teratogenesis. In this context, the regulation of CYP26 activity is mediated, at least in part, by RA signaling, which very dynamically up- or down-regulates *CYP26* expression in a tissue-dependent context[496]. To obtain insights into the regulation of *CYP26* genes by RA signaling in *B. lanceolatum*, we performed quantitative real time PCR (qPCR) analyses on amphioxus embryos at two developmental stages (mid neurula and early larva) that have been treated, at the early gastrula stage, with either 1µM RA or 1µM of the RAR antagonist BMS009. The qPCR experiments assessed the changes in relative expression of the three *B. lanceolatum CYP26* genes normalized by *RAR* expression in control embryos (Fig. R2.4 A, B).

Through these analyses, we found that the expression levels of *CYP26-1* and *CYP26-3* in control embryos were very low when compared to *CYP26-2* (about 45,0 to 50,0 times less in mid neurulae and 2,5 to 5,0 times less in early larvae) (Fig. R2.4 A, B). RA treatments very significantly increased the expression of both *CYP26-1* and *CYP26-3* at the mid neurula stage (an average of 28 and 352,6 fold, respectively), while *CYP26-2* levels increased merely by about 2,1 fold (Fig. R2.4 A). In contrast, when embryos were treated with the RAR antagonist BMS009, *CYP26-1* levels decreased by an average of 18,7 fold in mid neurulae, while the overall transcription of *CYP26-2* and *CYP26-3* remained relatively unchanged when compared to controls (a decrease of about 1,1 fold for *CYP26-2* and an increase of about 1,4 fold for *CYP26-3*) (Fig. R2.4 A). In early larvae, RA treatment also significantly increased the expression levels of *CYP26-1* and *CYP26-3* (by an average of 5,2 and 102,6 fold, respectively), while that of *CYP26-2* increased only by about 1,6 fold (Fig. R2.4 B). The RAR antagonist BMS009 had the opposite effect and strongly decreased *CYP26-1* levels by about 123,0 fold and *CYP26-3* levels by an average of 5,8 fold (Fig. R2.4 B). In contrast, *CYP26-2* expression dropped by only about 1,2 fold (Fig. R2.4 B). These results indicate that, although *CYP26*-2 is the amphioxus *CYP26* gene most strongly and broadly expressed during

embryogenesis, *CYP26-1* and *CYP26-3* are much more reactive to alterations of endogenous RA signaling levels.

Following this quantification, we next assessed, by ISH, the developmental expression profiles of amphioxus *CYP26* genes upon RA signaling-dependent pharmacological treatments. Using the same developmental stages as for the qPCR analyses (mid neurulae and early larvae), we found that the upregulation of *CYP26-1* expression upon RA treatment was not uniform (Fig. R2.5 A-F). At the mid neurula stage, *CYP26-1* was most strongly induced in the anterior half of the embryo in all tissue layers (i.e. in CNS, ectoderm, mesoderm, and endoderm), all along the CNS as well as in the posterior tip of the embryo in all tissue layers excepting the mesoderm (Fig. R2.5 A, B). In early larvae, the gene was upregulated in all tissue layers of the anterior half of the animal, but largely downregulated posteriorly, except in a few cells in the ectoderm (Fig. R2.5 D, E). As expected from the qPCR experiments, expression of *CYP26-1* in embryos and larvae treated with the RAR antagonist BMS009 was very inconspicuous, but after an extended coloration step was nonetheless detectable in mesodermal tissues (Fig. R2.5 C, F).

Treatment effects were similar, but not identical, for *CYP26-2*. At the mid neurula stage, RA strongly induced *CYP26-2* anteriorly and posteriorly in all tissue layers as well as all along the CNS and ectoderm (Fig. R2.5 H). However, following RA treatment, *CYP26-2* expression was stronger anteriorly and posteriorly in the embryo, when compared to that of *CYP26-1*. In early larvae, the effects of RA treatments were also much less pronounced for *CYP26-2* than for *CYP26-1*. Chiefly, *CYP26-2* expression expands slightly in the mesoderm and ectoderm in the anterior half of the animal (Fig. R2.5 K, L). Of note, the reduction of *CYP26-2* expression in the anterior endoderm of RA-treated larvae was due to the absence of pharyngeal structures normally expressing the gene[467]. Treatments with the RAR antagonist BMS009 generally weakened the *CYP26-2* signal, most noticeably in the CNS, the mesoderm, and both the anterior and posterior ectoderm (Fig. R2.5 J, M).

Finally, the expression of *CYP26-3* was also very strongly expanded by RA at mid neurula and early larva stages, even more strongly than that of *CYP26-1* with a general expansion of the staining observed throughout the embryo at both stages (Fig. R2.5 N, O). Following RA treatment, expression of the gene was induced anteriorly and posteriorly in all tissue layers, most conspicuously in the center of the embryo in the CNS, ectoderm, and mesoderm (Fig. R2.5 N, O). By the early larval stage, *CYP26-3* remained generally upregulated throughout the animal (Fig. R2.5 Q, R), which contrasts with a more restricted

distribution of *CYP26-1* and *CYP26-2* transcripts at this stage of development in response to RA treatments (Fig. R2.5 E, L). In embryos and larvae treated with the RAR antagonist BMS009, the *CYP26-3* signal was very weak and chiefly limited to mesodermal tissues (Fig. R2.5 P, S).

Altogether, these results show that treatments with RA and RAR antagonist induce similar, but not identical, tissue-specific responses of amphioxus *CYP26-1*, *CYP26-2*, and *CYP26-3*, thereby supporting the notion that each one of the three *CYP26* genes is required for a distinctive set of developmental functions.

RA regulates CYP26-1, CYP26-2, and CYP26-3 directly and via evolutionary conserved functional RAREs present in the CYP26 cluster

To assess, whether the observed effects of RA and RAR antagonist on amphioxus *CYP26* expression were mediated directly by RAR/RXR heterodimers, we first carried out pharmacological treatments in the presence of puromycin, a compound that efficiently blocks *de novo* protein synthesis in developing amphioxus[432]. Amphioxus mid neurulae were hence treated with puromycin for 5 min prior to adding RA or the RAR antagonist BMS009. The embryos were then sampled one hour later and, following RNA extraction, the expression levels of *CYP26-1*, *CYP26-2*, and *CYP26-3* were determined by qPCR (Fig. R2.6). The results showed that, in the presence of puromycin, RA treatments significantly upregulated expression of both *CYP26-1* and *CYP26-3* (by about 5,8 fold and 5,9 fold, respectively), while that of *CYP26-2* increased more modestly, by about 1,6 fold (Fig. R2.6). The RAR antagonist BMS009 had the inverse effect, reducing *CYP26-1* and *CYP26-3* levels by about 1,8 fold and 2,9 fold, respectively. In contrast, *CYP26-2* expression stayed relatively stable, decreasing only by about 1,1 fold (Fig. R2.6). Altogether, these results are consistent with the effects of RA pharmacology on the expression of *CYP26* genes described above. Furthermore, they indicate that the responses to RA and RAR antagonist treatments of all three amphioxus *CYP26* genes do not require *de novo* protein synthesis, thereby implying that they must be mediated directly by RAR/RXR protein heterodimers present in the target tissues.

In order to obtain deeper insights into the mechanisms of the direct regulation of *CYP26-1*, *CYP26-2*, and *CYP26-3* by RA signaling we then investigated the genomic environment of *CYP26* genes in the three amphioxus species with available genomes: *B. floridae*[31,123], *B. belcheri*[128], and *B. lanceolatum*[129]. Our analyses allowed us to reconstruct the entire genomic clusters for the three amphioxus *CYP26* genes, previously described only for

B. floridae[127] (Fig. R2.7). Although variable in size (about 85 Kbp in *B. floridae*, 56 Kbp in *B. belcheri*, and 79 Kbp in *B. lanceolatum* from the start codon of *CYP26-1* to the stop codon of *CYP26-3*), the order of the three *CYP26* genes within the continuous clusters is conserved between the three amphioxus species, lending further support to the notion that these genes originated by tandem duplications from a single ancestral *CYP26* gene at the base of the cephalochordates. Making use of the complete sequences of the three amphioxus *CYP26* clusters, we further searched systematically for conserved RAREs in the vicinity of the *CYP26-1*, *CYP26-2*, and *CYP26-3* genes. This *in silico* survey included a genomic region encompassing 20 Kbp upstream of the *CYP26-1* start codon and 20 Kbp downstream of *CYP26-3* stop codon (Fig. R2.7). In total, we identified 38 candidate RAREs in *B. floridae,* 30 in *B. belcheri*, and 21 in *B. lanceolatum*. Of these, 16 RAREs were conserved between all three amphioxus species, both in terms of DR motif sequences and their relative position within the *CYP26* cluster (Fig. R2.7 B and Appendix 2).

These conserved amphioxus *CYP26* RAREs, i.e. 13 DR5, 1 DR2, and 2 DR3 elements, were subsequently used in EMSA analyses to investigate their capacity, *in vitro*, to interact with the *B. lanceolatum* RAR/RXR heterodimer. Note that, even though DR3 elements are not considered as common RAREs, they were included in this survey. The results showed that the DR2 element and 12 of the 13 DR5 elements can be recognized and bound by the *B. lanceolatum* RAR/RXR heterodimer (Figs. R2.7 B and 8), indicating that most of these RAREs might be functional *in vivo*. Furthermore, the consensus signature of the *in vitro* validated amphioxus DR5 elements [(A/G)G(G/T)T(C/G)A NNN(A/G)(A/C/G)(A/G)G(G/T)(T/A)CA] is very similar to the classical vertebrate DR5 signature [(A/G)G(G/T)TCA $(N)_5$ (A/G)G(G/T)TCA] (Appendix 2), suggesting that, as in vertebrates [509,510], amphioxus *CYP26* genes are regulated directly by RA signaling via RAREs located in close vicinity of the open reading frames (ORFs).

Furthermore, comparisons of the conserved and validated amphioxus *CYP26* RAREs with functional RAREs associated with vertebrate *CYP26* genes revealed that the amphioxus DR5-6 sequence located upstream of *CYP26-2* (Fig. R2.7 B and Appendix 2) is identical in sequence to a DR5 RARE located upstream of vertebrate *CYP26A1* genes [AGTTCA $(N)_5$ AGTTCA][500,509]. To verify, whether this DR5 RARE motif is a chordate innovation or an ancestral signature of bilaterian *CYP26* genes, we screened the genomic regions surrounding *CYP26* genes in the annelid *C. teleta*, the mollusk *L. gigantea*, the echinoderm *S. purpuratus*, the hemichordate *S. kowalevskii*, and three vertebrates (*Takifugu rubripes*, *Mus musculus*, and

Homo sapiens). While no DR5 RARE with a similar motif could be identified in the annelid *C. teleta* and the mollusk *L. gigantea*, the conserved DR5 RARE was recovered in all three vertebrate species as well as in the hemichordate *S. kowalevskii*. Intriguingly, in the echinoderm *S. purpuratus*, instead of a DR5 RARE, we found a DR2 RARE with two similar DR sequences [AGTTCA] and in an inverse orientation relative to the conserved DR5 RAREs in other species (Table 1). Of note, the conserved amphioxus DR5 RARE is located significantly further away from the *CYP26* start codon than in the other studied species. Although the biological significance of this finding still remains to be explored, these results nonetheless suggest that the direct control of *CYP26* expression by RA signaling, mediated at least in part by a conserved DR5 RARE, is an ancestral feature that was most likely absent in the last common ancestor of all bilaterians and that was thus only subsequently acquired at the base of the deuterostomes.

Species	RARE name	Orientation relative to the *CYP26 gene*	Sequence (5'–3')	Upstream of	Distance to start codon	Reference
S.p.		+	AGTTCAATAGTTCA	*CYP26*	2503 bp	This study
S.k.		-	AGTTCATACCCAGTTCA	*CYP26a*	137 bp	This study
B.f.	DR5-6	-	AGTTCAACAAAAGTTCA	*CYP26-2*	6405 bp	This study
B.b.	DR5-6	-	AGTTCAACAAAAGTTCA	*CYP26-2*	6740 bp	This study
B.l.	DR5-6	-	AGTTCAACAAAAGTTCA	*CYP26-2*	4083 bp	This study
H.s.	R1	-	AGTTCACCCAAAGTTCA	*CYP26A1*	132 bp	[500]
M.m.	R1	-	AGTTCACCCAAAGTTCA	*CYP26A1*	134 bp	[500]
D.r.	R1	-	AGTTCACACAAAGTTCA	*CYP26A1*	166 bp	[500,509]
	R2	-	AGTTCAAGGATAGTTCA	*CYP26A1*	1963 bp	
T.r.	R1	-	AGTTCATTCAAAGTTCA	*CYP26A1*	103 bp	This study

Table 1. Conserved retinoic acid response elements (RAREs) with the characteristic sequence signature AGTTCA(N)$_5$AGTTCA identified in the vicinity of bilaterian *CYP26* genes. Species name abbreviations: B.b., *Branchiostoma belcheri*; B.f., *Branchiostoma floridae*; B.l., *Branchiostoma lanceolatum*; D.r., *Danio rerio*; H.s., *Homo sapiens*; M.m., *Mus musculus*; S.k., *Saccoglossus kowalevskii*; S.p., *Strongylocentrotus purpuratus*; T.r., *Takifugu rubripes*.

Discussion

CYP26 genes and their evolutionary involvement in RA-dependent A-P patterning in chordates

In this study, we assessed the developmental expression patterns and the biological functions of the three *CYP26* duplicates from the European amphioxus *B. lanceolatum*. Globally, our results suggest a sub-functionalization of these genes. Of these three genes, *CYP26-1* and *CYP26-3* are characterized by weak and disperse expression patterns, while *CYP26-2* displays a dynamic, tissue-specific pattern along the A-P axis. Considering that A-P patterning in early stages of amphioxus development is dependent on RA signaling[427,432], the expression profile of *CYP26-2*, at very distinct positions along the A-P axis of the gastrula and early neurula, is suggestive of a functional role for this gene in this RA-dependent patterning process. This notion is further supported by the results obtained by pharmacological inhibition of CYP26 action during gastrulation, which yielded embryos and larvae that resemble those treated with exogenous RA, displaying severe A-P patterning defects[467]. Thus, we propose that CYP26 enzymes assume two main functions during amphioxus development (Fig. R2.9), which are equivalent to those observed in vertebrates: (1) the mediation of RA-dependent developmental patterning and (2) the protection against fluctuations of RA levels. Due to its conspicuous and tissue-specific expression during development, we hypothesize that the former is mainly assumed by *CYP26-2*, while the latter is dependent on the activity of *CYP26-1* and *CYP26-3*.

During early development, the expression domains of *CYP26-2* along the A-P axis are inversely correlated with those of *Hox1* and *HNF3-1*, both of which are direct targets of RA signaling in amphioxus[432,511]. The posterior expression limit of *CYP26-2* in anterior tissues thus abuts the anterior border of both the *Hox1* and *HNF3-1* domains[432]. Concomitantly, the blastopore-associated expression of *CYP26-2* seems to delineate the posterior limits of both *Hox1* and *HNF3-1*[432,511]. Given that *RAR* expression is ubiquitous during gastrulation [418], the presence of *CYP26-2* in defined domains along the A-P axis might thus be crucial for the creation of RA sinks to subdivide the developing embryo into zones with and without active RA signaling. This is reminiscent of the conserved expression and function of *CYP26A1* during vertebrate gastrulation: *CYP26A1* is expressed in the anterior neural ectoderm and functions to establish A-P boundaries in the developing CNS[496]. It does so by establishing an

anterior sink for RA produced posteriorly in the paraxial mesoderm, hence creating a RA gradient along the A-P axis of the neuroectoderm[512,513].

By the mid neurula stage, the *CYP26-2* signal is present in the entire anterior CNS, with a posterior limit at the boundary between the first and second somite. This *CYP26* expression might be limited to a region homologous to the vertebrate forebrain and midbrain[108], which is considerably different from what is observed in the vertebrate CNS, where *CYP26* genes are expressed in the developing hindbrain and are fundamental for its patterning along the A-P axis[345,496]. In the amphioxus hindbrain homolog, RA signaling mediated by RAR/RXR has been shown to confer regional identity along the A-P axis by controlling the collinear expression of *Hox* genes[427]. However, our work suggests that, in contrast to the situation in vertebrates, this process does not involve the deployment of *CYP26* genes in the amphioxus hindbrain homolog. Thus, while the early role for *CYP26* in neural patterning was probably already present in the last common ancestor of amphioxus and vertebrates, a *CYP26*-dependent mechanism for subsequent hindbrain regionalization probably evolved in vertebrates, following the vertebrate-specific *CYP26* duplications.

Following neurulation, the three amphioxus *CYP26* genes are expressed in anterior mesoderm, most noticeably in the anterior-most somites. This is intriguing, as to date there is no convincing evidence for a requirement of RA signaling in amphioxus mesoderm development[427,430,468] and the inhibition of CYP26 function by pharmacological treatments does not yield a mesodermal phenotype[467]. These observations raise an important question: why are *CYP26* genes expressed in the anterior amphioxus mesoderm, as they are in vertebrate head mesoderm[499], if they are not required for A-P patterning of this tissue? We speculate that the conspicuous expression of *CYP26* in the anterior somites is an amphioxus innovation and is thus not comparable to the role of *CYP26* in the vertebrate head mesoderm. Instead, the *CYP26* genes might function in the anterior amphioxus mesoderm to establish a buffer zone between the CNS dorsally and the pharynx ventrally, which require distinct RA signaling cues for A-P regionalization[427,430]. Although further analyses aiming, for example, at the visualization of *in vivo* RA signaling levels[513] will be required to test this hypothesis, it would nonetheless be interesting to assess, whether this mechanism for separating two gradient-based A-P patterning systems is being used more widely during development of small-sized embryos.

In the course of development, the blastopore-associated expression of amphioxus *CYP26-2* becomes incorporated into the tail bud, a structure that, from the neurula stage on, plays a central role in posterior elongation of the embryo and larva and that is further

characterized by the expression of tissue-specific marker genes, such as *Wnt3*[514]. In vertebrates, both *CYP26A1* and *Wnt3a* are also co-expressed in the developing tail bud, and *CYP26A1* null mutants are characterized by a truncated tail and a spatial expression of *Wnt3a* that is abnormally restricted towards the midline of the posterior neural plate[292]. These observations suggest that, in the vertebrate tail bud, *CYP26A1* functions to keep RA levels low to create a permissive environment required for the posterior elongation of the embryo[403,515]. In contrast, in amphioxus, treatments with CYP26 inhibitor, exogenous RA or the RAR antagonist BMS009 do not affect expression of *Wnt3* in the tail bud[430]. Importantly, these pharmacological treatments, albeit affecting tail fin development, do not impact posterior elongation of the developing embryo[426,468]. Together, these findings support the notion that CYP26 function, and more generally the RA signaling system, is not required for tail bud-driven body extension in amphioxus and that this role for RA likely evolved in the vertebrate lineage.

CYP26 activity is required for the development of the amphioxus tail fin

Previous studies in the Florida amphioxus, *B. floridae*, have shown that the administration of excess RA during gastrulation results in a small anus[426] and that the continuous administration of RA to *B. floridae* larvae leads to closure of the anus and regression of the tail fin[429]. Our data on the European amphioxus, *B. lanceolatum*, are consistent with these previous findings and further suggest that RA signaling is required for proper tail fin outgrowth. Given that *CYP26-2* is the only *B. lanceolatum CYP26* gene expressed in the posterior ectoderm, it is very likely responsible for fine-tuning endogenous RA signaling levels in this territory.

The amphioxus tail fin is established by columnar epidermal cells that contain a large ciliary rootlet[507,508] and it has previously been shown that RA promotes the downregulation of a major component of this structure, the protein Rootletin[429]. This downregulation in turn is likely responsible for the induction of tail fin regression upon RA treatments in *B. floridae* larvae[429]. Our results of RA and CYP26 inhibitor treatments in *B. lanceolatum* are generally consistent with these previous observations, although we identified a major difference in the developmental timing of the involvement of RA signaling in tail fin outgrowth between the two amphioxus species. This fact is exemplified by the experimental setups required to obtain tail fin phenotypes. While in *B. floridae* a RA-dependent tail fin malformations can only be obtained by continuous treatment of pre-metamorphic larvae with 6-8 gill slits[429], changes in

tail fin morphology can be induced much earlier during *B. lanceolatum* development by treating in the course of gastrulation. This suggests that, albeit required for tail fin outgrowth in both amphioxus species, the attenuation of high RA signaling levels in the posterior ectoderm might take place earlier in *B. lanceolatum* than in *B. floridae* development. It is possible that this developmental difference is mediated by a delayed activation of the expression of *CYP26-2* in *B. floridae* relative to *B. lanceolatum*, although additional work is required to support this claim. Altogether, these results indicate that amphioxus tail fin development requires low levels of RA signaling, which are maintained by CYP26 activity. When compared to the known functions of CYP26 in vertebrates, where cells expressing CYP26 enzymes are said to be effectively devoid of RA[515], amphioxus tail fin outgrowth might represent a rare example of a developmental process, where CYP26 is deployed to merely reduce, and not to completely eliminate, RA signaling levels.

Amphioxus CYP26 genes are highly responsive RA signaling targets

In vertebrates, it has been established that, in RA-sensitive tissues, RA induces *CYP26* to generate a negative feedback loop that reduces the overall amount of available RA[496]. In mice, for example, expression of both *CYP26A1* and *CYP26B1* is upregulated upon RA treatment[489] and, at least for *CYP26A1*, this regulation is directly mediated by RAR/RXR binding to RAREs in the vicinity of the *CYP26A1* gene[500]. In contrast, the vertebrate *CYP26C1* gene is likely not contributing to this RA-dependent negative feedback system, as its expression is actually downregulated upon RA stimulation[489]. The data presented here suggest that expression of all three amphioxus *CYP26* genes, which are lineage-specific duplicates and located in a single cluster in the genome, is positively regulated by RA signaling. Accordingly, the transcriptional regulation of *B. lanceolatum CYP26-1*, *CYP26-2*, and *CYP26-3* is in all likelihood directly mediated by RAR/RXR heterodimers, given that one DR2-type sequence and 12 DR5-type elements within the *CYP26* cluster are recognized and bound by the *B. lanceolatum* RAR/RXR heterodimer *in vitro*. Of these 13 *in vitro* validated RAREs, four DR5 elements are located around the *CYP26-1* coding sequence, four DR5 and one DR2 are found in proximity of *CYP26-2*, and four DR5 elements are associated with the *CYP26-3* gene. Although each of the three amphioxus *CYP26* genes are likely to be directly regulated by RA signaling, the specific arrangement of RAREs relative to the *CYP26* ORFs nonetheless suggests that the expression of each of the three genes is regulated independently by different sets of RAREs. Although this fundamental difference in the regulation of

amphioxus and vertebrate *CYP26* genes remains to be demonstrated mechanistically *in vivo*, it is nonetheless tempting to speculate that alterations in the regulation by RA have been key for redefining the functions of the different *CYP26* genes following their duplication, hence leading to their genomic retention by sub-functionalization.

At least some of the RAREs within the amphioxus *CYP26* cluster might serve as hubs for long-range regulation of gene expression from shared RAREs. Although in both amphioxus and vertebrates it has previously been shown that functional RAREs are generally located in the proximity of genes directly regulated by RA signaling[403,440], long-range regulatory mechanisms of RA signaling have, for example, been implicated in the control of the rostral expansion of posterior *Hoxb* genes during mouse CNS development[516]. The *in vivo* validation of the RAREs located within the amphioxus *CYP26* cluster will shed light on their contribution to short- and/or long-range transcriptional control mechanisms exerted by RA signaling. Our *in silico* analyses further revealed the presence, within the amphioxus *CYP26* cluster, of two DR3 elements in close proximity of one of the *in vitro* validated DR5 elements. Although DR3 elements are generally not recognized by RAR/RXR heterodimers, they are bound by other nuclear receptors, such as vitamin D receptors[517]. The presence of DR3 elements within the amphioxus *CYP26* cluster thus hints at the possibility that additional nuclear receptors are involved in the regulation of amphioxus *CYP26* genes.

Interestingly, the consensus sequence of the *in vitro* validated amphioxus DR5 RAREs is identical to the classical vertebrate DR5 sequence consensus [(A/G)G(G/T)TCA $(N)_5$ (A/G)G(G/T)TCA][250], suggesting that the DNA binding properties of amphioxus and vertebrate RAR/RXR heterodimers are highly conserved. Along these lines, we found a similar, conserved DR5 RARE in the regulatory region of a hemichordate *CYP26* gene as well as an equivalent DR2 RARE close to the *CYP26* gene of a sea urchin. Whether these elements are recognized by RAR/RXR heterodimers in these two species is currently unknown. Their presence is nevertheless highly suggestive of a biological function[493], which thus indicates that direct regulation of *CYP26* genes by RA signaling is an ancestral feature that was already present in the last common ancestor of all deuterostomes. In contrast, the lack of conserved RAREs in the regulatory regions of the *CYP26* genes of lophotrochozoans, which generally encode *RAR* and *RXR* genes in their genomes[402], support the notion that the direct regulation of *CYP26* transcription by RA signaling is not an ancestral feature of bilaterian animals. This hypothesis is further strengthened by the fact that RARs of gastropod mollusks are unable to bind RA and hence to activate transcription in its presence[249,457].

The evolutionary history of CYP26 genes

In the animal kingdom, the basic molecular components of the RA machinery have previously been described in a wide variety of bilaterian animals, including both protostomes and deuterostomes[402,451]. Genes encoding the RA degrading enzyme CYP26 have, for example, been identified in the genomes of priapulids, brachiopods, mollusks, annelids, sea urchins, hemichordates, cephalochordates, ascidian tunicates, and vertebrates, but not in those of nematodes and arthropods, suggesting a secondary loss of *CYP26* subfamily genes in these two animal lineages[402,461]. We also found evidence for possible secondary losses of *CYP26* genes in different vertebrates, including the lamprey *P. marinus* (lacking *CYP26A1*), the chimera *C. milii* (lacking *CYP26A1*), the skate *L. erinacea* (lacking *CYP26C1*), the coelacanth *Latimeria chalumnae* (lacking *CYP26C1*), and the opossum *Monodelphis domestica* (lacking *CYP26B1*) (Appendix 2). Additional sequence information will be required to validate the absence of these genes from their respective genomes and to assess, at which point in evolution the confirmed gene losses occurred.

Our data further indicate that the evolutionary diversification of the vertebrate *CYP26* genes was a highly complex process. Given the presence of at least one *CYP26A1* and two *CYP26B1/C1* genes in lampreys, the last common ancestor of cyclostomes and gnathostomes probably already possessed at least two *CYP26* genes. Intriguingly, in the lamprey *L. japonicum*, *CYP26A1* and one of the two *CYP26B1/C1*, *CYP26B1/C1a*, are physically linked in a tandem cluster in the genome (on scaffold 19) (Appendix 2), just like *CYP26A1* and *CYP26C1* in gnathostome genomes[518]. This tandem cluster thus very likely originated before the cyclostome-gnathostome split, probably by a tandem duplication event predating the vertebrate-specific WGD. The two linked *CYP26* genes were then duplicated during the first WGD, which took place in the basal vertebrate lineage before the cyclostome-gnathostome split[34,37,519]. The subsequent loss of one of the duplicated *CYP26A1* genes yielded the *CYP26* complement of extant lampreys, such as *L. japonicum*: one *CYP26A1* and two *CYP26B1/C1* genes. In the gnathostome lineage, additional duplications[34,37,519] finally resulted in the diversification of *CYP26B1* and *CYP26C1* genes from the ancestral *CYP26B1/C1*. Although requiring additional scrutiny, this proposed scenario for vertebrate *CYP26* diversification suggests that, if two rounds of WGD occurred in the course of vertebrate evolution, the first likely took place before and the second after the cyclostome-gnathostome split[34,519].

Alternatively, a single, ancient WGD might have occurred before the cyclostome-gnathostome split and was followed by independent segmental duplications in the cyclostome and gnathostome lineages[37]. Interestingly, in teleost fish, which underwent an additional round of WGD[520], there are also only three *CYP26* genes, one member of each gnathostome paralogy group (*CYP26A1*, *CYP26B1*, and *CYP26C1*) and synteny analyses have shown that the teleost-specific *CYP26* duplicates have been non-functionalized in the course of evolution, in accordance with the DDC model[35,502,518].

Our phylogenetic analyses also provided evidence that *CYP26* genes underwent multiple duplication events, not only in vertebrates, but also in invertebrates, such as cephalochordates, ascidian tunicates, hemichordates, mollusks, annelids, brachiopods, and priapulids. Together, these results suggest that the ancestral bilaterian possessed a single *CYP26* gene that was subjected to independent duplication events in different animal lineages. By correlating the timing of the lineage-specific duplications of invertebrate *CYP26* genes with the geological timescale, we observed that the duplication events fall within three distinct time periods, i.e. the late Ordovician, the early Carboniferous, and the middle Permian/middle Triassic. Thus, the priapulid *CYP26* duplication took place during the late Ordovician, the first cephalochordate duplication and that identified in the hemichordate *S. kowalevskii* in the early Carboniferous, and the second cephalochordate duplication and that in the hemichordate *P.flava* in the middle Permian/middle Triassic. During the late Ordovician, a period marked by a mass extinction of marine species[521], earth was characterized by a rising level of atmospheric O_2[522] and by vast shallow and warm continental seas [521]. These conditions were very favorable for the appearance of cyanobacterial mats at moderate depths of the water column[523,524]. Similarly, during the early Carboniferous atmospheric O_2 levels were rising[522,525], and the fossil record accounts for one of the highest concentrations of calcified marine cyanobacteria[524]. The middle Permian/middle Triassic, in turn, saw the greatest biotic crisis in earth's history[526], and the fossil record suggests that, as a result of environmental changes, cyanobacteria became one of the most abundant life forms in both shallow and deep water environments[527].

Extant cyanobacteria that are known to create massive blooms under the exact same environmental conditions include *Trichodesmium*, *Anabaena*, and *Synechocystis*[528,529], all of which are also known to produce high levels of carotenoids and retinoids as anti-oxidative byproducts, which have been shown to induce teratogenic effects in the surrounding fauna[529–531]. The prevalence of cyanobacterial mats in marine environments during the late Ordovician,

early Carboniferous, and the middle Permian/middle Triassic might thus explain why *CYP26* duplications that occurred during these two geological periods were independently retained, either by neo- or sub-functionalization[35,502], in the genomes of different marine animal lineages. The independent duplication of *CYP26* genes might have increased the overall fitness of a given population by favoring individuals that were more efficient in buffering fluctuations of exogenous retinoid levels. Altogether, this finding represents an intriguing example of adaptive convergent evolution in response to environmental changes.

Conclusions

In the present study, we characterized the expression, function, and regulation of *CYP26* genes during amphioxus development. Our data suggest that, despite the independent origins of the *CYP26* gene repertoires in chordates, the *CYP26* genes of cephalochordates and vertebrates convergently evolved similar developmental functions: RA-dependent patterning and homeostatic regulation of RA levels. Moreover, by comparing the regulatory regions of *CYP26* genes in three amphioxus species with those from several different animal taxa, we identified a highly conserved, functional RARE, suggesting that negative feedback regulation of RA signaling is an evolutionary ancient mechanism for controlling endogenous RA levels. This mechanism of regulation was likely already present in the last common ancestor of all deuterostomes, but not in that of all bilaterians. Finally, the correlation between the timing of lineage-specific duplications of bilaterian *CYP26* genes and major environmental changes in the geological record suggest that the evolutionary diversification of the *CYP26* subfamily in bilaterians was strongly influenced by environmental pressures to buffer fluctuations of exogenous retinoid levels. In sum, this work thus sheds light on the evolution of the regulation of endogenous RA levels and establishes a framework for studying adaptive convergent evolutionary changes following gene duplication.

Material and Methods

Amphioxus adult husbandry, embryo rearing, and pharmacological treatments

Sexually mature animals of the European amphioxus (*Branchiostoma lanceolatum*) were collected by dredging in Argelès-sur-Mer, France, and retrieved from the sand by sieving. The collected animals were split evenly into tanks, with about 10-15 animals per aquarium, males and females together. The water temperature in the aquaria was kept at 16-17ºC, and the animals were kept under a spring-like day/night period, with 14 hours of light and 10 hours of absolute darkness. Spawning was induced by a 36-hour thermal shock at 23ºC, as previously described [148,196,213]. Following oocyte and sperm collection and *in vitro* fertilization, the embryos were raised in artificial seawater, in the dark, at 19ºC[174]. Pharmacological treatments of *B. lanceolatum* embryos were performed at the late blastula stage (6h of development) with all-*trans* RA (at 0,1 µM and 1 µM) (Sigma-Aldrich, Saint-Quentin Fallavier, France), the RAR antagonists BMS009 (at 1 µM) or BMS493 (at 1 µM) (Sigma-Aldrich, Saint-Quentin Fallavier, France), the CYP26 inhibitor R115866 (at 0,5 µM and 0,1 µM) (provided by Janssen Research & Development, a division of Janssen Pharmaceutica NV, Beerse, Belgium) or with a combination of both BMS493 (at 1 µM) and R115866 (at 0,5 µM). All compounds were initially dissolved in dimethyl sulfoxide (DMSO) to create 1000X stock solutions and subsequently added to the embryo cultures in artificial seawater in a 1:1000 dilution to yield the respective final concentrations. As a control, embryos were treated in separate dishes with DMSO alone to a final dilution of 1:1000 [418,426]. Puromycin (Sigma-Aldrich, Saint-Quentin Fallavier, France) treatments were performed at the mid neurula stage (19h of development) by adding the compound to embryo cultures at a final concentration of 200 µg/ml. After 5 min of incubation, all-*trans* RA (1 µM), the RAR antagonist BMS009 (1 µM) or DMSO (at a dilution of 1:1000) were added, and, 1 h thereafter, the embryos were frozen for RNA extraction.

Sequence analyses and gene cloning

The *B. lanceolatum CYP26-1, CYP26-2*, and *CYP26-3* sequences were obtained *in silico* from the *B. lanceolatum* genome by local BLAST using as template the sequences of *B. floridae* and *B. belcheri CYP26* (Appendix 2). RNA was extracted from embryos at different developmental stages according to the established protocols[532] and cDNA was synthesized using the SuperScriptIII reverse transcription kit (Invitrogen, Cergy Pontoise, France).

Complete coding sequences of *B. lanceolatum CYP26-1, CYP26-2*, and *CYP26-3* were subsequently cloned by PCR using gene-specific primers containing the start and stop codons of the three *CYP26* genes. RACE-PCR experiments using the SMARTer™ RACE cDNA amplification kit (Clontech, Saint-Germain-en-Laye, France) were then performed to amplify the 5' and 3' untranslated regions (UTRs) of *B. lanceolatum CYP26-1, CYP26-2*, and *CYP26-3*. The PCR and RACE-PCR products were cloned into the pCRII-TOPO vector (Invitrogen, Cergy Pontoise, France) and sequenced on both strands for validation. The *B. lanceolatum CYP26-1, CYP26-2*, and *CYP26-3* sequences were deposited in GenBank and their accession numbers are as follows: *CYP26-1* (KX118106), *CYP26-2* (KX118108), and *CYP26-3* (KX118107). Furthermore, a 2291-bp piece of the *B. lanceolatum Rootletin* gene was amplified by PCR, cloned into the pGEM-T Easy vector (Promega, Charbonnières-les-Bains, France), and verified by sequencing on both strands (GenBank accession number: KX118111), before being used as a marker for *in situ* hybridization experiments.

Phylogenetic analyses

Phylogenetic trees of the *CYP26* subfamily were calculated from amino acid and nucleotide sequences and the sequences included in the analysis are listed in Appendix 2, along with representative members of the *CYP51* subfamily, which were used as outgroup. Nucleotide sequences were translated into amino acid sequences, aligned with Muscle as implemented in SeaView v4.5.4[533], and refined by eye. The final amino acid alignment was subsequently retransformed into the final nucleotide alignment. The best-fit models of amino acid and nucleotide sequence evolution were selected based on the AIC score implemented, respectively, in ProtTtest 3[534] and jModelTest v2.1[535]. Molecular phylogenies were calculated with the Maximum Likelihood (ML) method using PhyML v3.1[536] under the models selected by ProtTtest 3 (LG+I+G+F) and jModelTest v2.1 (GTR+I+G). In addition, Bayesian Inference (BI) analyses were performed using the program MrBayes v3.2.6 [537] using the same models. The robustness of each node was estimated by bootstrap analyses (in 1000 pseudoreplicates) using PhyML v3.1 for the ML tree and by posterior probability for the BI tree, with two Monte Carlo Markov Chain (MCMC) analyses run independently for 100 million generations and trees sampled every 1000 generations. The burn-in was determined with Tracer v1.6[538], and the average standard deviation of split frequencies remained at <0.05 after the burn-in threshold. 10% and 50% of the trees were discarded, respectively, for the amino acid and nucleotide datasets. Consensus trees were visualized with Figtree v1.4[539].

Comparative dating analysis

Based on the nucleotide dataset, the divergence dates of the *CYP26* sequences were estimated with Beast v2.4[540] using divergence times in million years (Myr) estimated with the calibration intervals proposed by Benton and colleagues and dos Reis and colleagues [541,542]: origin of Vertebrata (457,5-636,1 Myr), Euarchontoglires (61,6-164,6 Myr), and Hemichordata (504,5-636,1 Myr) as well as the divergence between Holostei and Teleostei (250,0-331,1 Myr) and between Otocephala and Euteleostei (150,94-235 Myr). All calibration constraint sets were defined following a hard minimum, soft upper boundaries, and a lognormal prior[540]. Markov Chain Monte Carlo (MCMC) analyses were run on the nucleotide dataset for 100 million generations with trees sampled every 1000 generations[540]. Convergence of the calculations was verified and burn-in estimated with Tracer v1.6[538]. The results of the MCMC run were sampled with LogCombiner and a burn-in of 30%[540]. The trees were combined into a maximum clade credibility tree using TreeAnnotator with an estimation of the mean node height and highest posterior density intervals fixed at 95%[540].

B. lanceolatum CYP26 ***cluster reconstitution and identification of putative RAREs***

The full-length ORFs of *B. lanceolatum CYP26-1*, *CYP26-2*, and *CYP26-3* were used as templates for local BLAST searches of the *B. lanceolatum* genome to identify scaffolds containing UTRs and coding regions of *CYP26-1*, *CYP26-2*, and *CYP26-3*. Sequences of genome regions not obtained by these BLAST approaches (hence corresponding to gaps in the *B. lanceolatum CYP26* cluster) were subsequently identified by reciprocal BLAST searches of the *B. lanceolatum* genome sequence using short regions (less than 2 kb) of the *B. floridae* and *B. belcheri CYP26* clusters. Putative RARE sequences were identified using an automated pipeline[429] using as input sequences RAREs previously described as functional in amphioxus as well as all possible RARE combinations resulting from the canonical vertebrate RARE consensus: (A/G)G(G/T)TCA$(N)_{0-9}$(A/G)G(G/T)TCA. Detailed sequence and scaffold information is provided in Appendix 2.

Multiple expectation maximization algorithm for motif elicitation (MEME) analysis

Multiple expectation maximization algorithm for motif elicitation (MEME) logos were calculated with MEME Suite v4.10.1[543] using as input file the sequences of the *in vitro* validated *B. lanceolatum* DR5 RARE sequences including 5 nucleotides upstream and

downstream of the element. The following settings were used to obtain the MEME logos: nmotifs = 2, minwidth = 15, maxwidth = 27.

Electrophoretic mobility shift assay (EMSA) experiments

The *B. lanceolatum RAR* and *RXR* coding sequences were amplified by PCR and cloned into the pGEM-T Easy vector (Promega, Charbonnières-les-Bains, France). Subcloning into the pCS2+ vector[544] was performed using introduced EcoRI and XhoI restriction sites. Following verification by sequencing on both strands of the RAR- and RXR-pCS2+ constructs (GenBank accession numbers: *B. lanceolatum RAR*, KX118109; *B. lanceolatum RXR*, KX118110), RAR and RXR proteins were produced by *in vitro* translation using the TNT coupled *reticulocyte* lysate systems (Promega, Charbonnières-les-Bains, France). Electrophoretic mobility shift assays (EMSAs) were performed as previously described[249] using 4 µl of each receptor synthesized *in vitro* and 30-50 x 10^3 CPM of double-stranded oligonucleotide probe end-labeled with (γ-^{32}P)ATP, incubated for 30 minutes on ice in a final volume of 20 µl of binding buffer: 20 mM Tris-HCl pH8, 50 mM KCl, 2 mM DTT, 25 mM $MgCl_2$, 50 mM NaCl, 1 µg poly (dI-dC), and 10% glycerol. The samples were subsequently run on a 5% native acrylamide gel in 1X TAE for 2 hours.

In situ hybridization, histology, imaging, and tail fin measurements

Antisense riboprobe synthesis, *in situ* hybridization and Hoechst staining (Invitrogen, Cergy Pontoise, France) experiments were performed as previously described[545]. For *in situ* hybridization, Hoechst staining, and tail fin measurements, amphioxus (*B. lanceolatum*) embryos and larvae were fixed in 4% paraformaldehyde at different developmental stages, as previously described[545]. Following *in situ* hybridization, *B. lanceolatum* embryos and larvae were first photographed as whole mounts using Zeiss DIC (differential interference contrast) optics (Carl Zeiss SAS, Marly le Roi, France) and subsequently counterstained in Ponceau S (Sigma-Aldrich, Saint-Quentin Fallavier, France), embedded in Spurr's resin (Sigma-Aldrich, Saint-Quentin Fallavier, France), and prepared as 3 µm sections for light microscopy observations and photography. Following Hoechst staining, *B. lanceolatum* larvae were embedded in Mowiol mounting medium (Sigma-Aldrich, Saint-Quentin Fallavier, France) overnight at 4ºC, before being imaged using a Leica TCS SP5 confocal microscope (Leica Microsystems SAS, Nanterre, France). ImageJ was subsequently used for image processing and for the creation of maximal projections[546]. For tail fin measurements, normal and treated

B. lanceolatum larvae were photographed using Zeiss DIC optics (Carl Zeiss SAS, Marly le Roi, France). The length of the tail fin, defined as the distance between the anterior-most end of the tail fin ectoderm and the posterior-most tip of the tail fin (as represented in Fig. R2.3 G), was subsequently measured using the measurement tool of ImageJ[546] and ultimately represented as mean ± standard deviation, with n = 15 for each treatment and control conditions. Student's t-test was used to assess the statistical significance of the length differences measured in the treatment conditions relative to the DMSO control.

Quantitative real time PCR (qPCR) assays

Quantitative real time PCR (qPCR) experiments were performed at two developmental stages (mid neurula at 20h of development and early larva at 48h of development). The cDNAs from DMSO treatment controls as well as from embryos treated at the late blastula stage (6h of development) with either 1 µM all-*trans* RA or 1 µM of the RAR antagonist BMS009 were assayed on a MJ Research DNA Engine Opticon system (Bio-Rad, Marnes-la-Coquette, France) using the QuantiTect SYBR Green PCR reagent (Qiagen SAS, Courtaboeuf, France) and primers specific for *B. lanceolatum CYP26-1*, *CYP26-2*, *CYP26-3*, *RAR*, and *18S rRNA* (Appendix 2). Based on the lack of response to the different pharmacological treatments assayed, *18S rRNA* was selected as the reference for internal standardization of the starting quantity of RNA. Each qPCR experiment was performed in triplicates and the relative expression was normalized to *RAR* or *CYP26-2* expression levels, in Fig. R2.4 and 5, respectively. Normalized expression levels are shown as ΔΔCT means ± standard deviation, with n = 3. Furthermore, fold change of expression relative to the control is represented as the mean of the ΔΔCT ratios of all possible combinations of the three replicates of a given condition over the three controls ± standard deviation, with n = 9.

Figures

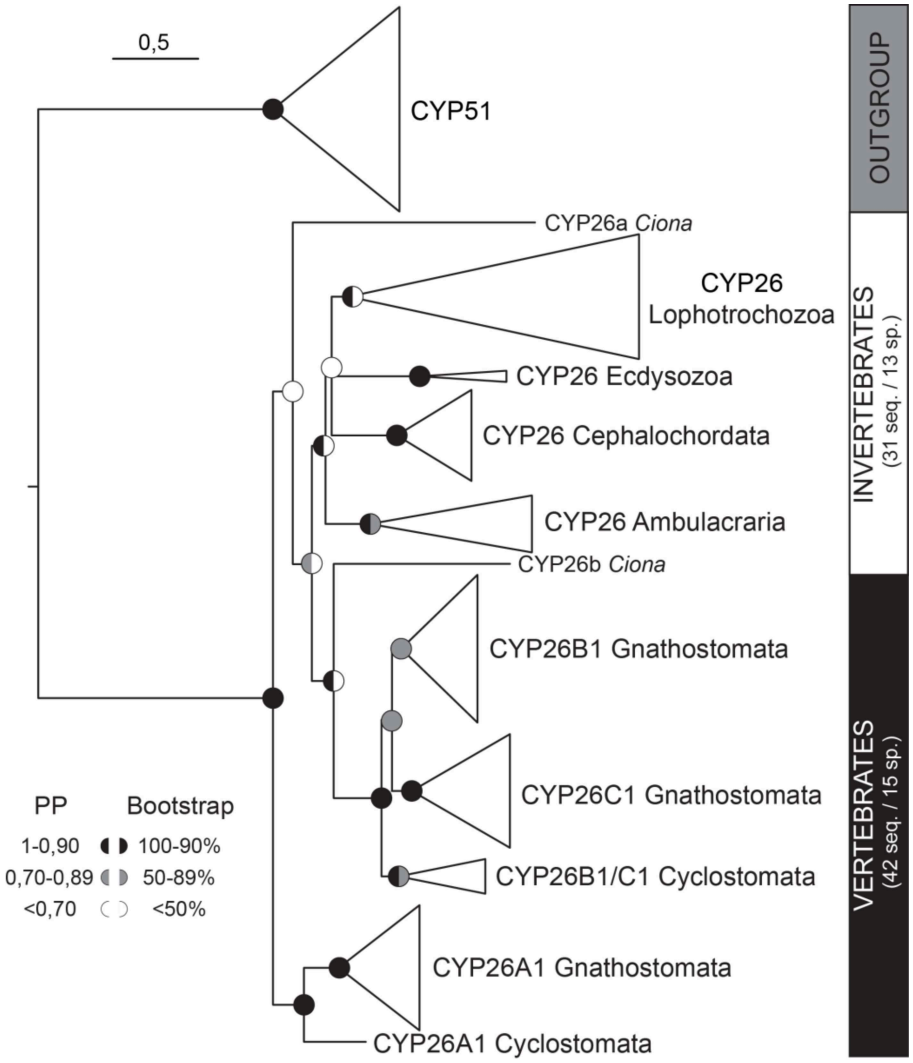


Fig. R2.1 – Phylogenetic analysis of the CYP26 subfamily. *Diagrammatic summary of Bayesian Inference (BI) and Maximum Likelihood (ML) analyses of the phylogenetic relationships within the CYP26 subfamily, with CYP51 used as outgroup. Detailed tree and sequence information in Appendix 2. Branch lengths are representative of the amino acid substitution rate and branch support is indicated at each major node as posterior probabilities (PP) for the BI tree and as bootstrap percentages for the ML tree. Furthermore, the total number of sequences (seq.) and species (sp.) is stated for the CYP26 subfamily.*

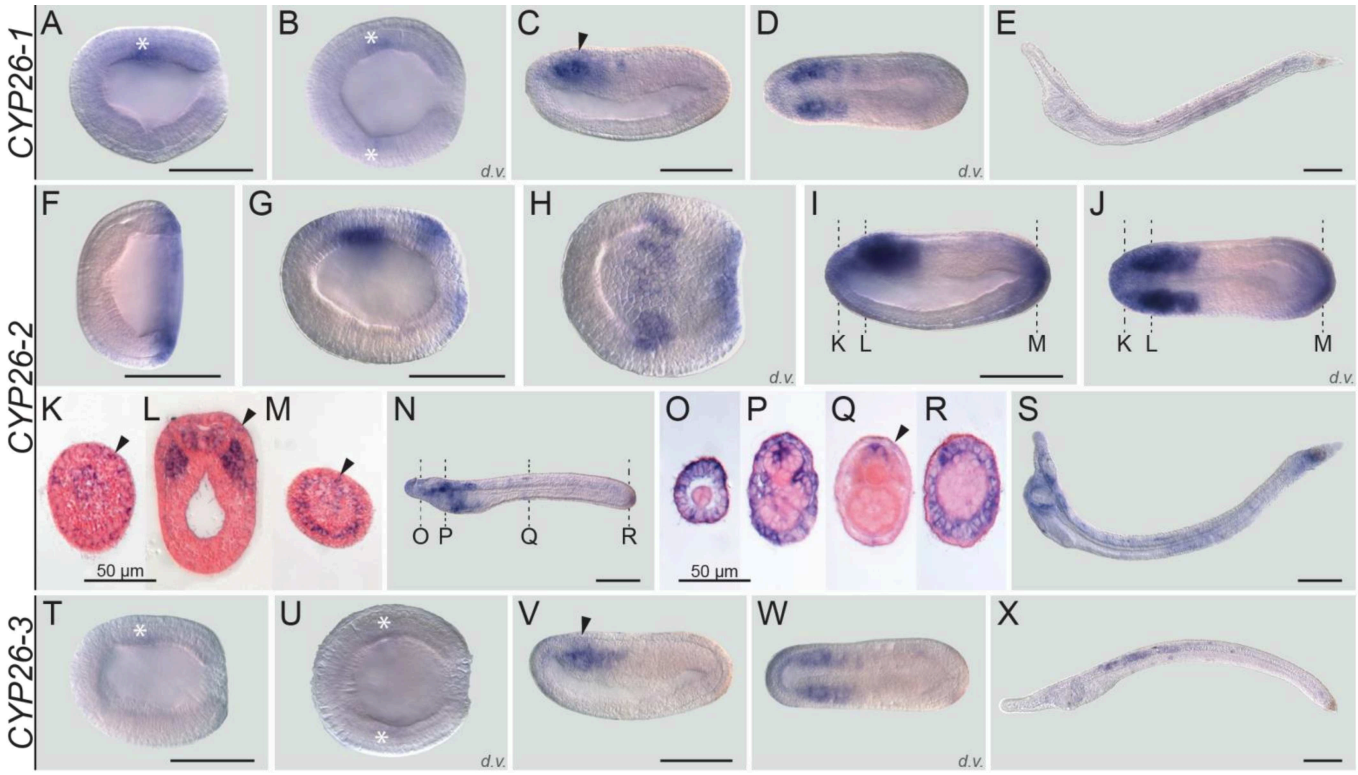


Fig. R2.2 – Developmental expression patterns of amphioxus CYP26 genes. *Whole mount in situ hybridization experiments were carried out for CYP26-1 (A-E), CYP26-2 (F-S), and CYP26-3 (T-X). Amphioxus (Branchiostoma lanceolatum) embryos and larvae are shown as lateral views with anterior to the left and the dorsal side up, excepting for (B, D, H, J, U, W), which are dorsal views (d.v.) with anterior to the left, and for (K-M) and (O-R), which are cross-sections viewed from the front. Cross-sections in (K-M) are through the embryo shown in (I, J) at the levels indicated by the dashed lines and cross-sections in (O-R) are through the embryo shown in (N) at the levels indicated by the dashed lines. White asterisks in (A, B, T, U) highlight inconspicuous early expression domains of CYP26-1 and CYP26-3. Arrowheads in (C, V) indicate central nervous system expression and in (K-M, Q) ectodermal signal. Developmental stages shown are: early gastrula (F), late gastrula (A, B, G, H, T, U), mid neurula (C, D, I-M, V, W), very early larva (N-R), early (60h) larva (E, S, X). Scale bars are 100 µm for the whole mounts and 50 µm for the cross-sections.*

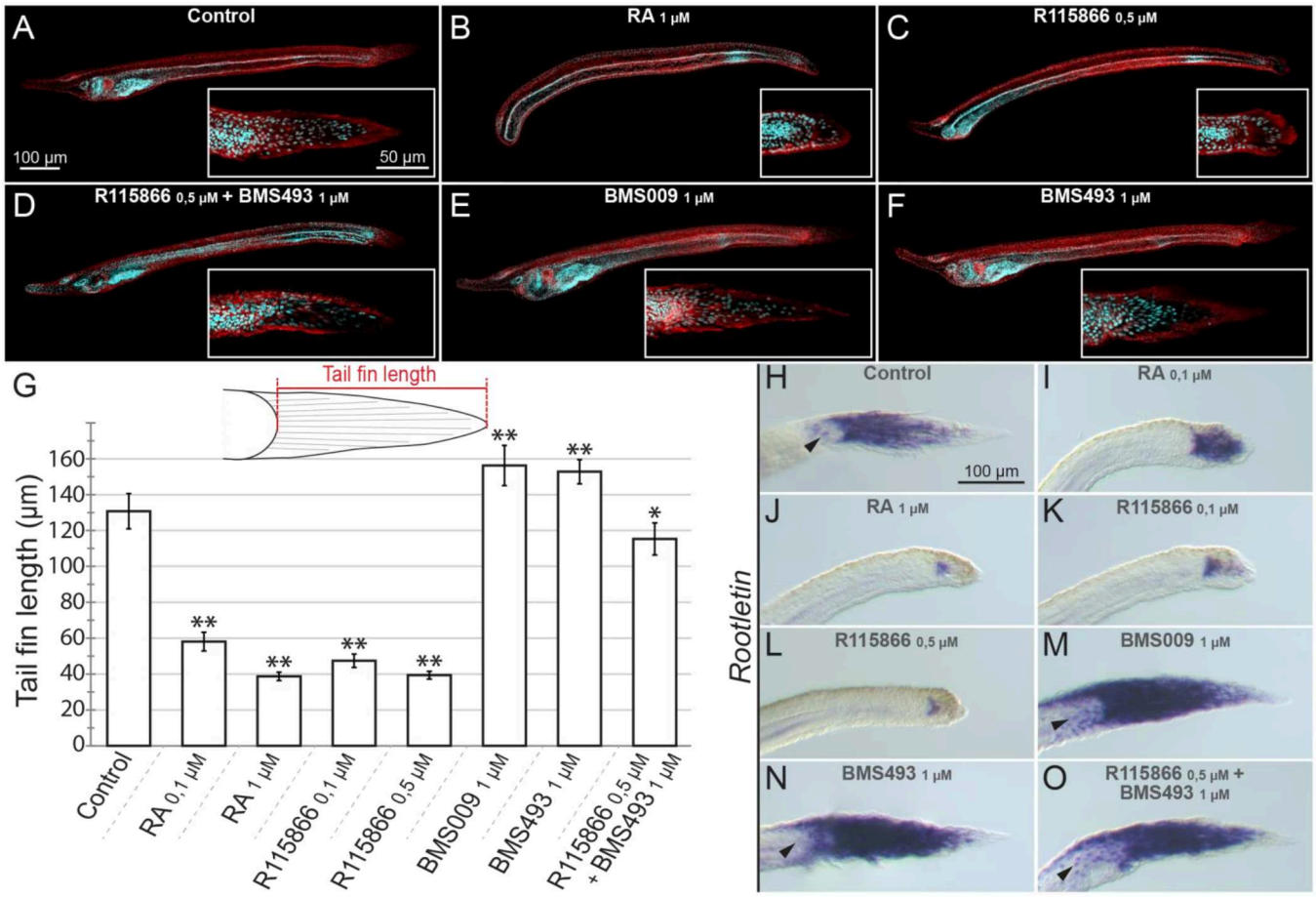


Fig. R2.3 – Effects of retinoic acid (RA) signaling alterations on the development of amphioxus. *(A-F) Early amphioxus (Branchiostoma lanceolatum) larvae at 60h of development with anterior to the left and the dorsal side up are shown as maximal projections of confocal microscopy scans displaying auto-fluorescence (in red) and Hoechst nuclear staining (in cyan). Magnifications of the tail region are boxed, with anterior to the left and the dorsal side up. Scale bars are 100 µm for the larvae and 50 µm for the tail fin. (G) Tail fin length measurements following pharmacological treatments. The length of the tail fin of B. lanceolatum larvae at 60h of development was measured, as indicated in the schematics. The graph shows the tail fin lengths ± standard deviation, with n = 15, for the different treatment conditions indicated. Tail fin lengths that are significantly different from the DMSO control are indicated as "*", for a cutoff value of p=0,005, and as "**", for a cutoff value of p=0,001. (H-O) Rootletin expression in the tail fin of early B. lanceolatum larvae at 60h of development, with anterior to the left and the dorsal side up. Arrowheads in (H, M-O) indicate expression in lateral ectodermal cells. Pharmacological treatments are as indicated. Scale bars are 100 µm.*

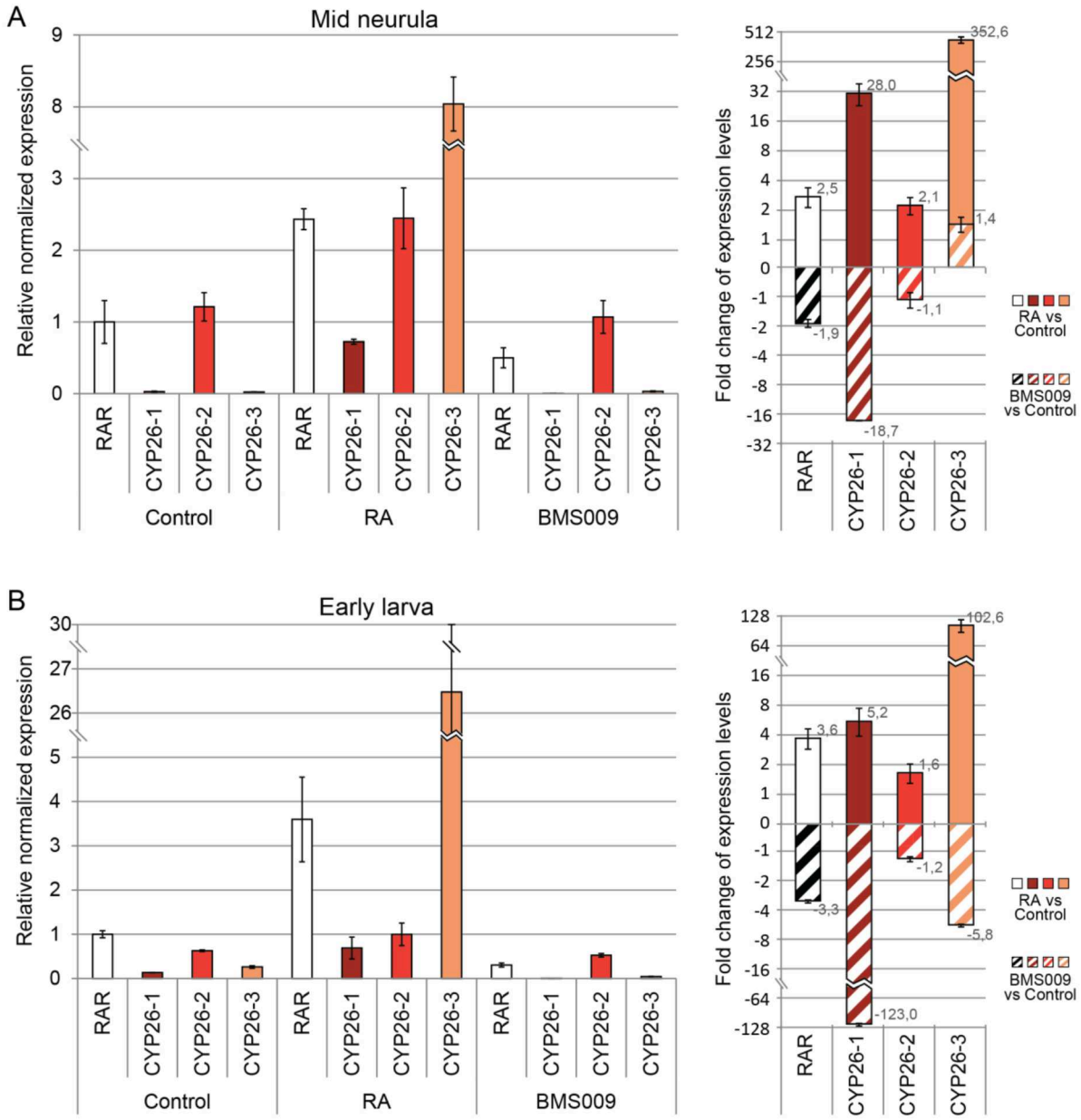


Fig. R2.4 – Quantitative changes of amphioxus CYP26 expression in response to retinoic acid (RA) signaling alterations. *Expression of CYP26-1, CYP26-2, and CYP26-3 in amphioxus (Branchiostoma lanceolatum) was assessed by quantitative real time PCR (qPCR) at two developmental stages (mid neurula at 20h of development and early larva at 48h of development). RAR expression was also established and used as a normalized positive control. (A) Relative normalized expression (left panel) and fold change of expression levels (right panel) at the mid neurula stage, using a Log_2 scale. (B) Relative normalized expression (left panel) and fold change of expression levels (right panel) at the early larval stage, using a Log_2 scale.*

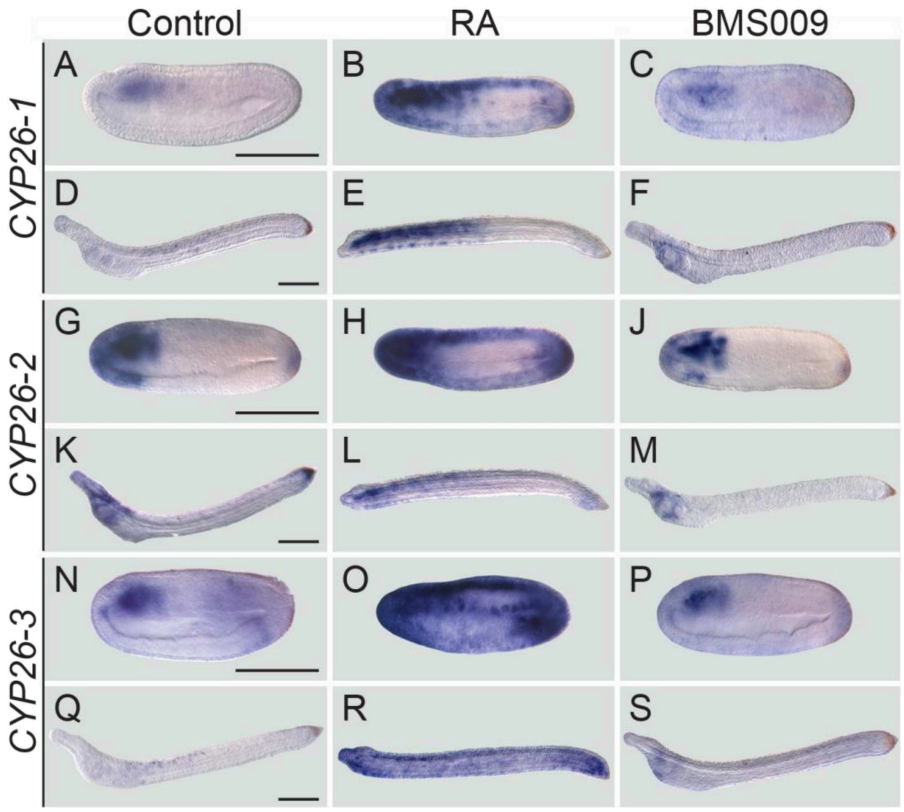


Fig. R2.5 – Spatial changes of amphioxus CYP26 expression in response to retinoic acid (RA) signaling alterations. *Whole mount in situ hybridization experiments were carried out for CYP26-1 (A-F), CYP26-2 (G-M), and CYP26-3 (N-S) on amphioxus (Branchiostoma lanceolatum) embryos treated at the late blastula stage with either DMSO (as control), all-trans RA (1 µM) or RAR antagonist BMS009 (1 µM). The BMS009-treated embryo in (C) and larva in (F) were subjected to an extended coloration period. Mid neurula embryos at 20h of development (A-C, G-J, N-P) and early larvae at 48h of development (D-F, K-M, Q-S) are shown as lateral views with anterior to the left and the dorsal side up. Scale bars are 100 µm.*

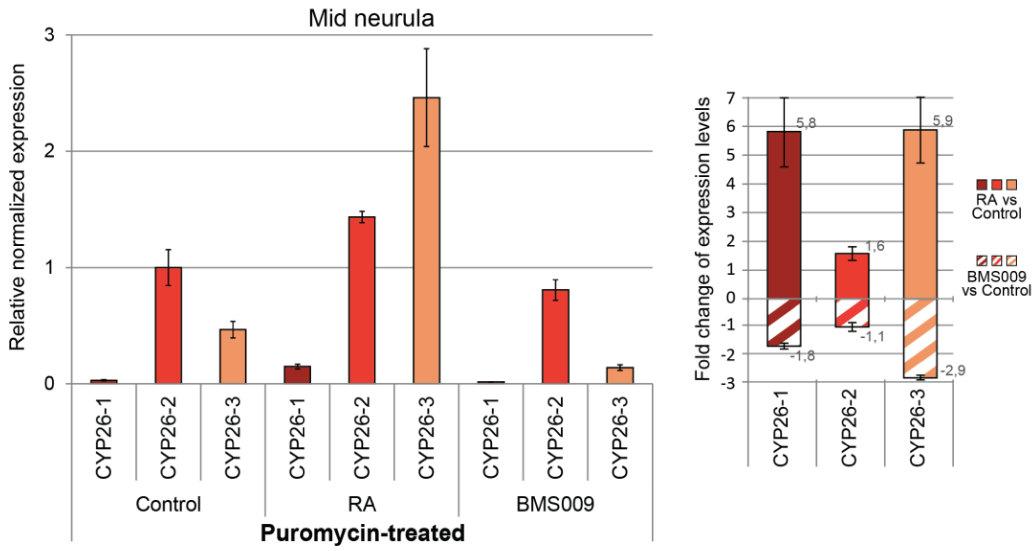


Fig. R2.6 – Dynamics of amphioxus CYP26 gene expression following protein synthesis inhibition and retinoid treatments. *Expression of CYP26-1, CYP26-2, and CYP26-3 was assessed in amphioxus (Branchiostoma lanceolatum) by quantitative real time PCR (qPCR) at the mid neurula stage (at 20h of development) following protein synthesis inhibition by puromycin treatment (200 µg/ml) at 19h of development and subsequent treatment with DMSO (as control), all-trans RA (1 µM) or RAR antagonist BMS009 (1 µM) for one hour. Relative normalized expression (left panel) and fold change of expression levels (right panel) relative to the controls are shown.*

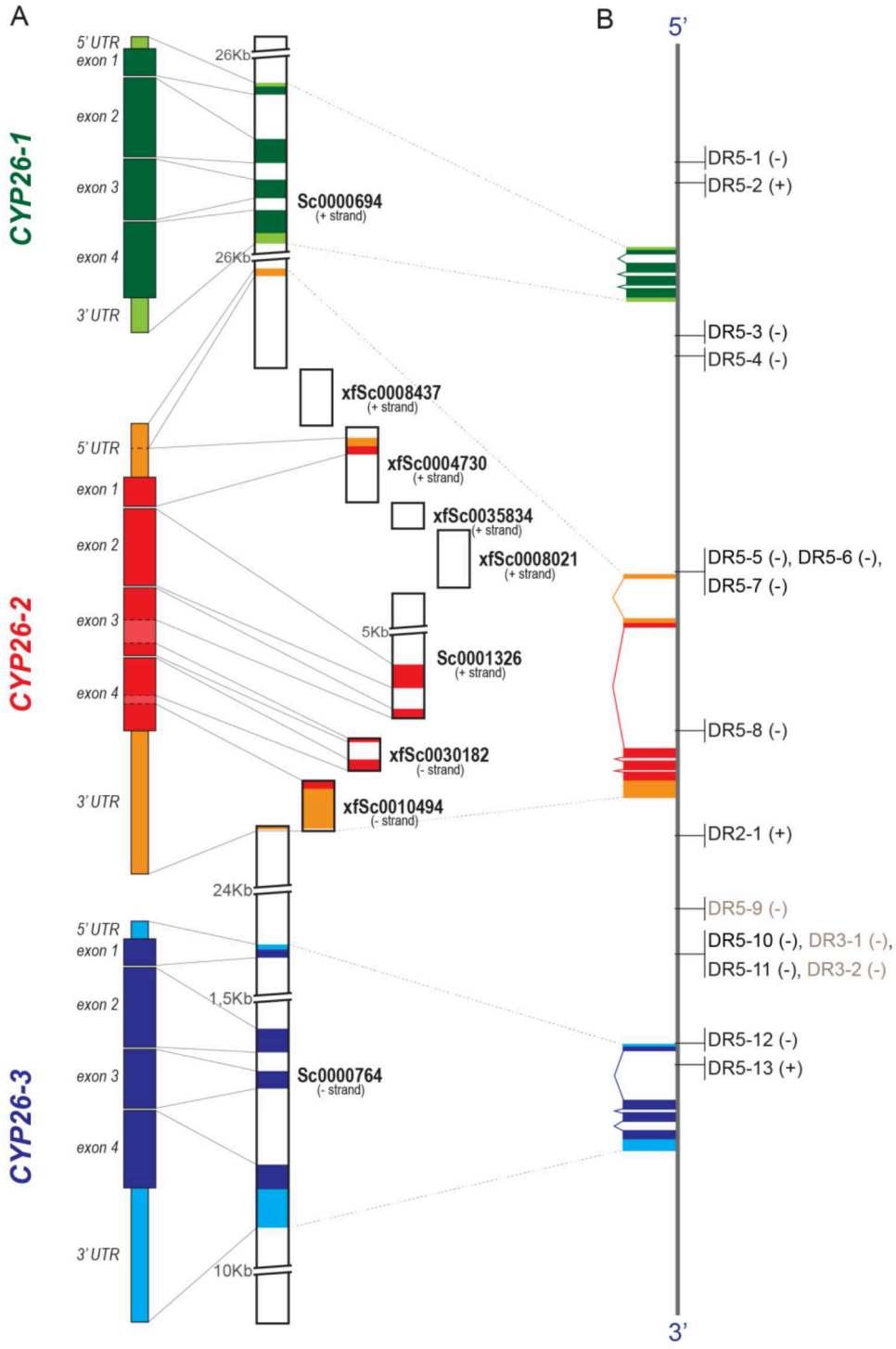


Fig. R2.7 – Organization and distribution of conserved retinoic acid response elements (RAREs) in the amphioxus CYP26 cluster. *(A) Schematic representation of the experimentally verified amphioxus (Branchiostoma lanceolatum) CYP26-1, CYP26-2, and CYP26-3 gene sequences, including the 5' and 3' untranslated regions (UTRs) and the exons of the coding region, as well as of the corresponding scaffolds of the B. lanceolatum genome that cover the complete CYP26 cluster and thus also include intronic and intergenic regions. The names of the B. lanceolatum genome scaffolds are indicated, as is their orientation relative to the CYP26 cluster (+/- strand). (B) Distribution of*

conserved amphioxus RAREs in the B. lanceolatum CYP26 cluster. The orientation of each RARE is indicated relative to the CYP26 cluster (+/-). RAREs recognized and bound by the B. lanceolatum RAR/RXR heterodimer in vitro are indicated in black, RAREs that do not associate with the B. lanceolatum RAR/RXR heterodimer in vitro are shown in grey. The RARE sequences are given in Appendix 2 and the in vitro assays are shown in Fig. R2.8.

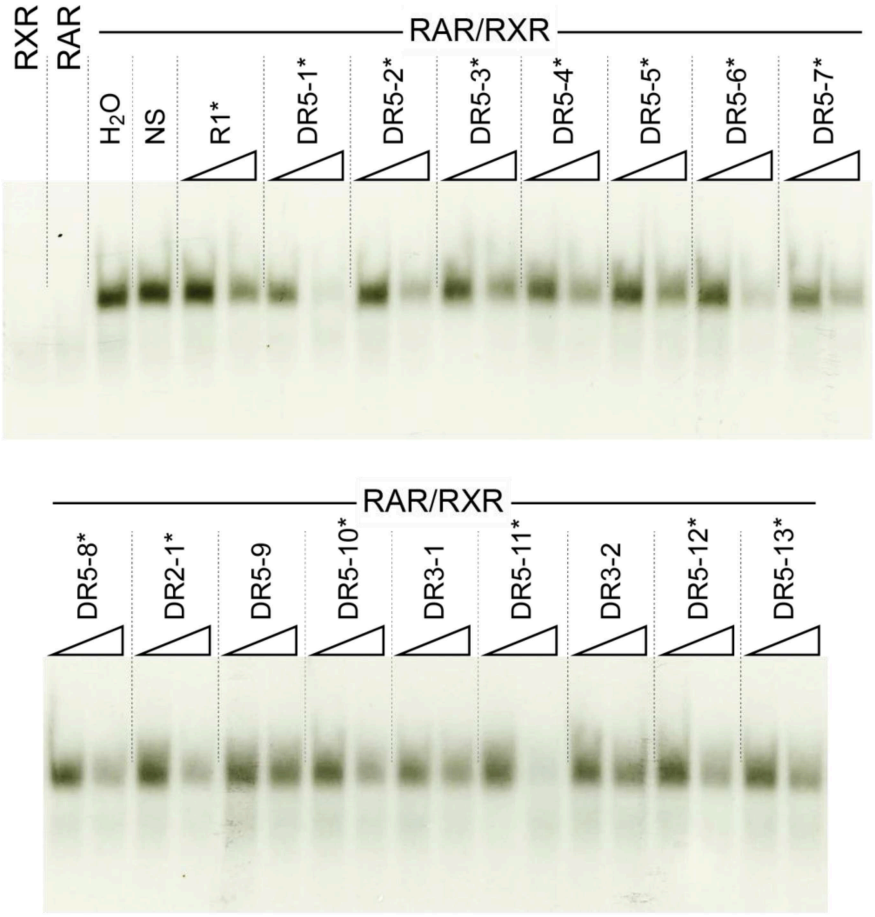


Fig. R2.8 – In vitro validation of putative retinoic acid response elements (RAREs) in the amphioxus CYP26 cluster. *Using in vitro synthesized Branchiostoma lanceolatum RAR and RXR proteins, the binding of the cognate RAR/RXR heterodimer to putative RAREs identified in the amphioxus CYP26 cluster was assessed. The mouse CYP26A1 DR5 sequence (R1)^{500} (Table 1) was used as ^{32}P-radiolabeled double-stranded probe. The RXR and RAR lanes respectively contain only in vitro synthesized RXR or RAR proteins, together with the radiolabeled probe. The H_20 (water) lane contains radiolabeled probe as well as RAR and RXR proteins. The NS (non-specific) lane contains a non-specific and unlabeled double-stranded DR element in 100-fold molar excess relative to the*

radiolabeled probe as well as RAR and RXR protein. All other lanes contain one of the 16 putative amphioxus DR elements (for nomenclature explanations see Fig. R2.7B) at 10-fold (left lane) or 100-fold (right lane) molar excess, along with RAR and RXR protein and the radiolabeled mouse R1 element, to test the binding specificity of the putative amphioxus RAREs. A black asterisk indicates that a given DR element can be recognized by the B. lanceolatum RAR/RXR heterodimer, as it outcompetes binding to the mouse R1 element.

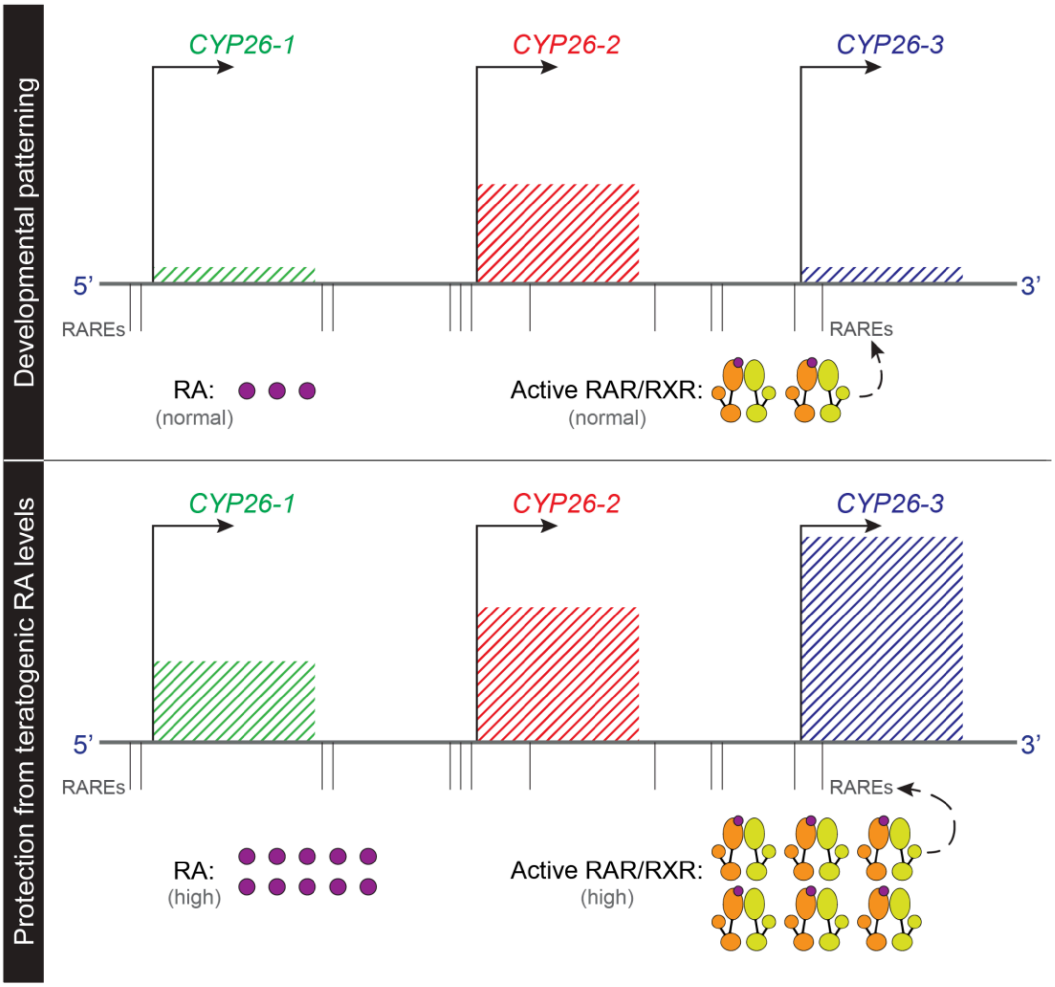


Fig. R2.9 – Model for regulation and function of CYP26 genes in amphioxus. *The position of the functionally validated retinoic acid response elements (RAREs) within the amphioxus CYP26 cluster is indicated relative to the start codons of three amphioxus CYP26 genes (CYP26-1, CYP26-2, and CYP26-3). The relative expression of each of the genes is indicated during normal development (normal levels of RA and active RAR/RXR receptor heterodimers) as well as during exposure to RA (high levels of RA and active RAR/RXR receptor heterodimers), hence highlighting the dominance of CYP26-2 during normal development and the importance of CYP26-1 and CYP26-3 for protection against RA teratogenesis.*

R3: RNAseq and ATACseq data reveals new targets and mechanisms of action of the RA-signaling pathway in amphioxus development [4]

Introduction

The vitamin A bioactive metabolite, retinoic acid (RA) is a well-known regulator of embryonic development, organogenesis, cell proliferation, cell differentiation and tissue homeostasis[232]. Globally, the distribution of RA in a developing embryo is controlled by specific enzymatic pathways that in accordance with the receptors of RA generate intricate regulatory loops. For instance, RA signaling is known to directly control the expression of RAR as well as to stimulate directly the expression of RA catabolizing enzymes from the Cytochrome P450 subfamily 26, denominated CYP26s[495]. RA exerts its highly diverse organismal action through specific nuclear receptors (NR), classically the heterodimer RA receptor (RAR) and retinoid X receptor (RXR)[237]. Although, in vertebrates RA can also bind with high affinity to other members of the NR family, for instance PPARβ/δ (peroxisome proliferator-activated receptor)[547,548], and with a lower affinity with RORs[549], COUP-TFs[324] and TR2/4[323]. Nevertheless, the biological implications of these non-classical RA receptors during development remain elusive.

Typically the multiple RAR/RXR heterodimers, generally represented by combinatory binding of one the three RARs (RARα, RARβ and RARγ) with one the three different RXRs (RXRα, RXRβ and RXRγ), act as ligand dependent regulators of transcription with a central DNA binding domain linked to a ligand binding domain[550]. In the canonical model for the action of RAR/RXR heterodimer, this heterodimer constitutively binds specific response elements (REs) located in of promoters of target genes. Retinoic acid response elements (RAREs) are composed of two direct repeats of a core hexameric motif (A/G)G(G/T)TCA. In vertebrates, the classical RARE is a 5 bp spaced direct repeat (referred as DR5), but RXR/RAR heterodimers can also bind to direct repeats separated by 2 bp (DR2), 1 bp (DR1)

[4] Article in preparation. Authorship: João E. Carvalho, Sandra J. Gancedo, Ferdinand Marlétaz, François Lahaye, Hector Escriva, Jenifer C. Croce, José L. Skarmeta, Michael Schubert.

or 0 bp (DR0)[310,494,551,552]. In the absence of ligand, bound RAR/RXR associate with a set corepressors, which serves as adaptors for complexes that have an histone deacetylase activity, ultimately leading to chromatin condensation and prevention of gene transcription[553]. Additionally, DNA regions occupied by RAR/RXR heterodimer in the absence of the ligand are also coated with the repressive complexes of the polycomb group proteins that by deposition of repressive marks induce the condensation of the chromatin[554]. Contrasting with the classical view, recent genomic studies revealed that only a small subset of RAR/RXR binding sites are constitutively occupied in the absence of ligand and that RA treatment induces widespread changes in RAR/RXR binding[312,555]. Nevertheless, in the presence of ligand, pre-existing as well as de novo DNA recruited RAR/RXR are associated with coactivators, which have an intrinsic histone acetyl transferase activity and thus facilitates the recruitment of the transcription machinery. These epigenetic changes have very broad impact in downstream signaling cascades and in the deployment of specific transcription factors (TF), such as Hox[349,556]. Furthermore, it has been suggested that TF cooperativity is an inherent feature of TF-DNA binding[557,558], and that many genomic loci are bound by a large number of different TFs[559–561] that could have divergent roles in regulating gene expression, i.e., they contribute to activation on some genes but repression on other genes[561]. Within those groups of interacting TF, we can identify also hotspots of NRs binding motifs[559,562].

Despite the large quantity of data obtained in vertebrates in terms of the RA signaling targets[312,493,556,563–565], so far no study has established an hierarchical organization able to point which of those are affected from a direct point of view and which are downstream effectors of different levels. In vertebrates, several combinations of the RAR (α, β and γ) with the RXR (α, β and γ) conduct spatio-temporally the different steps of RA-dependent development[317]. This very organized and defined process also requires several other players, like the enzymes involved in RA synthesis and degradation, which control the bioavailability of the ligand at the right time and at the right place, as well as other signaling pathways that interact synergistically or antagonistically with RA signaling pathway[317]. Due to this highly complex multilevel property of the RA signaling pathway in vertebrates, the use of a model that possesses a proxy of this signaling cascade is fundamental. In the invertebrate chordate amphioxus, RA signaling pathway is controlled by a single RAR/RXR heterodimer and thus allowing the study of this signaling pathway in an organismal context, making it easier to establish a framework that characterizes the hierarchy of targets genes during development. Albeit controlled only by single RAR and RXR genes, RA signaling in amphioxus seems to

be controlled by a vertebrate-like gene hierarchy, as well as it drives developmental functions similar to those in vertebrates[232]. So far in amphioxus, RA signaling has been studied from a candidate gene perspective and some direct targets of RA signaling have been identified (e.g. *RAR*, the Hox genes *Hox1* and *Hox3*, *FoxA-b* and the 3 *CYP26* genes)[432]. Furthermore, knockdown of *Hox1* showed that RA acts via *Hox1* to establishing the posterior limit of the pharynx[430], to specify motor neurons as well as in the patterning of the amphioxus central nervous system (CNS)[427]. Indirect targets include for instance *Otx*, *Pax1/9*, *Pax2/5/8* and *Wnt3*, although quite some genes described as modulated by RA signaling have not clearly been validated as direct or indirect targets, for example *Hox6*, *Wnt5*, *Xlox* and *Cdx*[427,430–432]. However, not much is known about other RA signaling targets during amphioxus development, nor about the potential variety of TF partners that controls or interacts with this pathway.

Here, we first characterized how different retinoid-dependent treatments (RAR antagonist, RA and CYP26 inhibition) induce transcriptional changes in the developing neurula of amphioxus (N0, N3 and N4 stages) by RNA high throughput sequencing (RNAseq). We then correlated the different treatment results and characterized them based in their described molecular function, revealing a strong link between RA signaling modulated targets with transcription factor and metabolic as well as regulatory functions. In order to establish potential genes being directly or indirectly regulated by the intermediary of the RAR/RXR heterodimer, we systematically predicted the occurrence of REs (ranging from DR0s to DR5s) in the vicinity of the modulated set of transcripts and thus considering DR2 and DR5 as proxy for direct regulation mediated by this heterodimer of NRs. We further extended these findings by characterizing the open regions of the chromatin induced by RAR antagonist and RA treatment at the N3 stage via Assay for Transposase-Accessible Chromatin combined with high-throughput sequencing (ATACseq) and identified the enriched TF binding motifs in differentially represented peaks. Thus, we verified that these motifs enriched represented binding motifs for NRs, Forkhead box (Fox) and basic helix-loop-helix (bHLH) factors specifically in over-represented peaks and homeodomain, high-mobility group (HMG) from the Sry box (Sox) family and even the nuclear receptor NR2F6 binding motifs were specifically enriched in RA-dependent under-represented peaks.

We finally intersected RA modulated targets with ATACseq peaks assigned genes to identify potential directly regulated genes marked by neighboring regions of open chromatin, hence we identified a list of potential 80 genes that fit these criteria. The open regions of the chromatin around these genes exhibits highly enriched NRs binding sites and in particular

DR5 elements, typical signatures of RAREs, as well as multiple Fox binding sites. Strikingly, we verified a very strong inverse correlation between the central position of NR response elements and the more lateral localization of Fox binding sites, thus suggesting a potential interaction between these two families of fundamental TFs into regulating gene specific transcription. In sum, here we show very robust results of the retinoid induced transcriptional landscape changes and we identify a mechanism that could be playing a major role in the regulation of RA-dependent epigenetic changes during chordate development.

Results

Transcriptional landscape of the amphioxus developing neurula is affected by retinoid-dependent treatments

In order to obtain a comprehensive vision of the RA dependent transcriptional landscape during the development of amphioxus we designed an RNAseq-based set of experiments. We characterized the transcriptional profiles of 3 different neurula stages of amphioxus under 4 different conditions, DMSO control that was used to compare changes induced by pharmacological treatments with BMS493, an antagonist of RAR, RA and R115866, an inhibitor of the CYP26 activity and thus increasing locally the level of RA (Fig. R3.1 A). The treatments were applied to the developing amphioxus embryos at the onset of gastrulation (G1 stage) that were exposed to the pharmacological compound for a period of 6, 12 or 18h accordingly to the stage of development studied: N0, N3 and N4 respectively (Fig. R3.1 A). Total RNA was extracted and poly-A dependent libraries were generated for Illumina high-throughput sequencing (pair-ended 50 bp), yielding an average of 61,3 million reads per sample (min 51,8 million and max 76,6 million reads). Reads were then mapped against the gene models for *Branchiostoma lanceolatum*.

Differential expression analysis of the mapped reads, comparing each of the treatment conditions against the control, within each of the time points studied, was filtered with a threshold of 0,05 for the adjusted p-value obtained. Furthermore, to avoid a bias associated with a low count effect, genes with a base mean expression value inferior to 50 were not considered in the final dataset. Under these principles a given gene was considered as modulated by the treatment (BMS493, RA or R115866) if its expression level was at least 50% higher or lower when compared to the control condition (log_2 fold change of at least +/- 0,585).

At the N0 stage, we verified that most of the transcripts modulated by the action of the BMS493 treatment were being downregulated, while most of the transcripts affected by either exogenous RA or R115866 were being upregulated (Fig. R3.1 B). Accordingly, later in development, at the N3 stage, we observed a great amount of transcripts that were up and downregulated by all the treatments. The same tendency as in the previous developmental stage in terms of distribution of regulation by the different treatment was observed, although the fold change exhibits values that are much higher or lower when compared to the previous stage (Fig. R3.1 B). As expected, BMS493-treated embryos exhibited a high number of downregulated targets, contrasting with the important number of transcripts upregulated by RA and R115866 (Fig. R3.1 B). Then at the N4 stage, the number of both upregulated and downregulated targets in all conditions increased and also their relative variance between the lowest and highest fold change, even though, the overall tendency associated with each treatment was maintained.

Considering the modulation of genes during development, the pharmacology treatments used here had a gradual cumulative effect, since of all the different 1244 transcripts scored as being differentially expressed in the entire experiment, 209 were regulated at the N0 stage, 849 at the N3 stage and 1188 at the N4 stage (Fig. R3.1 C). Furthermore, our results clearly showed that while RA treatment had a rather strong early effect, with 202 transcripts being modulated at the N0 stage, followed by a rapid increase of these numbers to 808 and 872 targets in the following stages, the effect of both BMS493 and R115866 were more gradual (Fig. R3.1 C). Treatment with BMS493 initially influences 28 transcripts, followed by 342 at the N3 stage and 635 targets by the later stage, thus rather similarly R115866 modulates the expression of 49 targets at the N0 stage, 220 at the N3 stage and 777 transcripts at the N4 stage (Fig. R3.1 C). Globally, of all the modulated targets, RAR antagonist regulated transcripts represented 57,4%, 84,5% were regulated by exogenous RA and 64,8% were being modulated by the action of CYP26 inhibitor treatment (Fig. R3.1 C). By comparing the number of targets that were shared by either BMS493 and RA or RA and R115866 treatments, we observed that at the N0 (25 and 45 respectively) and N3 stage (314 and 202 respectively) these number were rather similar to the number of individually modulated transcripts for BMS493 or R115866, respectively. Although, at the N4 stage we verified that each individual treatment exhibited a larger number of targets that were specifically induced. In our survey, 43,1% of all targets identified were modulated by both BMS493 and RA treatment, and 51,8% were differentially regulated in the context of RA and R115866 treatments.

When analyzing the results from a treatment perspective, most of the transcripts that were modulated as early as the N0 stage were concomitantly modulated later during development (Fig. R3.1 D). Differently, for all the treatment regimes there were quite some targets differentially expressed only at the N3 or at the N4 stage and also a large number of transcripts that were regulated in all the three developmental stages characterized. For instance, BMS493 treatment regulated specifically 76 transcripts at the N3 stage and 369 at the later stage. Under this condition, 23 targets were commonly modulated in the N0, N3 and N4 stages while 241 transcripts were differentially regulated in the N3 and N4 stage. Treatment with exogenous RA specifically regulated 146 targets during the N3 stage and 237 at the N4 stage, while R115866 treatment modulated exclusively 26 and 582 transcripts at the N3 and N4 stages, respectively (Fig. R3.1 D). Furthermore, RA commonly controlled the expression of 167 genes in all the three developmental stages and 466 different transcripts were modulated in both N3 and N4 stage by the action of RA treatment. R115866 treatment, by his turn, differentially influenced the expression of 39 and 151 genes at all the three stages or at the two later stages, respectively. Of all BMS493 and RA treatment shared targets, 95 and 220 are exclusively regulated at the N3 and N4 stage, respectively. 21 were commonly modulated in all the developmental stages studied and 196 transcripts were differentially controlled in both the N3 and N4 stages. Consistently, 38 and 437 transcripts shared between those regulated by RA and R115866 treatment were specifically regulated at the N3 stage for the former and at the N4 stage for the latter, as well as 28 targets were jointly modulated in all developmental stages and 125 only in N3 and N4 stages. Globally, the majority of the targets identified were modulated in the N3 and N4 stages (i.e. 598 transcripts), although 195 targets were differentially regulated in all the developmental stages. 46 and 391 transcripts were specifically regulated at N3 or N4 stages, respectively. Altogether, these data shows that the transcriptional landscape of the amphioxus developing neurula is differently affected by each of the retinoid-dependent treatments applied. Furthermore, the response to these pharmacological treatments possesses a strong component associated with the specific stage of development, namely N0, N3 and N4 stages.

Modulated genes are involved in metabolic processes, DNA binding, and chromatin rearrangements

To assess the relationship between differentially regulated targets for each of the three different treatments and the three different developmental times points, we further calculated

the overall association between different set of genes and defined clusters of modulated transcripts based on their expression profiles (Fig. R3.2 A). The clusters defined greatly vary in terms of the gene number, showing that some tendencies are more robustly associated than others. Comprehensively and as expected, the defined clusters denoted a strong inverse correlation between BMS493 treatment and both RA and R115866 treatments. Despite the differences in terms of range between the max in min value for the fold change between the clusters 1–7, they roughly represented transcripts that were affected similarly. Globally, in clusters 1–7 BMS493 regulated targets were mostly downregulated at least during one time point, while R115866 treatments were mostly upregulated at one of the developmental stages, similarly RA treatment upregulated all the transcripts in these clusters more strongly at the N3 stage than at the N4 stage (Fig. R3.2 A). Accordingly, in clusters 8–10 we observed similar profiles verified in clusters 1–7 for both BMS493 and R115866 treatments, although RA treatment denoted a stronger influence in targets at the N4 stage than at the preceding developmental time points, the N0 and N3 stages. Differently, in the clusters 11–13, transcripts influenced by RA and R115866 were mostly downregulated, while RAR antagonist treatment, with minor exceptions, upregulated the set of genes grouped in these clusters (Fig. R3.2 A). Cluster 14 represents a special case where most of the genes were not differentially regulated by BMS493, although both RA and R115866 induced upregulation of these transcripts. Cluster 15 included the majority of the genes that did not behave accordingly to the inverse correlation between BMS493 treatment and both RA and R115866 treatments, since the transcripts included in this cluster were to be globally upregulated by all the three treatments. Last but not least, grouped in the cluster 16 we found transcripts that were upregulated by R115866 treatment and some of those were also upregulated by BMS493 treatment, while RA treatment have only minor effects in this gene cluster (Fig. R3.2 A).

When classifying the gene ontology (GO) affinities of the transcripts for each cluster, we found that two main groups of represented GO terms are recurrent: catalytic activity and binding. Our results show that within the catalytic activity group of terms both the transferase activity and hydrolase activity sub-terms were highly enriched. For instance, terms like transferase activity relative to acyl groups and N-acetyltransferase as well as carbohydrate and sugar phosphatase activity are highly enriched in our set of regulated transcripts (Fig. R3.2 B). Importantly, the enriched sets included components of retinoid-related metabolism (e.g. *DHRS*s, *RDH10*, *ALDH1A* and *CYP26*s). Furthermore, the molecular function of binding was also enriched, specifically with terms associated with regulatory regions (Fig. R3.2 B). Even though binding related terms were less enriched than the terms that refer to catalytic activity,

several of these groups contain essential members associated with key developmental processes and chromatin rearrangements. For instance, we identified BAHD1 (bromo adjacent homology domain-containing 1), ACTB (actin beta), a large set of Homeobox domain containing genes (e.g. *Hmx*, *Hnf1, Hox*s, *Meis*, *Pbx*, etc.) and several others TFs (e.g. *TBX6/16, Sox7, etc.*) (Appendix 3).

Globally, our GO results thus suggest that the genes being modulated by retinoid-dependent treatments were associated with metabolic processes, including retinoid metabolism, and TF binding sites as well as regulatory regions that are responsible for mediating a great variety of molecular processes during development, such as the regulation of transcriptional activity and chromatin remodeling.

Presence of putative response elements is associated with complex regulation of transcriptional activation

Genes that are directly regulated by RA signaling often have in their vicinity or in their promoter regions one or even several RAREs[440,510,566]. To establish a potential framework that distinguishes between modulated direct targets and indirect or downstream effector genes, we next searched systematically for the presence of putative REs (DR0, DR1, DR2, DR3, DR4 and DR5) in the upstream and downstream vicinity (10Kb range) as well as in intragenic regions for all the 1244 transcripts modulated by the three different treatments. By doing so, we identified 387 out of the 1244 that do not have any predicted RE in their neighboring region, and we were able to localize a single RE in the vicinity of 340 genes. For the remaining transcripts we identified at least 2 potential REs: 242 genes with 2 REs predicted, 127 transcripts with 3 putative REs and 148 that have 4 or more predicted REs, with 22 being the maximal number of putative REs nearby a single gene (Fig. R3.3 A).

Globally, the putative REs identified are in majority of the type DR1 (i.e. about 33%), followed by DR0 and DR5 elements which, together, had about the same number (i.e. 17-18% each) and then DR4, DR2, and DR3 (14%, 10% and 8%, respectively). Furthermore, no major deviation of the global distribution of the type of response elements was found when focusing only on the number of REs per gene, i.e. considering 1, 2, 3 or 4 and more REs identified per genes (Fig. R3.3 B).

Then, we established a correlation between the type of putative RE and its location towards the respective transcript, and thus we classified them into 3 categories: upstream (up to 10kb upstream of the transcript), downstream (up to 10kb downstream of the transcript)

and intragenic (between both regions aforementioned). Globally, the results showed that all the REs studied are more or less equally distributed in the three categories defined: 31,15% were situated upstream, 33,47% were located in intragenic regions and 35,38% were found upstream of the targets genes (Fig. R3.3 C). Proceeding with the same type of analysis for each of the individual sets of response elements (i.e. DR0, DR1, DR2, DR3, DR4, DR5) we verified that only DR5 elements have a deviation from the global distribution that is statistically significant (χ^2 test, P-value> 0,05). We find 40,53% of DR5 putative elements in upstream domains, 37,28% in intragenic regions and 22,19% of these elements that are specifically located downstream (Fig. R3.3 C).

We also defined the concept of potential response elements proximal hotspots as the presence of at least two predicted REs within a maximal span of 100 bp. These proximal hotspots containing REs represent potential regions of interaction between two heterodimers of NRs. Surprisingly, most of the hotspots identified had a span inferior to 30bp. We identified 62 regions that had putative response elements proximal hotspots of multiple configurations. Interestingly, like for the DR5 elements, these REs hotspots also exhibited a deviation from the global distribution of the putative response elements that was statistically significant (41,33% upstream, 22,67% intragenic and 36% downstream) (Fig. R3.3 C). More than half of these hotspots included the DR1 motif and we identified the composite DR1-DR1 in 16 out of 62 hits. Nevertheless, several other combinations of RE were observed, such as the composite hotspots including only DR0/DR2/DR5s permutations with 13 occurrences, the combination of DR1s with DR0/DR2/DR5s (13 domains identified), the close proximity of DR3s and DR0/DR1s that accounted for 4 regions, as well as 11 hotspots that had DR4 elements in close proximity with DR0/1/2/5s, a single RE hotspot that was exclusively composed by DR3s and 4 that included only DR4 composites (Fig. R3.3 D).

These results suggested that the expression of almost a third of the modulated genes (i.e. 387 of 1244) is probably not directly regulated by the action of the RAR/RXR receptor since they did not have any RE in their vicinity, although we cannot exclude a potential long range mode of action[516]. Moreover, for what concerns the other two thirds of modulated genes, the presence of one or more putative response elements indicated a very complex system based in interaction between different nuclear receptors in a tissue context environment. Considering that only DR2 or DR5 elements are generally bound by the amphioxus RAR/RXR heterodimer (Carvalho et al. in press), these data suggest that about 30% of all transcripts modulated might be bound, and thus directly regulated. Moreover, 5% of all transcripts modulated have REs hotspots, of which 41% contain at least one DR2- or

DR5-type RAR/RXR binding site indicating that the transcription of some genes modulated by RA is possibly controlled by the interaction between RAR/RXR and others NRs.

RA signaling changes drive chromatin remodeling

NRs and in particular the couple RAR/RXR is well known to function in recruiting coactivators upon ligand biding which will modify the chromatin structure surrounding the promoter of the target genes, leading to gene specific transcription[317]. Thus, to address how the RA transcriptional landscape induces changes that translate in chromatin remodeling during amphioxus development, we characterized the differentially represented regions of open chromatin in BMS493 and RA treated embryos at the N3 stage, using ATACseq (Fig. R3.4 A). In our analysis we decided not to consider enriched or depleted peaks that overlap with the transcription start sites (TSS), since this could create a bias in our analysis due to the increased transcriptional activity of a given set of genes. Our results showed that, for instance in the neighboring region of known direct targets of RA, such as *RAR* and *CYP26*s, there was significant over-representation of ATACseq peaks following RA treatment (Fig. R3.4 B and Appendix 3). Some of these specific peaks overlapped with the location of functional validated RAREs, at least in the case of the *CYP26* cluster domain (Carvalho et al. in press). Globally, when comparing ATACseq peaks between RA treated embryos and control ones we identified 436 peaks that were either over-represented, 358 peaks, or under-represented, 78 peaks. Not surprisingly, BMS493 treatment yielded much lower numbers of differentially represented peaks: 10 over-represented peaks (ORP) and 2 under-represented peaks (URP) (Fig. R3.4 C).

We next investigated the potential enriched binding motifs present in both RA ORPs and URPs to demonstrate the regulatory elements being controlled and recruited by RA activity. Of note, considering the low amount of differentially represented peaks obtained for embryos treated with BMS493 (i.e. 12 peaks in total) no statistical analyses were possible, thus the prediction of binding motifs analysis could not be performed for this set of BMS493-dependent peaks. In RA ORPs, our results showed a specific enrichment of binding sites of the RAR/RXR couple (i.e. DR5-type RAREs) and potentially for other NRs, since DR half-sites were also highly enriched (Fig. R3.4 D). Accordingly to the complete set of results enriched DR half-sites could correspond to binding motifs for: NR2F1 (COUP-TF), ROR (NR1F1 or RAR-related orphan receptor), ESR (estrogen receptor) and for instance also for the ultraspiracle (usp) receptor of *Drosophila* (i.e. RXR related NR), suggesting the presence

of a wide variety of potential NR biding motifs. Additionally, in RA ORPs we found the enrichment of binding sites specifically for the winged helix-turn-helix Fox family, e.g. *Drosophila* fork head gene (fkh), FoxA1, FoxB1 and FoxP1, as well as for basic helix-loop-helix (bHLH) factors, like insulin-degrading enzyme (IDE) and transcription factor (TCF) 3 and 4 (Fig. R3.4 D). Differently, in RA URPs we identified the enrichment of motifs associated with homeodomain containing factors (e.g. POU6F1, EMX, LHX and related factors like the *Drosophila* ventral veins lacking – vvl), high-mobility group domain factors, mainly from the Sox family as well as the nuclear receptor NR2F6 (Fig. R3.4 D).

In order to correlate the ATACseq results on chromatin induced changes with the differential regulated genes of the RNAseq analysis, we assigned to each differential represented peak (ORPs and URPs) potential genes influenced. To do so we established a search window that extended 10kb up and downstream of each peak and every gene's ORF present in that region will be considered an assigned gene. In total and considering both conditions (RA and BMS493 treatment) 1043 genes were assigned to our ATACseq differentially represented peaks. The 358 RA ORPs corresponded to 752 assigned genes, to the 78 RA URPs 290 genes were assigned, BMS493 ORPs matched to 23 potential genes and 3 genes were assigned to the BMS493 URPs (Fig. R3.4 C). By intersecting all the ATACseq peaks assigned genes with the RA modulated targets at the N3 stage and at the N4 stage, we identified 80 common transcripts that corresponded to 129 peaks (Fig. R3.4 E). This subset of peaks reflected regions that were being directly influenced by the chromatin conformational changes induced by RA signaling and that had a direct influence on the transcription level of a neighboring gene.

NR binding sites and Fox factor motifs strongly associate in RA-induced chromatin open regions

Following the association of potential RA target genes with RA-dependent regulated regions, we further reanalyzed the enriched motifs in the subset of peaks obtained from the intersection between RNAseq targets and ATACseq assigned genes (Fig. R3.4 E), hence identifying potential interacting TFs that directly mediate RA signaling. Not surprisingly, we were able to identify a highly enriched set of RAR/RXR binding sites of the DR5 type or other NR half-sites in 78 out of the 129 peaks (i.e. 63 assigned genes) (Fig. R3.5 A). Surprisingly, though, we also found a specific enriched binding motif generally associated with Fox transcription factors, with a total of 39 peaks that had a positive match for Fox-

related binding sites (i.e. 37 assigned genes) (Fig. R3.5 A). Furthermore, when verifying the distribution of these predicted biding sites in the sequence of the peaks, we found a clear positional inverse correlation between the NR response element location and the Fox binding sites. Fox biding sites were more likely found at a distance of about 100bp of the center of the peak, while NR response elements were mostly located at the central region of the peaks (Fig. R3.5 B). These results clearly revealed the existence, in close proximity of functional RAR/RXR binding sites, of binding motifs for Fox transcription factors, which is strongly suggestive of a regulatory interaction between Fox and RA signaling.

Discussion

Robust data supports retinoid influence on transcriptional landscapes

Here, we explored the interaction of transcriptomics (RNAseq) and genomics (ATACseq) to reveal new mechanisms of action of the RA signaling pathway during the development of amphioxus. For the first time in amphioxus, we demonstrated at the scale of the entire developing embryo that RA signaling plays its roles by interacting with a multitude of targets genes. By employing an unbiased pharmacology-dependent approach, addressing both the transcriptomic landscape as well as the chromatin permissiveness, we established a list of potential direct RA signaling targets and we identified a new mechanism that might play major roles in the regulation of RA-driven transcription.

Despite some minor drawbacks of the approaches chosen, we were able to set a new standard for this kind of high throughput sequencing studies and obtained a very robust set of final results. First, in our study, the choice of using entire embryos of amphioxus instead of dissected specialized tissues as a very technical component associated with the reduced size of the embryo. Nevertheless, we estimated that despite the potential increase of the ratio signal/noise, our data would still reveal key aspects of both the genes being targeted by the RA signaling during development and the regulation of such a set or subset of genes. Also, by doing this kind of combined analysis we aimed at establishing a framework approach that will, at the expense of the details in local vs global regulation, lay the ground for the evolutionary study of the mechanisms that govern RA-driven transcription in an entire developing organism.

Globally, very stringent sets of thresholds were applied throughout the different statistical analyses, to avoid misinterpretation of our large dataset and thus to integrate our

data into a biological context. Although this choice of criteria can mean the loss of some confirmed positive results, it avoids the presence of false positives in the final dataset and thus strengthens the robustness of the results. Thus few genes that have been shown in other amphioxus studies to be regulated by the RA signaling pathway were not reteived by our final dataset[427,430–432]. For instance, *FoxAb* had been previously shown to be affected by RA treatments in the amphioxus neurula stage and considered as a direct target[430], however in our final dataset we could not retrieve this gene. Although, when looking more carefully in the raw data, we identified *FoxAb* as a gene highly expressed and modulated by both RA and CYP26 inhibitor, but the values of the $\log_2$ fold change of the overall expression level was lower than the threshold applied (Appendix 3). Similarly, and despite the lack of information concerning relationship between *FoxAa* and RA, this highly expressed gene is also under the control of RA signaling but once again the $\log_2$ fold change in expression induced by the treatments is beneath the threshold (Appendix 3). Two more genes, *Cdx* and *Otx*, that were before established as targets of RA signaling[430,431] were missing in our final list of genes, and, as for the two amphioxus *FoxA* genes, this fact is associated with the cutoff value implemented to the $\log_2$ fold change. Interestingly, previously measured fold change for *Cdx*, upon RA and RAR antagonist treatment, were in accordance with the fold change results obtained[431]. Another described target of the RA signaling pathway that was not included in the final dataset was the gene *Hox6*[427], since this gene was considered as lowly expressed (base mean value lower than 50) and therefore removed from the final analysis, even though it is highly upregulated by RA treatment. By doing this comparison of the known RA targets with the raw list of genes we were able to identify and clearly understand why these genes are not present in the final dataset. Importantly the non-included genes mentioned above behaved like what have been described in the literature. All in all, these indications of potentially regulated genes that are not included in our final RNAseq dataset are reassuring, mainly since we can identify the reasons why these genes did not make it through our analysis. Hence, our choice of criteria was very conservative and the final results can be regarded as extremely robust.

Pharmacology-based interference with amphioxus RA signaling: a transcriptomic perspective

Our RNAseq results partially translate the intrinsic property of the RA signaling pathway controlling a large variety target genes during development[232,335]. Comparing the different pharmacological treatment regimes, we have observed that both RA treatment and

CYP26 inhibition, and thus the tissue-specific increase of RA, generated similar profiles of differentially regulated fold changes that could be grouped into specific cohorts of highly or lowly up- or downregulated genes. In contrast, treatment with RAR antagonist globally induced the opposite effect of what is observed for RA treatments and CYP26 inhibition. Even though, there is a very marked difference between the ways of functioning of these three drugs that is manifestly verified by the progression of modulated targets during development. The combined results of these three drugs thus delineate a very complementary and clear picture of transcripts regulated by the RA signaling pathway.

RAR antagonist treatment modulates initially only a low number of targets and a then a fast progression of modulated genes is observed in the following stages. These results can be explained by a generalized effect in potential directly regulated targets, which need some time to be clearly identified by the means of differential expression studies, those of which will further impact all the downstream effectors and cumulatively keep augmenting the number of differential expressed genes. Most of the targets modulated by RAR antagonist treatment are downregulated. Complementarily, RA has a very strong initial effect that progresses very fast to a high number of modulated genes already by the N3 stage, and differently the effect induced by inhibition of CYP26 represents a more linear cumulative effect. This can be explained by the global action that RA treatment produces, and thus inducing a consequent fold change of target genes that are both located in regions normally controlled by RA and in tissues where RA signaling is not usually actively controlling expression. Differently, the CYP26 inhibitor treatment induces an abnormal and progressive accumulation of RA in tissues that are typically depleted of this morphogen. Thus, depending on the sensitivity of the modulated genes or their upstream regulators, only after a long exposure to the drug an effective action is noticeable. These two treatments induce many genes to be upregulated and while the fold change induced by RA treatments rapidly reach high values, CYP26 inhibitor treatments induce fold changes that gradually increase through development. In summary, by implementing a set of complementary pharmacological treatments and by contrasting their global vs local effects in the amphioxus development we were able to recapitulate the intrinsic multilevel interactive feature of the RA signaling pathway, which coordinates very tightly multiple regulatory loops and other signaling pathways.

RA signaling controls chromatin rearrangements and drives expression by the intermediary of TF interactions in amphioxus

In the light of the observations made in vertebrates, which suggest that RAR/RXR heterodimers are partially bound and at the same time newly recruited to specific transcription sites[312,555,556], in amphioxus the set of identified genes modulated by the action of retinoid-specific treatments and their main associated molecular functions are consistent. In vertebrates, this action of RAR/RXR is supported by the simultaneous recruitment of coactivators and other TFs that facilitate, at the chromatin level, the action mediated by the RA signaling[567]. GO analysis indicates that great number of genes are intimately associated with typical housekeeping enzymatic reactions, such as retinoids metabolism. Although, our dataset also have an association with binding to regulatory regions, probably due to the presence of several TFs differentially regulated by the action of the RA signaling. Some of these identified TFs are known to interact with chromatin regions and induce conformational changes[561,568].

Moreover, by characterizing the open regions of the chromatin using ATACseq at the N3 stage, we can add to our analysis a new level of regulation and therefore bring to light potential epigenetic mechanisms that govern gene specific transcription mediated by RA signaling. Regions of the chromatin that were specifically induced into a closed state by RA treatment (i.e. URPs) show an enrichment for motifs that are associated with homeodomain containing factors, the Sox family of HMG, and for the orphan nuclear receptor NR2F6. In both vertebrates and amphioxus, homeodomain containing genes are mostly know to be activated by RA, for instance the well studied case of *Hox* genes[232,349,427,432]. In contrast, in vertebrates some reports also suggest that RA signaling can antagonize directly or indirectly the expression and activity of some members of this very large family, like for instance Pou2/Pou5f1, Hoxb-1 and Brn-3.2 in the developing CNS, Msx-1 in the limb bud and Evx1 in differentiating F9 cells[569–573]. This antagonizing interaction between the RA signaling and the homeodomain family of genes is responsible for a very specific set of actions during development[569–573] and thus it would be interesting to further explore these surprising result, both in terms of the specific binding sites and proteins involved, as well as the specific mechanism that governs these interactions. Along the same lines the Sox family (motifs binding also enriched in URPs) have been shown to interact with the homeodomain Pou subfamily, co-acting to regulate regions of active chromatin[574,575]. Moreover, Sox17 specific binding regions have been shown to have a slight negative correlation with the binding of the

RAR/RXR heterodimer in F9 cells[312]. This suggests that RA signaling is acting potentially in specific tissues or group of cells that are being patterned normally by the action of TFs from these two families. Similar situation have been described for the interaction of the orphan nuclear receptor NR2F6 and RAR/RXR couple, although this interaction has been only describe for the specific case of the gene *Renin*[576,577].

In contrast, the characterization of TF motifs in regions of chromatin that were induced by RA treatment into an open configuration (ORPs) lead to the identification of the bHLH E-box binding motifs, Fox associated binding sites as well as NRs binding motifs signature, mainly the REs for the RAR/RXR heterodimer. E-box sites are known to be bound by elements mediating cell-type-specific gene transcription, and for instance they have been identified in promoter and enhancer elements that regulate gene expression specifically in vertebrate muscles, neurons, and pancreas[578]. Fox genes represent a large family of transcription factors that arose early in evolution, and they are key components of many gene regulatory networks and essential for embryonic development[579,580]. Fox proteins can act as regular transcription factors, although they are also known for their pioneer activity, since they are able to interact with compacted chromatin to facilitate the binding of other transcriptional regulators[581]. Thus, the presence of these two binding sites families in open regions of the chromatin denotes that RA signaling induces downstream effects, by interacting with these two classes of TF.

The observation that ORPs host several NRs binding sites raises the questions of how RAR/RXR controls transcriptional variations and whether the presence of RAREs is by itself an indication of directly regulated expression. Thus, to address the potential interaction and regulation mediated directly by RAR/RXR heterodimer in a broader scale, we characterized the location and distribution of REs in the vicinity of all the modulated genes in the RNAseq. For now, though it is interesting to emphasize that DR5 elements, shown to be bound by the amphioxus RAR/RXR(Carvalho et al. in press) are preferentially located upstream of coding regions. Remarkably, though is the finding that some genes possess in their vicinity potential REs hotspots that could host more than one pair of nuclear receptors, suggesting that some genes are regulated either by an antagonizing balance between neighboring nuclear receptors or in the contrary by a coordinate action of multiple nuclear receptors. For instance, in vertebrates heterodimers of RXR and other NRs have been described as having specific DR binding sites: DR1 can be bound by PPAR, DR3s are typically associated with thyroid hormone receptor and DR4s are known binding sites for the Vitamin D receptor[517]. Although,

in amphioxus the specificity of biding sites for these NRs are largely unknown[582]. In our results, for instance, *Wnt5*, a described target of the RA signaling and seemingly affected in the anterior endoderm[427], possess a RE hotspot located upstream of the ORF composed by a DR4 and a DR5 element. Similarly, in close proximity to the gene *WIF*, also modulated by RA signaling, specifically downregulated in the endoderm (Appendix 3), we identified a composite DR1-DR1 arrangement. Thus, suggesting that the presence of hotspots of regulation indicates a tissue specific regulation by the combined action of RAR and other NRs. To address the specificity of binding sites that can host the amphioxus RAR/RXR heterodimer we are currently testing the capacity of the heterodimer to bind *in vitro* the consensus of the different elements identified: DR0, DR1, DR3 and DR4 since DR2 and DR5 were already tested and validated(Carvalho et al. in press).

In sum, by including multilevel prediction of binding sites associated with chromatin regions regulated by RA, and of REs elements located in the vicinity of differentially regulated transcripts we can point interesting features of the RA signaling pathway in amphioxus. First, RA is involved in chromatin rearrangements that specifically silence a set of TFs (i.e. Sox genes and homeodomain factors), which are possibly interacting into an uncharacterized process in amphioxus development. In contrast, RA signaling pathway activates a different specific set of TFs (NRs, Fox and E-box factors) that working in collaboration with NRs can regulate dynamically transcription of downstream genes. Furthermore, some of the genes being directly regulated by RA signaling can also be differentially modulated by the interaction of neighboring bound NRs. Further working addressing these interactions will allow the clarification of the mechanistic details involved in these regulatory loops involving RA signaling.

RAR/RXR and Fox factors interact to directly regulate RA-mediated transcription

Strikingly, we identified specific enrichment of RAR/RXR and Fox binding motifs in potential RA directed regulated chromatin open regions. Even more surprising was the mutually exclusive correlation that we verified for the relative distribution of these binding motifs within a specific set of peaks. This data clearly suggests a mechanistic interaction between RAR/RXR, and potentially even NRs, and Fox proteins in RA induced open chromatin. But which Fox factor mediates this mechanistic interaction and which other NRs, apart from RAR/RXR, might be involved?

Amphioxus has 32 *Fox* genes described and a great part of them can be identified during development[583,584]. In our RNAseq final dataset, all the *Fox* genes identified were downregulated by RA treatment, and *FoxAa* and *FoxA*b were upregulated by RA treatments, albeit not strongly enough to be retained in our final dataset of differently expressed genes(Appendix 3). Curiously, amphioxus *FoxAb* is a known direct target of the RA signaling pathway, although its paralog *FoxAa* seems not to be affected, at least when studying expression profiles in *B. floridae*[430,432]. In vertebrates, *FoxA* genes (*FoxA1 and FoxA2*) have been widely described as acting in organogenesis and to regulate target genes in a cell-type and stage-specific manner, mainly through either compensatory activity or by differently recruiting specific sets of activating or repressing regulatory co-factors in a highly context-specific manner[585–587]. The most striking feature of Fox proteins, though, is their capacity to recruit other TFs to a given regulatory region to exert a synergistic effect on the transcriptional activity of the targeted locus. Fox proteins are thus pioneer factors priming specific regulatory regions for transcription. For instance, FoxA1 preferentially binds chormatin regions marked with H3K4me2 and renders them transcriptionally active by recruiting other TFs, including a number of different NRs, such as glucocorticoid, androgen, and estrogen receptors as well as RAR/RXR heterodimers [585,588–590]. A recent study in breast cancer cells has thus demonstrated a fundamental role that FoxA1 play in recruiting RAR to specific target sites[567], suggesting a close relationship of RAR/RXR and FoxA factors in vertebrates, with FoxA factors acting as pioneers to pre-occupying regulatory regions primed for RA-specific activation by RAR/RXR heterodimers. One could imagine that, since RA treatment induces expression of FoxAa and FoxAb in amphioxus, these factors act as pioneers in binding to the regulatory DNA of RA target genes and recruit RAR/RXR to those specific domains to initiate transcription[567]. Additional work will be required to test this hypothesis[5] and to describe the mechanisms controlling the relationships between RAR and Fox proteins in amphioxus.

[5] We are currently still assessing the relationships between RAR/RXR and different Fox proteins during amphioxus development.

Conclusion

In the present study, we described for the first time in amphioxus a pharmacology-based study that combines the characterization of the transcriptome-dependent landscapes with changes in the chromatin permissiveness. Our results make a strong point on the global relationship between RA and its signaling targets during the development of amphioxus. Furthermore, we demonstrate that by relating the results obtained with different complementary pharmacological compounds acting locally as well as globally, a very robust set of genes can be obtained and that those results can easily be interpreted in a biological context. Moreover, the characterization of potential binding sites and their relative distribution can highlight undescribed mechanisms that drive RAR/RXR-dependent transcription. We uncovered a potential regulatory interaction of RAR/RXR and Fox factors during amphioxus development that has so far only been described in a very specialized cell system, in MCF-7 breast cancer cells, suggesting that Fox pioneering might be a previously unrecognized general mechanism of RA signaling acting during development that is evolutionarily conserved between amphioxus and vertebrates and has thus already been present in the last common ancestor of both lineages. This highlights not only the utility of amphioxus as a simple model to study the complexity of vertebrate gene regulatory networks, but also indicates the need for future work addressing the combined action of RAR and Fox factors in mediating specific development processes in both amphioxus and vertebrates systems.

Material and Methods

Amphioxus adult husbandry, embryo cultures and pharmacological treatments

Ripe European amphioxus (*Branchiostoma lanceolatum*) were collected by dredging in Argelès-sur-Mer, France, and retrieved from the sand by sieving. The collected animals, invariantly of their sex, were kept evenly into tanks with running natural sea water. The water temperature in the aquaria was kept at 16-17ºC under a spring-like day/night period (14 hours of light and 10 hours of absolute darkness). 36-hour thermal shock was used to induce spawning [148,196,213]. Following gametes collection and *in vitro* fertilization, the embryos were raised in artificial seawater, in the dark, at 19ºC [148]. Pharmacological treatments of *B. lanceolatum* embryos were performed at the G1 stage (6h of development) with all-*trans* RA (at 0,1 µM) (Sigma-Aldrich, Saint-Quentin Fallavier, France), RAR antagonist BMS493 (at 1 µM) (Sigma-Aldrich, Saint-Quentin Fallavier, France) and CYP26 inhibitor R115866 (at 0,5 µM) (provided by Janssen Research & Development, a division of Janssen Pharmaceutica NV, Beerse, Belgium). As a control, embryos were treated in separate dishes with dimethyl sulfoxide (DMSO) alone to a final dilution of 1:1000, to match the dilution used for each of the other compounds [418,426].

Embryos collection, sample preparation and high throughput sequencing

Embryos used to perform RNAseq analysis were raised in parallel for each of the 3 biological replicates (eggs from 3 different females were fertilized with the sperm originated from a single male). At the correct time point, N0, N3 and N4 stages (12, 18 and 24h of development, respectively), the embryos of each treatment and for each biological replicate (a total of 12 embryonic samples per time point) were collected and fixed in 4% paraformaldehyde as previously described [545] as well as simultaneously deep freeze in liquid nitrogen. Embryos were homogenized in tubes containing MagNA Lyser specific Green Beads (Roche Diagnostics, Meylan, France) and the total RNA was extracted using the RNeasy Mini Kit following the fabricant instructions (Qiagen SAS, Courtaboeuf, France). The overall quality of the final RNA obtained was verified using the Nanodrop (check company) and the Agilent 2100 Bioanalyzer (Agilent Technologies, Les Ulis, France). mRNA-focused sequencing libraries were prepared using Truseq RNA Sample Prep Kit v2 (Illumina, Eindhoven, The Netherlands). The libraries were sequenced on the Illumina Hiseq 2500 as paired-end 50 base reads following Illumina's instructions.

Embryos used to perform ATACseq were raised as mentioned above for each of the 2 biological replicates. At the N3 stage (18h of development) about 45 embryos of each treatment set and for each biological replicate (a total of 6 embryonic samples) were collected. The embryonic cells were dissociated by vividly pipetting the embryos up and down in lysis buffer (10mM Tris-HCl pH 7,4, 10mM NaCl, 3mM $MgCl_2$ and 0,1% NP40), yielding an average of 133000 cells per sample. Tagmentation was conducted using the Nextera DNA Library Preparation Kit (Illumina, Eindhoven, The Netherlands) accordingly to the manufacture instructions, followed by purification with MinElute Kit (Qiagen SAS, Courtaboeuf, France). DNA amplification was performed as previously described [591,592]. Purified amplification products were sequenced the Illumina Hiseq 2500 as single-end 50 base reads following Illumina's instructions.

Data processing and analyses

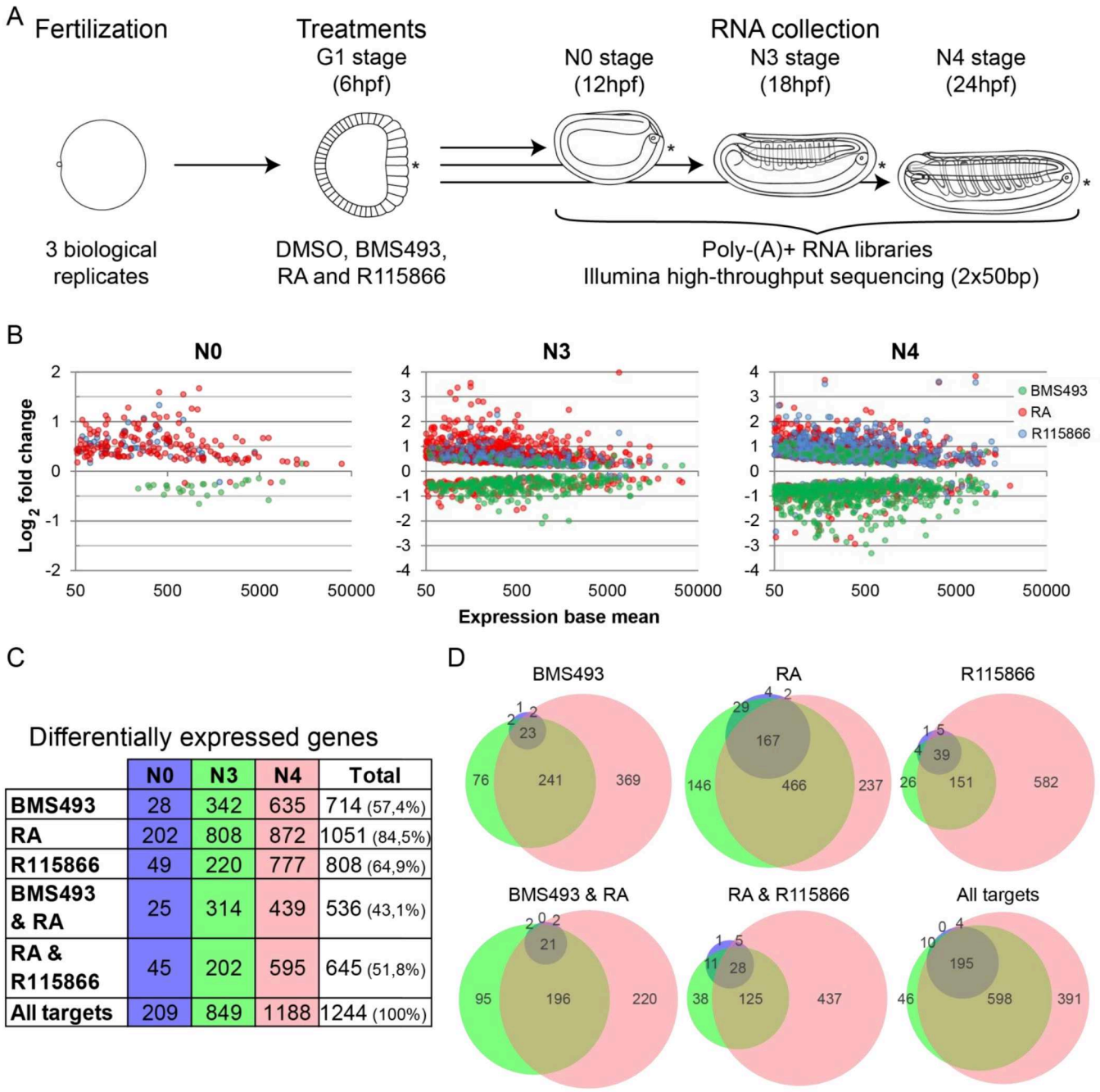
RNAseq reads were mapped with STAR[593] using intron junctions set inferred from gene model as an assistance and default parameters. Read counts for each gene model were obtained with featureCounts[594] using the "EVM/PASA+cufflinks" gene models for *B. lanceolatum*. Differential analysis of count data were performed with DESeq2 implemented in R[595]. A gene was declared modulated if it displayed a significant difference between any treatment condition and control in at least a time point (the cut-off was fixed at $5.10^{-2}$of (Bonferroni-Hochberg) adjusted P-value), a estimated fold change of +/- 50% (log2 fold change inferior to -0,585 and superior to 0,585), and at least in one given point a estimated level over 50 (baseMean superior to 50). Hierarchical clustering based in Ward's test was use to define gene clusters. GO analysis were performed using TopGO R package[596] and network graphs were created in Cytoscape v3.4.0[597].

For each ATACseq samples reads were mapped to the *B. lanceolatum* v1 genome 'Bl71nemr' using Bowtie2[598]. Bam files were processed to select 'nucleosome free' read pairs (insert <150bp) and peaks were called using MACS2[599] using the parameters: --nomodel --extsize 100 --shift -50. Peaks without overlap with gene model TSS were considered as putative distal regulatory elements. All peaks were intersected using the mergePeaks script from the HOMER toolkit to define a non-redundant peak set[600]. Tag density was estimated using BEDTools[601]. Differential activation analysis was then conduced using DESeq2[595] and only peaks exhibiting adjusted p-value >0,1 were considered.

Retinoic acid response elements and enriched motifs discovery

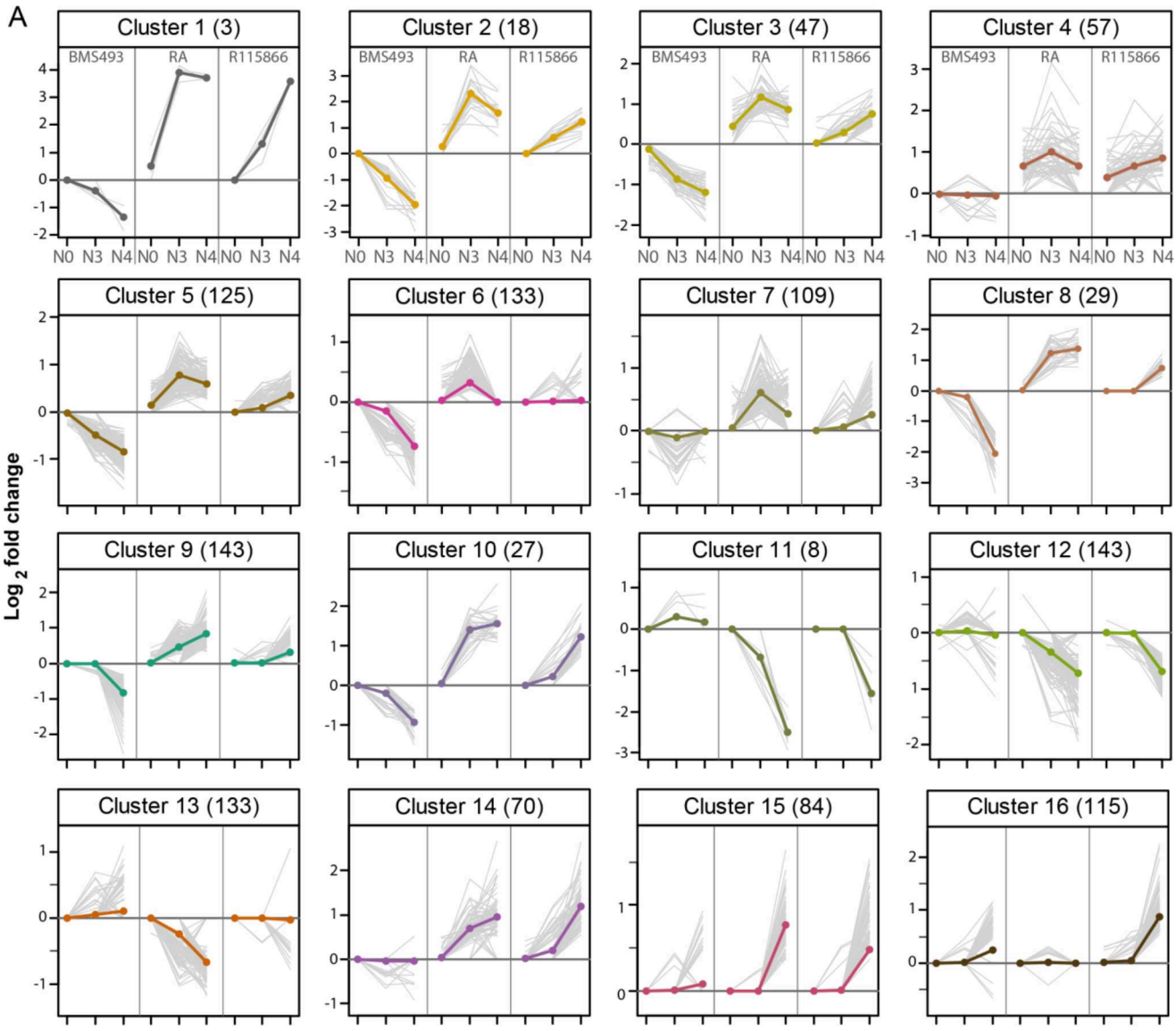
Putative RARE and related motif sequences were identified using an automated pipeline [429]Carvalho et al. in press) using as input sequences RAREs previously described as functional in amphioxus as well as all possible RARE combinations resulting from the canonical vertebrate RARE consensus: (A/G)G(G/T)TCA$(N)_{0-5}$(A/G)G(G/T)TCA. Furthermore, ATACseq peaks were scanned for de novo enriched motifs by the combination of two different method: the online version of MEME Suite [543], specifically MEME-chip protocol and with the online tool RSAT peak-motifs[602,603]. Enriched motifs were considered when E-value >0,05.

Figures



	N0	N3	N4	Total
BMS493	28	342	635	714 (57,4%)
RA	202	808	872	1051 (84,5%)
R115866	49	220	777	808 (64,9%)
BMS493 & RA	25	314	439	536 (43,1%)
RA & R115866	45	202	595	645 (51,8%)
All targets	209	849	1188	1244 (100%)

Fig. R3.1 – Global results of the retinoid treatment dependent RNAseq. *(A) Experimental setup of the RNAseq experiment. * indicates the posterior side (B) MA plots for the differential expressed genes in each time point: N0, N3 and N4. (C) Number of differential expressed genes by time point and treatment (BMS493, RA, R115866), combination of treatments (BMS493 and RA, RA and R115866) or in all targets. (D) Veen diagrams showing the distribution of differentially expressed genes per treatment. Circle colors correspond to the column colors in C, for N0 (blue), N3 (green) and N4 (pink).*


A
Cluster 1 (3)
Cluster 2 (18)
Cluster 3 (47)
Cluster 4 (57)
Cluster 5 (125)
Cluster 6 (133)
Cluster 7 (109)
Cluster 8 (29)
Cluster 9 (143)
Cluster 10 (27)
Cluster 11 (8)
Cluster 12 (143)
Cluster 13 (133)
Cluster 14 (70)
Cluster 15 (84)
Cluster 16 (115)
BMS493
RA
R115866
N0 N3 N4
Log 2 fold change

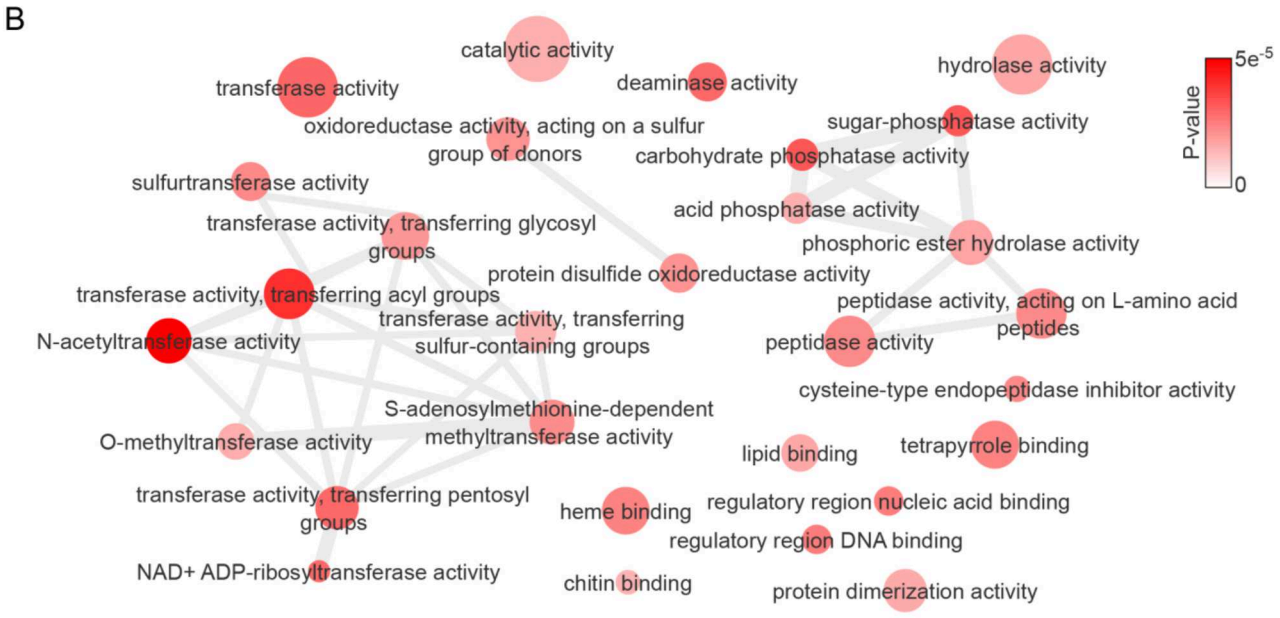

B
catalytic activity
transferase activity
deaminase activity
hydrolase activity
oxidoreductase activity, acting on a sulfur group of donors
sugar-phosphatase activity
carbohydrate phosphatase activity
sulfurtransferase activity
acid phosphatase activity
transferase activity, transferring glycosyl groups
phosphoric ester hydrolase activity
protein disulfide oxidoreductase activity
transferase activity, transferring acyl groups
peptidase activity, acting on L-amino acid peptides
transferase activity, transferring sulfur-containing groups
N-acetyltransferase activity
peptidase activity
cysteine-type endopeptidase inhibitor activity
S-adenosylmethionine-dependent methyltransferase activity
O-methyltransferase activity
lipid binding
tetrapyrrole binding
transferase activity, transferring pentosyl groups
heme binding
regulatory region nucleic acid binding
regulatory region DNA binding
NAD+ ADP-ribosyltransferase activity
chitin binding
protein dimerization activity
P-value
5e-5
0

Fig. R3.2 – Clusters of differential expressed genes and gene ontology (GO).analysis. *(A) Clusters grouping differential expressed genes based on their expression profiles. Number of genes per cluster are indicated. For a question of simplification, only the first 4 clusters have indicated the treatments (from left to right in each cluster graph: BMS493, RA and R115866) and the time points (for each of the treatments: N0, N3 and N4) that are systematically the same for every cluster representation. Each grey line represents the log2 fold change of a single gene and colorful lines correspond to the mean representation of the log2 fold change for the entire cluster. (B) Molecular function GO results. Circle colors represent the P-value associated with the presence of the respective GO term, the circle size represents the relative number of hits in the GO database and lines connect closely related terms.*

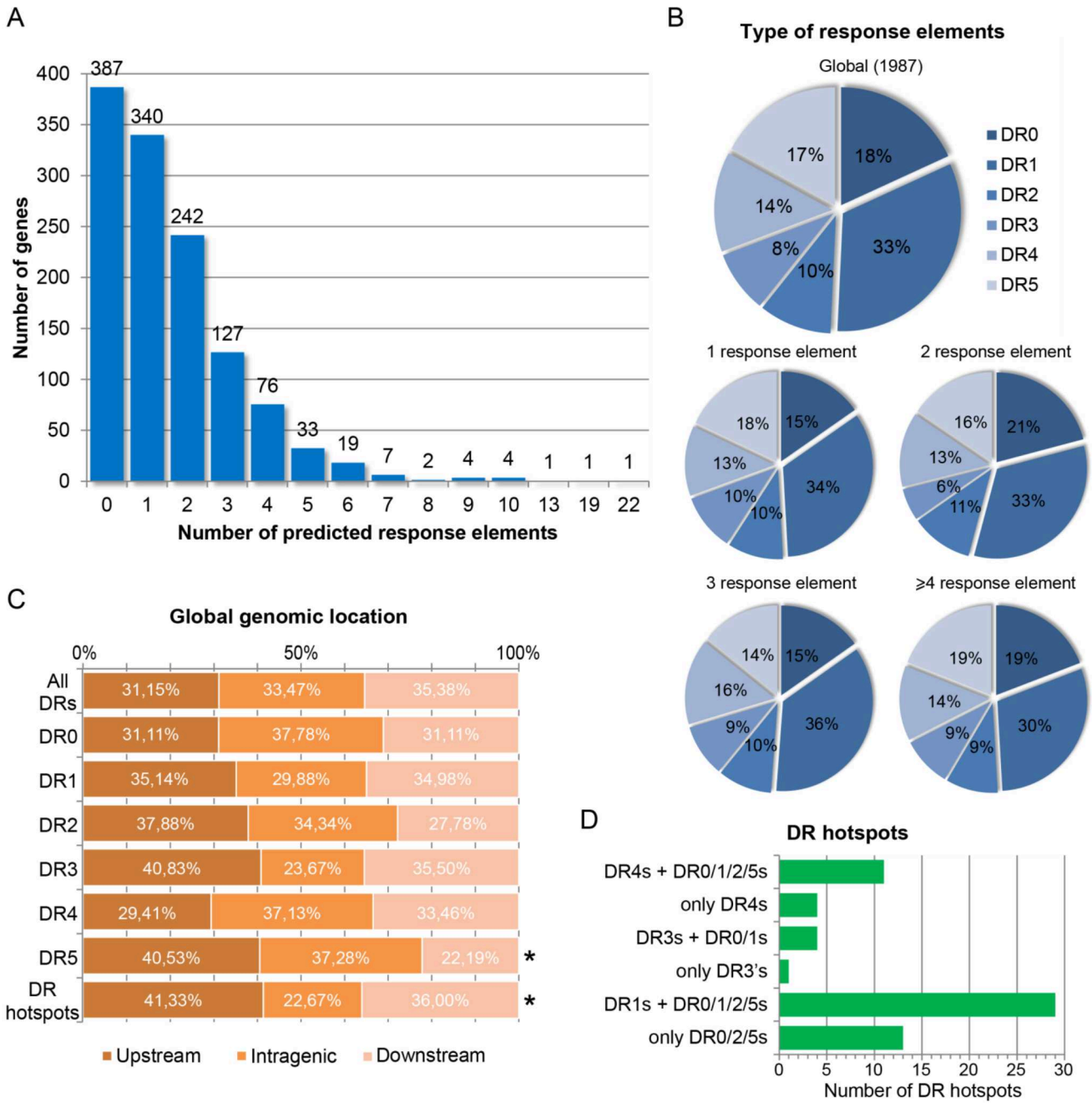


Fig. R3.3 – Putative response elements (REs) occurrence and distribution in the RNAseq modulated targets. *(A) Number of occurrences of REs per transcript. (B) Frequency of the REs identified in global, considering transcripts that have 1 response element, 2 response elements, 3 response elements and 4 or more response elements. (C) Relative genomic location of the identified REs. (D) Number of the different REs hotspots identified.*

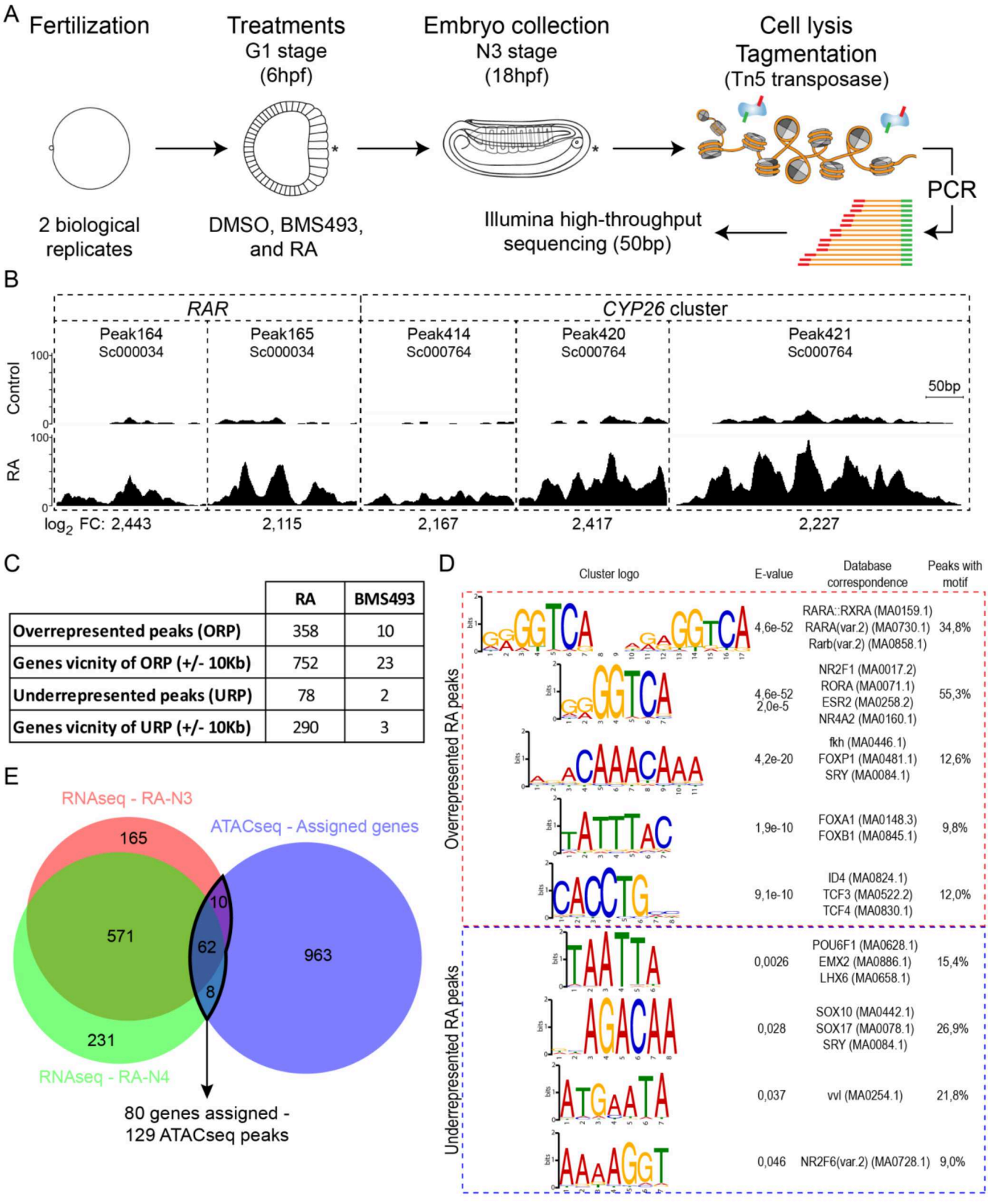


Fig. R3.4 – ATACseq results. *(A) Experimental setup of the ATACseq experiment. * indicates the posterior side. (B) Example of the 5 top overrepresented peaks (ORPs) induced by RA treatment. (C) Overall number for overrepresented and underrepresented peaks (URPs) and the respective*

number of assigned genes for each treatment condition: RA and BMS493. (D) Enriched motifs in ORPs and URPs and the respective top hits in the JASPAR database. (E) Veen diagram showing the intersection between all the ATACseq assigned genes and the RA modulated RNAseq identified targets at the N3 and N4 stage.

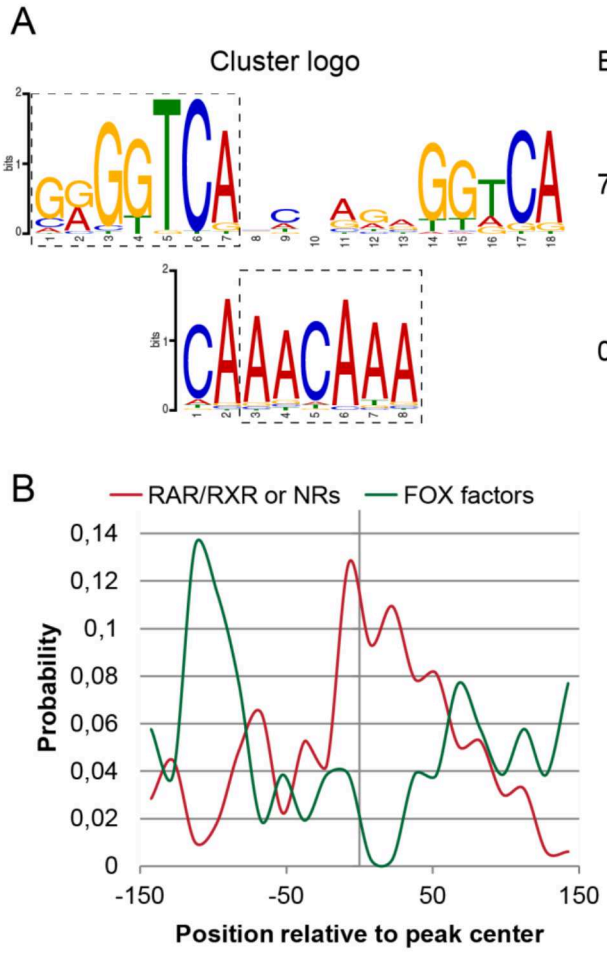


Fig. R3.5 – Enriched motifs and distribution in potentially directly regulated open regions of the chromatin. *(A) Enriched motifs with respective top hits in the JASPAR database, as well as the number of potential peaks with the match and corresponding conserved assigned genes. Squared box represents the more conservative identified domains (B) Probabilistic distribution of the identified motifs within the 300 bp peaks.*

Conclusion and perspectives

The work presented in this manuscript, addressed multiple questions concerning both the development of amphioxus and the evolution of the RA signaling pathway in deuterostomes, focusing on chordate diversification. The multidisciplinary and integrative set of approaches used, ranging from the ecology and technical developments applied to the amphioxus aquaculture system, the molecular tools developed, such as the microinjection of amphioxus eggs, the evo-devo strategy combining expression profiles with pharmacological treatments as well as the implementation of high throughput sequencing methods to study RNA profiles, genomic and epigenetic landscapes, allowed me to shed new light on several aspects that once again revealed the importance of studying "non-model" organisms.

I thus described the evolutionary diversification of CYP26 genes in deuterostomes and I used amphioxus *CYP26* expression, function and mechanisms that govern the feedback loop controlled directly by RA to establish a proxy for potential roles assumed by RA degradation in chordates. Therefore, I defined in more detail how RA distribution is set during amphioxus development and how this tissue-specific depletion of RA is fundamental to pattern certain organs, using as example the amphioxus larval tail fin. Furthermore, I revealed a new set of genes that are under the control of the RA signaling pathway as well as new potential regulatory loops driving RAR and eventually other NR-mediated expression. Moreover, I established a framework that allows the characterization of gene hierarchies during development that can be widely applied to other signaling pathways and organisms.

Furthermore, following up exactly on the work described in here, our lab is currently preparing a complete description of the influence that RA degradation promoted by CYP26 has during development, focusing on specific tissues. This study will characterize the influence of CYP26 inhibition on specific key genes, and thus relate with the developmental patterning of certain tissues, like the general ectoderm and potential ectodermal sensory neurons, CNS, anterior endoderm and mesoderm.

Concerning this last point, additional verifications are still to be made. In the near future, it will be extremely important to correlate the distribution of *RAR*, *Fox* genes modulated by RA treatments as well as of other potential interesting NR (e.g. estrogen receptor), and verify the dynamics of these potential interactions with the identified likely direct targets of RA (e.g. *Pbx*, *RDH10*, *FABP7*, *ZNF503*). With these insights, one should be able to produce a very strong statement about the evolution and diversification of the RA

signaling pathway as well as about the evolutionary conservation of regulatory mechanisms driven by this signaling cascade.

To expand further the impact of the research conducted on the regulation of RA signaling targets during development, a next possible step would address the chromatin state at different other stages of development and the results obtained under the same experimental conditions would certainly either confirm the interaction of NR and Fox factors as a global mechanism or classify it as a fundamental mark of the mid-neurula RA-dependent specification. To complement such a study and to reinforce the potential role that RAR plays in this interaction, the establishment of a RAR-directed ChIPseq (Chromatin Immunoprecipitation followed by high throughput sequencing) during normal development and after disruption of endogenous RA signaling by either pharmacological treatments or microinjection-based approaches would be of exquisite importance. Not only that, but in the context of the described epigenetic signals allowing specific binding of Fox factors, it would be interesting to include also histone modification ChIPseq, hence providing at the same time the epigenetic landmarks that specifically recruit Fox factors, RAR and eventually both in parallel. Furthermore, and to place the amphioxus system as a potential new tool to characterize the ancestral epigenetic RA-driven mechanism, one would use the results obtained from amphioxus and the same set of analyses coming from vertebrate embryonic systems (e.g. *Xenopus*, mouse, etc.) and tissue-specific cell lines (e.g. F9 embryonal carcinoma, MCF-7 breast cancer, etc.). With the addition of comparative studies on vertebrate models, another layer of complexity can then easily be added: (1) the tissue specificity of the RA signaling pathway, since dissection of tissues can be easily achieved for downstream applications; and (2) the independent RAR paralog (RARα, RARβ and RARγ) functions, which can be targeted by paralog-specific drugs. Thus, comparisons between tissue-specific and paralog-specific data with global data obtained from amphioxus studies would highlight the potential roles of the RA signaling pathway in the basal chordate and which RA functions were acquired, maintained or lost during the independent evolutionary story of these two phyla, the cephalochordates and the vertebrates.

In sum, the research project described in this manuscript uses a large set of complementary approaches to shed light on the evolution of the RA signaling pathway in chordates and raises a very large spectrum of questions about the ancestral mode of action as well as about the conservation of the gene hierarchy controlled by the RA signaling pathway.

Bibliography

1. Buffon, G.-L. L. (1707-1788 ; comte de) A. du texte & Daubenton, L.-J.-M. (1716-1800) A. du texte. *Histoire naturelle générale et particulière : avec la description du Cabinet du Roy*. (Imprimerie Royale, 1749).
2. Lamarck, J.-B. de M. de (1744-1829). *Philosophie zoologique*. (Dentu, 1809).
3. Saint-Hilaire, É. G. *Principes de philosophie zoologique*. (Pichon et Didier, 1830).
4. Darwin, C. *On the origin of species by means of natural selection*. (J. Murray, 1859).
5. Futuyma, D. J. *Evolution*. (Sinauer Associates Inc., 2009).
6. Haeckel, E. H. P. A. *The evolution of man*. (D. Appleton and company, 1897).
7. Sansom, R. S., Gabbott, S. E. & Purnell, M. A. Non-random decay of chordate characters causes bias in fossil interpretation. *Nature* **463,** 797–800 (2010).
8. Sansom, R. S. Preservation and phylogeny of Cambrian ecdysozoans tested by experimental decay of *Priapulus*. *Sci. Rep.* **6,** 32817 (2016).
9. Janvier, P. Facts and fancies about early fossil chordates and vertebrates. *Nature* **520,** 483–489 (2015).
10. Losos, J. B. *et al.* Evolutionary biology for the 21st century. *PLOS Biol* **11,** e1001466 (2013).
11. Gilbert, S. F. & Singer, S. *Developmental biology*. (Sinauer Associates Inc., 2006).
12. Newman, S. A. Physico-genetic determinants in the evolution of development. *Science* **338,** 217–219 (2012).
13. Baer, K. E. von. *Über entwickelungsgeschichte der thiere. Beobachtung und reflexion*. (Gebrüdern Bornträger, 1828).
14. Haeckel, E. H. P. A. *Generelle morphologie der organismen*. (Georg Reimer, 1986).
15. Gould, S. J. *Ontogeny and phylogeny*. (Harvard University Press, 1977).
16. Alberch, P., Gould, S. J., Oster, G. F. & Wake, D. B. Size and shape in ontogeny and phylogeny. *Paleobiology* **5,** 296–317 (1979).
17. Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276,** 565–570 (1978).
18. de Rosa, R. *et al.* Hox genes in brachiopods and priapulids and protostome evolution. *Nature* **399,** 772–776 (1999).
19. Diogo, R. Where is the Evo in Evo-Devo (evolutionary developmental biology)? *J. Exp. Zoolog. B Mol. Dev. Evol.* **326,** 9–18 (2016).
20. Hall, B. K. Guest Editorial: Evo-devo or devo-evo — does it matter? *Evol. Dev.* **2,** 177–178 (2000).
21. Takechi, M. & Kuratani, S. History of studies on mammalian middle ear evolution: a comparative morphological and developmental biology perspective. *J. Exp. Zoolog. B Mol. Dev. Evol.* **314B,** 417–433 (2010).
22. Cracraft, J. Phylogeny and evo-devo: characters, homology, and the historical analysis of the evolution of development. *Zoology* **108,** 345–356 (2005).
23. Wagner, G. P. The biological homology concept. *Annu. Rev. Ecol. Syst.* **20,** 51–69 (1989).
24. Kleisner, K. The formation of the theory of homology in biological sciences. *Acta Biotheor.* **55,** 317–340 (2007).
25. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430,** 85–88 (2004).
26. Takahasi, K. R., Matsuo, T. & Takano-Shimizu-Kouno, T. Two types of *cis-trans* compensation in the evolution of transcriptional regulation. *Proc. Natl. Acad. Sci.* **108,** 15276–15281 (2011).
27. Gonçalves, A. *et al.* Extensive compensatory *cis-trans* regulation in the evolution of mouse gene expression. *Genome Res.* **22,** 2376–2384 (2012).
28. Landry, C. R. *et al.* Compensatory *cis-trans* evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics* **171,** 1813–1822 (2005).
29. Ohno, S. *Evolution by gene duplication*. (Springer-Verlag, 1970).
30. Sidow, A. Gen(om)e duplications in the evolution of early vertebrates. *Curr. Opin. Genet. Dev.* **6,** 715–722 (1996).
31. Putnam, N. H. *et al.* The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453,** 1064–1071 (2008).
32. Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10,** 725–732 (2009).
33. Kuraku, S. Insights into cyclostome phylogenomics: pre-2R or post-2R. *Zoolog. Sci.* **25,** 960–968 (2008).
34. Kuraku, S., Meyer, A. & Kuratani, S. Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Mol. Biol. Evol.* **26,** 47–59 (2009).
35. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151,** 1531–1545 (1999).

36. MacCarthy, T. & Bergman, A. The limits of subfunctionalization. *BMC Evol. Biol.* **7,** 213 (2007).
37. Smith, J. J. & Keinath, M. C. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. *Genome Res.* **25,** 1081–1090 (2015).
38. Asrar, Z., Haq, F. & Abbasi, A. A. Fourfold paralogy regions on human HOX-bearing chromosomes: role of ancient segmental duplications in the evolution of vertebrate genome. *Mol. Phylogenet. Evol.* **66,** 737–747 (2013).
39. Abbasi, A. A. Diversification of four human HOX gene clusters by step-wise evolution rather than ancient whole-genome duplications. *Dev. Genes Evol.* **225,** 353–357 (2015).
40. Hafeez, M., Shabbir, M., Altaf, F. & Abbasi, A. A. Phylogenomic analysis reveals ancient segmental duplications in the human genome. *Mol. Phylogenet. Evol.* **94, Part A,** 95–100 (2016).
41. Dennis, M. Y. & Eichler, E. E. Human adaptation and evolution by segmental duplication. *Curr. Opin. Genet. Dev.* **41,** 44–52 (2016).
42. Morgan, A. P. *et al.* The evolutionary fates of a large segmental duplication in mouse. *Genetics* **204,** 267–285 (2016).
43. Akam, M. Hox and HOM: homologous gene clusters in insects and vertebrates. *Cell* **57,** 347–349 (1989).
44. Pascual-Anaya, J., D'Aniello, S., Kuratani, S. & Garcia-Fernàndez, J. Evolution of *Hox* gene clusters in deuterostomes. *BMC Dev. Biol.* **13,** 26 (2013).
45. Holland, P. W. H. Evolution of homeobox genes. *Wiley Interdiscip. Rev. Dev. Biol.* **2,** 31–45 (2013).
46. Gaunt, S. J. The significance of *Hox* gene collinearity. *Int. J. Dev. Biol.* **59,** 159–170 (2015).
47. Alexander, T., Nolte, C. & Krumlauf, R. *Hox* genes and segmentation of the hindbrain and axial skeleton. *Annu. Rev. Cell Dev. Biol.* **25,** 431–456 (2009).
48. Foronda, D., de Navas, L. F., Garaulet, D. L. & Sanchez-Herrero, E. Function and specificity of *Hox* genes. *Int. J. Dev. Biol.* **53,** 1404–1419 (2009).
49. Postlethwait, J. H. & Schneiderman, H. A. A clonal analysis of determination in *Antennapedia* a homoeotic mutant of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* **64,** 176–183 (1969).
50. Emerald, B. S. & Cohen, S. M. Spatial and temporal regulation of the homeotic selector gene *Antennapedia* is required for the establishment of leg identity in *Drosophila*. *Dev. Biol.* **267,** 462–472 (2004).
51. Telford, M. J., Budd, G. E. & Philippe, H. Phylogenomic insights into animal evolution. *Curr. Biol.* **25,** R876–R887 (2015).
52. Dunn, C. W., Giribet, G., Edgecombe, G. D. & Hejnol, A. Animal phylogeny and its evolutionary implications. *Annu. Rev. Ecol. Evol. Syst.* **45,** 371–395 (2014).
53. Lanna, E. Evo-devo of non-bilaterian animals. *Genet. Mol. Biol.* **38,** 284–300 (2015).
54. Dohrmann, M. & Wörheide, G. Novel scenarios of early animal evolution — Is it time to rewrite textbooks? *Integr. Comp. Biol.* **53,** 503–511 (2013).
55. Philippe, H. *et al.* Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19,** 706–712 (2009).
56. Leys, S. P. & Riesgo, A. Epithelia, an evolutionary novelty of Metazoans. *J. Exp. Zoolog. B Mol. Dev. Evol.* **318,** 438–447 (2012).
57. Ryan, J. F. *et al.* The genome of the Ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* **342,** 1242592 (2013).
58. Moroz, L. L. *et al.* The ctenophore genome and the evolutionary origins of neural systems. *Nature* **510,** 109–114 (2014).
59. Hejnol, A. & Martindale, M. Q. in *Animal evolution: genomes, fossils, and trees* (eds. Telford, M. J. & Littlewood, D. T. . J.) (Oxford University Press, 2009).
60. Martín-Durán, J. M., Janssen, R., Wennberg, S., Budd, G. E. & Hejnol, A. Deuterostomic development in the protostome *Priapulus caudatus*. *Curr. Biol.* **22,** 2161–2166 (2012).
61. Janssen, R., Jörgensen, M., Lagebro, L. & Budd, G. E. Fate and nature of the onychophoran mouth–anus furrow and its contribution to the blastopore. *Proc. R. Soc. Lond. B Biol. Sci.* **282,** 20142628 (2015).
62. Rouse, G. W., Wilson, N. G., Carvajal, J. I. & Vrijenhoek, R. C. New deep-sea species of *Xenoturbella* and the position of Xenacoelomorpha. *Nature* **530,** 94–97 (2016).
63. Hejnol, A. *et al.* Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc. R. Soc. Lond. B Biol. Sci.* rspb20090896 (2009). doi:10.1098/rspb.2009.0896
64. Hejnol, A. & Pang, K. Xenacoelomorpha's significance for understanding bilaterian evolution. *Curr. Opin. Genet. Dev.* **39,** 48–54 (2016).
65. Cannon, J. T. *et al.* Xenacoelomorpha is the sister group to Nephrozoa. *Nature* **530,** 89–93 (2016).
66. Philippe, H. *et al.* Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. *Nature* **470,** 255–258 (2011).
67. Giribet, G. Genomics and the animal tree of life: conflicts and future prospects. *Zool. Scr.* **45,** 14–21 (2016).

68. Harzsch, S., Müller, C. H. G. & Perez, Y. in *Evolutionary developmental biology of invertebrates* (ed. Wanninger, A.) **Vol. 1,** 215–240 (Springer Vienna, 2015).
69. Perez, Y., Müller, C. H. & Harzsch, S. in *Deep Metazoan Phylogeny: The Backbone of the Tree of Life* (eds. Wägele, J. W. & Bartolomaeus, T.) 49–77 (2014).
70. Chapman, A. D. *Numbers of living species in Australia and the world*. (Australian Biological Resources Study, 2009).
71. *Evolutionary developmental biology of invertebrates*. **Vol. 2,** (Springer Vienna, 2015).
72. Swalla, B. J. & Smith, A. B. Deciphering deuterostome phylogeny: molecular, morphological and palaeontological perspectives. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363,** 1557–1568 (2008).
73. Cannon, J. T., Rychel, A. L., Eccleston, H., Halanych, K. M. & Swalla, B. J. Molecular phylogeny of hemichordata, with updated status of deep-sea enteropneusts. *Mol. Phylogenet. Evol.* **52,** 17–24 (2009).
74. Röttinger, E. & Lowe, C. J. Evolutionary crossroads in developmental biology: hemichordates. *Development* **139,** 2463–2475 (2012).
75. Brown, F. D., Prendergast, A. & Swalla, B. J. Man is but a worm: Chordate origins. *genesis* **46,** 605–613 (2008).
76. Lowe, C. J. Molecular genetic insights into deuterostome evolution from the direct-developing hemichordate *Saccoglossus kowalevskii*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363,** 1569–1578 (2008).
77. McClay, D. R. Evolutionary crossroads in developmental biology: sea urchins. *Development* **138,** 2639–2648 (2011).
78. Arnone, M. I., Byrne, M. & Martinez, P. in *Evolutionary Developmental Biology of Invertebrates* (ed. Wanninger, A.) **Vol. 6,** 1–58 (Springer Vienna, 2015).
79. Telford, M. J. *et al.* Phylogenomic analysis of echinoderm class relationships supports Asterozoa. *Proc. R. Soc. Lond. B Biol. Sci.* **281,** 20140479 (2014).
80. Bourlat, S. J., Nielsen, C., Lockyer, A. E., Littlewood, D. T. J. & Telford, M. J. *Xenoturbella* is a deuterostome that eats molluscs. *Nature* **424,** 925–928 (2003).
81. Bourlat, S. J. *et al.* Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444,** 85–88 (2006).
82. Stemple, D. L. Structure and function of the notochord: an essential organ for chordate development. *Development* **132,** 2503–2512 (2005).
83. Annona, G., Holland, N. D. & D'Aniello, S. Evolution of the notochord. *EvoDevo* **6,** (2015).
84. Paris, M. *et al.* Amphioxus postembryonic development reveals the homology of Chordate metamorphosis. *Curr. Biol.* **18,** 825–830 (2008).
85. Satoh, N., Rokhsar, D. & Nishikawa, T. Chordate evolution and the three-phylum system. *Proc R Soc B* **281,** 20141729 (2014).
86. Haeckel, E. *Anthropogenie oder entwickelungsgeschichte des menschen*. (W. Engelmann, 1910).
87. Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439,** 965–968 (2006).
88. Koop, D. & Holland, L. Z. The basal chordate amphioxus as a simple model for elucidating developmental mechanisms in vertebrates. *Birth Defects Res. Part C Embryo Today Rev.* **84,** 175–187 (2008).
89. Ota, K. G., Fujimoto, S., Oisi, Y. & Kuratani, S. Identification of vertebra-like elements and their possible differentiation from sclerotomes in the hagfish. *Nat. Commun.* **2,** 373 (2011).
90. Ota, K. G., Oisi, Y., Fujimoto, S. & Kuratani, S. The origin of developmental mechanisms underlying vertebral elements: implications from hagfish evo-devo. *Zoology* **117,** 77–80 (2014).
91. Shimeld, S. M. & Donoghue, P. C. J. Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). *Development* **139,** 2091–2099 (2012).
92. Stolfi, A. & Brown, F. D. in *Evolutionary Developmental Biology of Invertebrates* (ed. Wanninger, A.) **Vol. 6,** 135–204 (Springer Vienna, 2015).
93. Lemaire, P. Evolutionary crossroads in developmental biology: the tunicates. *Development* **138,** 2143–2152 (2011).
94. Tsagkogeorga, G. *et al.* An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. *BMC Evol. Biol.* **9,** 187 (2009).
95. Berná, L. & Alvarez-Valin, F. Evolutionary genomics of fast evolving tunicates. *Genome Biol. Evol.* **6,** 1724–1738 (2014).
96. Holland, L. Z. in *Evolutionary developmental biology of invertebrates* (ed. Wanninger, A.) **Vol. 6,** 91–133 (Springer, 2015).
97. Bertrand, S. & Escriva, H. Evolutionary crossroads in developmental biology: amphioxus. *Development* **138,** 4819–4830 (2011).
98. Holland, L. Z., Laudet, V. & Schubert, M. The chordate amphioxus: an emerging model organism for developmental biology. *Cell. Mol. Life Sci. CMLS* **61,** 2290–2308 (2004).

99. Schubert, M., Escriva, H., Xavier-Neto, J. & Laudet, V. Amphioxus and tunicates as evolutionary model systems. *Trends Ecol. Evol.* **21,** 269–277 (2006).
100. Louis, A., Crollius, H. R. & Robinson-Rechavi, M. How much does the amphioxus genome represent the ancestor of chordates? *Brief. Funct. Genomics* **11,** 89–95 (2012).
101. Janvier, P. Vertebrate characters and the Cambrian vertebrates. *Comptes Rendus Palevol* **2,** 523–531 (2003).
102. Holland, N. D. & Chen, J. Origin and early evolution of the vertebrates: new insights from advances in molecular biology, anatomy, and palaeontology. *BioEssays* **23,** 142–151 (2001).
103. Mallatt, J. & Chen, J. Fossil sister group of craniates: predicted and found. *J. Morphol.* **258,** 1–31 (2003).
104. Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. & Worm, B. How many species are there on Earth and in the ocean? *PLOS Biol* **9,** e1001127 (2011).
105. Sommer, R. J. The future of evo–devo: model systems and evolutionary theory. *Nat. Rev. Genet.* **10,** 416–422 (2009).
106. Jarvela, A. M. C. & Pick, L. in *Current Topics in Developmental Biology* (ed. Wassarman, P. M.) **117,** 253–274 (Academic Press, 2016).
107. Jenner, R. A. & Wills, M. A. The choice of model organisms in evo–devo. *Nat. Rev. Genet.* **8,** 311–314 (2007).
108. Holland, L. Z. *et al.* Evolution of bilaterian central nervous systems: a single origin? *EvoDevo* **4,** 27 (2013).
109. Ellegren, H. Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* **29,** 51–63 (2014).
110. Pantalacci, S. & Sémon, M. Transcriptomics of developing embryos and organs: a raising tool for evo–devo. *J. Exp. Zoolog. B Mol. Dev. Evol.* **324,** 363–371 (2015).
111. Reis, A., Hornblower, B., Robb, B. & Tzertzinis, G. CRISPR/Cas9 and targeted genome editing: a new era in molecular biology. *NEB Expr.* **1,** 3–6 (2014).
112. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346,** 1258096 (2014).
113. Moczek, A. P. *et al.* The significance and scope of evolutionary developmental biology: a vision for the 21st century. *Evol. Dev.* **17,** 198–219 (2015).
114. Willey, A. *Amphioxus and the ancestry of the vertebrates*. (Macmillan and Co., 1894).
115. Gill, T. The lowest of the vertebrates and their origin. *Science* **1,** 645–649 (1895).
116. Herdman, W. A. in *The Cambridge natural history* **Vol. 7,** (Macmillan and Co., 1904).
117. Shimeld, S. M. & Holland, P. W. H. Vertebrate innovations. *Proc. Natl. Acad. Sci.* **97,** 4449–4452 (2000).
118. Schlosser, G. Vertebrate cranial placodes as evolutionary innovations - the ancestor's tale. *Curr. Top. Dev. Biol.* **111,** 235–300 (2015).
119. Holland, L. Z. & Onai, T. Early development of cephalochordates (amphioxus). *Wiley Interdiscip. Rev. Dev. Biol.* **1,** 167–183 (2012).
120. Holland, L. Z. Heads or tails? Amphioxus and the evolution of anterior-posterior patterning in deuterostomes. *Dev. Biol.* **241,** 209–228 (2002).
121. Hatschek, B. (1854-1941). Leuckart Chart, Series I, Chart 72: Vertebrata; Pisces; Acranii; *Amphioxus lanceolatus*. *Mar. Biol. Lab. Arch. Leuckart Chart 114* (1892).
122. Yu, J. K. S. & Holland, L. Z. Cephalochordates (amphioxus or lancelets): a model for understanding the evolution of chordate characters. *Cold Spring Harb. Protoc.* **2009,** pdb.emo130 (2009).
123. Holland, L. Z. *et al.* The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res.* **18,** 1100–1111 (2008).
124. Pascual-Anaya, J., D'Aniello, S. & Garcia-Fernàndez, J. Unexpectedly large number of conserved noncoding regions within the ancestral chordate Hox cluster. *Dev. Genes Evol.* **218,** 591–597 (2008).
125. Huang, S. *et al.* Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res.* **18,** 1112–1126 (2008).
126. D'Aniello, S. *et al.* Gene expansion and retention leads to a diverse tyrosine kinase superfamily in amphioxus. *Mol. Biol. Evol.* **25,** 1841–1854 (2008).
127. Albalat, R., Brunet, F., Laudet, V. & Schubert, M. Evolution of retinoid and steroid signaling: vertebrate diversification from an amphioxus perspective. *Genome Biol. Evol.* **3,** 985–1005 (2011).
128. Huang, S. *et al.* Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. *Nat. Commun.* **5,** 5896 (2014).
129. Acemel, R. D. *et al.* A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. *Nat. Genet.* **48,** 336–341 (2016).
130. Yue, J.-X., Yu, J.-K., Putnam, N. H. & Holland, L. Z. The transcriptome of an amphioxus, *Asymmetron lucayanum*, from the Bahamas: a window into chordate evolution. *Genome Biol. Evol.* **6,** 2681–2696 (2014).

131. Stach, T. The ontogeny of the notochord of *Branchiostoma lanceolatum*. *Acta Zool.* **80,** 25–33 (1999).
132. Olsson, R. The skin of Amphioxus. *Z. Für Zellforsch. Mikrosk. Anat.* **54,** 90–104
133. Sahlin, K. & Olsson, R. The wheel organ and Hatschek's groove in the lancelet, *Branchiostoma lanceolatum* (Cephalochordata). *Acta Zool.* **67,** 201–209 (1986).
134. Grassé, P.-P. *Traité de zoologie: anatomie, systématique, biologie - Tome XI, Echinodermes - Stomocordes - Procordes*. (Masson & Cie, 1948).
135. Ogasawara, M. Overlapping expression of amphioxus homologs of the thyroid transcription factor-1 gene and thyroid peroxidase gene in the endostyle: insight into evolution of the thyroid gland. *Dev. Genes Evol.* **210,** 231–242 (2000).
136. Fan, C. *et al.* Identification and expression of a novel class of glutathione-S-transferase from amphioxus *Branchiostoma belcheri* with implications to the origin of vertebrate liver. *Int. J. Biochem. Cell Biol.* **39,** 450–461 (2007).
137. Li, H.-Y. & Zhang, S.-C. Hepatic caecum of amphioxus and origin of vertebrate liver. *Hereditas* **32,** 437–442 (2010).
138. Garcia-Fernàndez, J. & Benito-Gutiérrez, È. It's a long way from amphioxus: descendants of the earliest chordate. *BioEssays* **31,** 665–675 (2009).
139. Mansfield, J. H., Haller, E., Holland, N. D. & Brent, A. E. Development of somites and their derivatives in amphioxus, and implications for the evolution of vertebrate somites. *EvoDevo* **6,** 21 (2015).
140. Rähr, H. The circulatory system of amphioxus (*Branchiostoma lanceolatum* (Pallas)). *Acta Zool.* **60,** 1–18 (1979).
141. Walter, H. E. & Sayles, L. P. *Biology of the vertebrates : a comparative study of man and his animal allies*. (New York : Macmillan Co., 1949).
142. Wicht, H. & Lacalli, T. C. The nervous system of amphioxus: structure, development, and evolutionary significance. *Can. J. Zool.* **83,** 122–150 (2005).
143. Kotpal, R. L. *Modern text book of zoology: vertebrates*. (Rastogi Publications, 2010).
144. Ruppert, E. E. Evolutionary origin of the vertebrate nephron. *Am. Zool.* **34,** 542–553 (1994).
145. Ruppert, E. E. Morphology of Hatschek's nephridium in larval and juvenile stages of *Branchiostoma virginiae* (cephalochordata). *Isr. J. Zool.* **42,** S161–S182 (1996).
146. Langeland, J. A. An amphioxus LIM-homeobox gene, *AmphiLim1/5* , expressed early in the invaginating organizer region and later in differentiating cells of the kidney and central nervous system. *Int. J. Biol. Sci.* 110–116 (2006). doi:10.7150/ijbs.2.110
147. Castro, A., Becerra, M., Manso, M. J. & Anadón, R. Neuronal organization of the brain in the adult amphioxus (*Branchiostoma lanceolatum*): a study with acetylated tubulin immunohistochemistry. *J. Comp. Neurol.* **523,** 2211–2232 (2015).
148. Theodosiou, M. *et al.* Amphioxus spawning behavior in an artificial seawater facility. *J. Exp. Zoolog. B Mol. Dev. Evol.* **316B,** 263–275 (2011).
149. Holland, L. Z. & Holland, N. D. Early development in the lancelet (= Amphioxus) *Branchiostoma floridae* from sperm entry through pronuclear fusion - presence of vegetal pole plasm and lack of conspicuous ooplasmic segregation. *Biol. Bull.* **182,** 77–96 (1992).
150. Holland, N. D. & Holland, L. Z. Fine structural study of the cortical reaction and formation of the egg coats in a lancelet (= amphioxus), *Branchiostoma floridae* (phylum Chordata: subphylum Cephalochordata = Acrania). *Biol. Bull.* **176,** 111–122 (1989).
151. Wu, H.-R. *et al.* Asymmetric localization of germline markers *Vasa* and *Nanos* during early development in the amphioxus *Branchiostoma floridae*. *Dev. Biol.* **353,** 147–159 (2011).
152. Hirakow, R. & Kajita, N. An electron microscopic study of the development of amphioxus, *Branchiostoma belcheri tsingtauense*: Cleavage. *J. Morphol.* **203,** 331–344 (1990).
153. Hatschek, B. *Studien über entwicklung des Amphioxus.* (A. Hölder, 1881).
154. Hirakow, R. & Kajita, N. Electron microscopic study of the development of amphioxus, *Branchiostoma belcheri tsingtauense*: The gastrula. *J. Morphol.* **207,** 37–52 (1991).
155. Tung, T. C., Wu, S. C. & Tung, Y. Y. F. Experimental studies on neural induction in Amphioxus. *Sci. Sin.* **7,** 263–270 (1961).
156. Tung, T. C., Wu, S. C. & Tung, Y. Y. F. The presumptive areas of the egg of amphioxus. *Sci. Sin.* **11,** 629–644 (1962).
157. Yu, J.-K. *et al.* Axial patterning in cephalochordates and the evolution of the organizer. *Nature* **445,** 613–617 (2007).
158. Onai, T., Yu, J.-K., Blitz, I. L., Cho, K. W. Y. & Holland, L. Z. Opposing Nodal/Vg1 and BMP signals mediate axial patterning in embryos of the basal chordate amphioxus. *Dev. Biol.* **344,** 377–389 (2010).
159. Bertrand, S. *et al.* Amphioxus FGF signaling predicts the acquisition of vertebrate morphological traits. *Proc. Natl. Acad. Sci.* **108,** 9160–9165 (2011).

160. Lu, T.-M., Luo, Y.-J. & Yu, J.-K. BMP and Delta/Notch signaling control the development of amphioxus epidermal sensory neurons: insights into the evolution of the peripheral sensory system. *Development* **139,** 2020–2030 (2012).
161. Hatschek, B. *The Amphioxus and its development*. (Swan Sonnenschein & Co., 1893).
162. Hirakow, R. & Kajita, N. Electron microscopic study of the development of amphioxus, *Branchiostoma belcheri tsingtauense*: the neurula and larva. *Kaibogaku Zasshi* **69,** 1–13 (1994).
163. Conklin, E. G. The embryology of amphioxus. *J. Morphol.* **54,** 69–151 (1932).
164. Kozmik, Z. *et al.* Characterization of an amphioxus paired box gene, *AmphiPax2/5/8*: developmental expression patterns in optic support cells, nephridium, thyroid-like structures and pharyngeal gill slits, but not in the midbrain-hindbrain boundary region. *Development* **126,** 1295–1304 (1999).
165. Kozmik, Z. *et al. Pax–Six–Eya–Dach* network during amphioxus development: conservation in vitro but context specificity in vivo. *Dev. Biol.* **306,** 143–159 (2007).
166. Holland, N. D. & Holland, L. Z. Stage- and tissue-specific patterns of cell division in embryonic and larval tissues of amphioxus during normal development. *Evol. Dev.* **8,** 142–149 (2006).
167. Kaji, T., Reimer, J. D., Morov, A. R., Kuratani, S. & Yasui, K. Amphioxus mouth after dorso-ventral inversion. *Zool. Lett.* **2,** (2016).
168. Stach, T. On the preoral pit of the larval amphioxus (*Branchiostoma lanceolatum*). *Ann. Sci. Nat. Zool. Biol. Anim.* **17,** 129–134 (1996).
169. Candiani, S., Holland, N. D., Oliveri, D., Parodi, M. & Pestarino, M. Expression of the amphioxus Pit-1 gene (*AmphiPOU1F1/Pit-1*) exclusively in the developing preoral organ, a putative homolog of the vertebrate adenohypophysis. *Brain Res. Bull.* **75,** 324–330 (2008).
170. Goodrich, E. S. Memoirs: The development of the club-shaped gland in amphioxus. *J. Cell Sci.* **s2-74,** 155–164 (1930).
171. Jefferies, R. P. S. *The ancestry of the vertebrates*. (Cambridge University Press, 1987).
172. Olsson, R. Club-shaped gland and endostyle in larval *Branchiostoma lanceolatum* (Cephalochordata). *Zoomorphology* **103,** 1–13 (1983).
173. Urata, M., Yamaguchi, N., Henmi, Y. & Yasui, K. Larval development of the Oriental lancelet, *Branchiostoma belcheri*, in laboratory mass culture. *Zoolog. Sci.* **24,** 787–797 (2007).
174. Holland, L. Z. & Yu, J.-K. Cephalochordate (amphioxus) embryos: procurement, culture, and basic methods. *Methods Cell Biol.* **74,** 195–215 (2004).
175. Wickstead, J. H. *Branchiostoma lanceolatum* larvae: some experiments on the effect of thiouracil on metamorphosis. *J. Mar. Biol. Assoc. U. K.* **47,** 49–59 (1967).
176. Kovalevsky, A. O. Entwicklungsgeschichte des *Amphioxus lanceolatus*. *Mém. Académie Impérial Sci. St Pétersbourg* **XI,** (1867).
177. Cerfontaine, P. *Recherches sur le développement de l'Amphioxus*. (Impr. H. Vaillant-Carmanne, 1906).
178. Stokes, M. D. & Holland, N. D. Embryos and larvae of a lancelet, *Branchiostoma floridae*, from hatching through metamorphosis: growth in the laboratory and external morphology. *Acta Zool.* **76,** 105–120 (1995).
179. Holland, N. D. & Holland, L. Z. Laboratory spawning and development of the Bahama lancelet, *Asymmetron lucayanum* (cephalochordata): fertilization through feeding larvae. *Biol. Bull.* **219,** 132–141 (2010).
180. Holland, N. D., Holland, L. Z. & Heimberg, A. Hybrids between the Florida amphioxus (*Branchiostoma floridae*) and the Bahamas lancelet (*Asymmetron lucayanum*): developmental morphology and chromosome counts. *Biol. Bull.* **228,** 13–24 (2015).
181. Zhang, Q.-J., Luo, Y.-J., Wu, H.-R., Chen, Y.-T. & Yu, J.-K. Expression of germline markers in three species of amphioxus supports a preformation mechanism of germ cell development in cephalochordates. *EvoDevo* **4,** 17 (2013).
182. Kaji, T., Hoshino, Y., Henmi, Y. & Yasui, K. Longitudinal observation of Japanese lancelet, *Branchiostoma japonicum*, metamorphosis. *Dataset Pap. Biol.* **2013,** 1–6 (2013).
183. Kon, T. *et al.* Phylogenetic position of a whale-fall lancelet (cephalochordata) inferred from whole mitochondrial genome sequences. *BMC Evol. Biol.* **7,** 127 (2007).
184. Nishikawa, T. A new deep-water lancelet (Cephalochordata) from off Cape Nomamisaki, SW Japan, with a proposal of the revised system recovering the genus *Asymmetron*. *Zoolog. Sci.* **21,** 1131–1136 (2004).
185. Poss, S. G. & Boschung, H. T. Lancelets (cephalochordata: Branchiostomatidae): how many species are valid? *Isr. J. Zool.* **42,** S13–S66 (1996).
186. Zhang, Q.-J., Zhong, J., Fang, S.-H. & Wang, Y.-Q. *Branchiostoma japonicum* and *B. belcheri* are distinct lancelets (cephalochordata) in Xiamen waters in China. *Zoolog. Sci.* **23,** 573–579 (2006).
187. Nohara, M., Nishida, M., Manthacitra, V. & Nishikawa, T. Ancient phylogenetic separation between pacific and atlantic cephalochordates as revealed by mitochondrial genome analysis. *Zoolog. Sci.* **21,** 203–210 (2004).

188. Bettany, G. T. The missing link between the vertebrates and invertebrates. *Nature* **14,** 195–196 (1876).
189. Goldschmidt, R. in *Wissenschaftliche Ergebnisse der Deutschen Tiefsee-Expedition auf dem Dampfer 'Valdivia' 1898-1899.* (ed. Chun, C.) **12,** 1–90 (G. Fischer, 1905).
190. Yuan, S., Ruan, J., Huang, S., Chen, S. & Xu, A. Amphioxus as a model for investigating evolution of the vertebrate immune system. *Dev. Comp. Immunol.* **48,** 297–305 (2015).
191. Somorjai, I. M. L., Escrivà, H. & Garcia-Fernàndez, J. Amphioxus makes the cut—Again. *Commun. Integr. Biol.* **5,** 499–502 (2012).
192. Kon, T., Nohara, M., Nishida, M., Sterrer, W. & Nishikawa, T. Hidden ancient diversification in the circumtropical lancelet *Asymmetron lucayanum*. *Mar. Biol.* **149,** 875–883 (2006).
193. Pallas, P. S. *Spicilegia zoologica : quibus novae imprimis et obscurae animalium species iconibus, descriptionibus atque commentariis illustrantur*. (Vol 1, Fascicle 10. GA Lange, 1774).
194. Costa, O. G. *Cenni zoologici ossia descrizione sommaria delle specie nuove di animali discoperti in diverse contrade del regno nell' anno 1834: con illustrazioni sopra talune altre meno ovvie*. (Azzolino E Comp., 1834).
195. Yarrell, W. *A history of British fishes*. **2,** (J. Van Voorst, 1836).
196. Fuentes, M. *et al.* Insights into spawning behavior and development of the European amphioxus (*Branchiostoma lanceolatum*). *J. Exp. Zoolog. B Mol. Dev. Evol.* **308B,** 484–493 (2007).
197. Gray, J. E. Description of a new species of amphioxus from Borneo. *Ann. Mag. Nat. Hist.* **19,** 463–464 (1847).
198. Li, G., Shu, Z. & Wang, Y. Year-round reproduction and induced spawning of Chinese amphioxus, *Branchiostoma belcheri*, in laboratory. *PLOS ONE* **8,** e75461 (2013).
199. Tung, T. C., Wu, S. C. & Tung, Y. Y. F. The development of isolated blastomeres of Amphioxus. *Sci. Sin.* **7,** 1280–1320 (1958).
200. Tung, T. C., Wu, S. C. & Tung, Y. Y. F. The developmental potencies of the blastomere layers in Amphioxus egg at the 32-cell stage. *Sci. Sin.* **9,** 119–141 (1960).
201. Tung, T. C., Wu, S. C. & Tung, Y. Y. F. Differentiation of the prospective ectodermal and entodermal cells after transplantation to new surroundings in Amphioxus. *Sci. Sin.* **7,** 253–262 (1961).
202. Tung, T. C., Wu, S. C. & Tung, Y. Y. F. The interchangeability of ectodermal and mesodermal cells in Amphioxus. *Sci. Sin.* 408–414 (1963).
203. Wu, S. C. & Cai, N. E. Further studies on the interactions of various parts of the developing Amphioxus egg. *Shi Yan Sheng Wu Xue Bao* 119–129 (1964).
204. Terazawa, K. & Satoh, N. Formation of the chordamesoderm in the amphioxus embryo: analysis with *Brachyury* and *fork head/HNF-3* genes. *Dev. Genes Evol.* **207,** 1–11 (1997).
205. Suzuki, M. M. & Satoh, N. Genes expressed in the amphioxus notochord revealed by EST analysis. *Dev. Biol.* **224,** 168–177 (2000).
206. Hubbs, C. L. A list of the lancelets of the world with diagnosis of five new species of Branchiostoma. *Occas. Pap. Mus. Zool. Univ. Mich.* **105,** (1922).
207. Howell, W. M. & Boschung, H. T. Chromosomes of the lancelet, *Branchiostoma floridae* (order amphioxi). *Experientia* **27,** 1495–1496 (1971).
208. Moller, P. C. & Ellis, R. A. Fine structure of the excretory system of Amphioxus (*Branchiostoma floridae*) and its response to osmotic stress. *Cell Tissue Res.* **148,** 1–9 (1974).
209. Moller, P. C. & Philpott, C. W. The circulatory system of Amphioxus (*Branchiostoma floridae*) I. Morphology of the major vessels of the pharyngeal area. *J. Morphol.* **139,** 389–406 (1973).
210. Andrews, E. A. *An undescribed acraniate:* Asymmetron lucayanum. (1893).
211. Holland, N. D. Spawning periodicity of the lancelet, *Asymmetron lucayanum* (Cephalochordata), in Bimini, Bahamas. *Ital. J. Zool.* **78,** 478–486 (2011).
212. Stokes, M. D. & Holland, N. D. Life-history characteristics of the Florida lancelet, *Branchiostoma floridae*: some factors affecting population dynamics in Tampa Bay. *Isr. J. Zool.* **42,** S67–S86 (1996).
213. Fuentes, M. *et al.* Preliminary observations on the spawning conditions of the European amphioxus (*Branchiostoma lanceolatum*) in captivity. *J. Exp. Zoolog. B Mol. Dev. Evol.* **302B,** 384–391 (2004).
214. Henmi, Y. & Yamaguchi, T. Biology of the amphioxus, *Branchiostoma belcheri* in the Ariake Sea, Japan I. Population structure and growth. *Zoolog. Sci.* **20,** 897–906 (2003).
215. Kubokawa, K., Azuma, N. & Tomiyama, M. A new population of the amphioxus (*Branchiostoma belcheri*) in the Enshu-Nada Sea in Japan. *Zoolog. Sci.* **15,** 799–803 (1998).
216. Kubokawa, K., Mizuta, T., Morisawa, M. & Azuma, N. Gonadal state of wild amphioxus populations and spawning success in captive conditions during the breeding period in Japan. *Zoolog. Sci.* **20,** 889–895 (2003).
217. Nishino, A., Kubokawa, K., Sekifuji, M., Azuma, N. & Morisawa, M. A survey of amphioxus (Cephalochordata: *Branchiostoma belcheri*) in the offing of Misaki. *Benthos Res.* **54,** 29–35 (1999).

218. Sanders, H. L., Hessler, R. R. & Hampson, G. R. An introduction to the study of deep-sea benthic faunal assemblages along the Gay Head-Bermuda transect. *Deep Sea Res. Oceanogr. Abstr.* **12,** 845–867 (1965).
219. Benito-Gutiérrez, È., Weber, H., Bryant, D. V. & Arendt, D. Methods for generating year-round access to amphioxus in the laboratory. *PLOS ONE* **8,** e71599 (2013).
220. Li, G., Yang, X., Shu, Z., Chen, X. & Wang, Y. Consecutive spawnings of Chinese amphioxus, *Branchiostoma belcheri*, in captivity. *PLOS ONE* **7,** e50838 (2012).
221. Somorjai, I. M. L., Camasses, A., Rivière, B. & Escrivà, H. Development of a semi-closed aquaculture system for monitoring of individual amphioxus (*Branchiostoma lanceolatum*), with high survivorship. *Aquaculture* **281,** 145–150 (2008).
222. Yasui, K., Urata, M., Yamaguchi, N., Ueda, H. & Henmi, Y. Laboratory culture of the oriental lancelet *Branchiostoma belcheri*. *Zoolog. Sci.* **24,** 514–520 (2007).
223. Yasui, K., Igawa, T., Kaji, T. & Henmi, Y. Stable aquaculture of the Japanese lancelet *Branchiostoma japonicum* for 7 years. *J. Exp. Zoolog. B Mol. Dev. Evol.* **320,** 538–547 (2013).
224. Yong, L. W. & Yu, J.-K. Tracing the evolutionary origin of vertebrate skeletal tissues: insights from cephalochordate amphioxus. *Curr. Opin. Genet. Dev.* **39,** 55–62 (2016).
225. Zhang, Q.-J. *et al.* Continuous culture of two lancelets and production of the second filial generations in the laboratory. *J. Exp. Zoolog. B Mol. Dev. Evol.* **308B,** 464–472 (2007).
226. Li, G., Wang, J., Yuan, L., Wang, H. & Wang, Y.-Q. A simple method for selecting spawning-ready individuals out from laboratorial cultured amphioxus population. *J. Exp. Zoolog. B Mol. Dev. Evol.* **324,** 629–635 (2015).
227. Desdevises, Y., Maillet, V., Fuentes, M. & Escriva, H. A snapshot of the population structure of *Branchiostoma lanceolatum* in the Racou beach, France, during its spawning season. *PLOS ONE* **6,** e18520 (2011).
228. Wu, X., Zhang, B., Guo, Z. & Qu, Y. Artificial culture of amphioxus (*Branchiostoma belcheri tsingtauense*). *Chin. J. Oceanol. Limnol.* **18,** 334–337 (2000).
229. Ruppert, E. E., Nash, T. R. & Smith, A. J. The size range of suspended particles trapped and ingested by the filter-feeding lancelet *Branchiostoma floridae* (Cephalochordata: Acrania). *J. Mar. Biol. Assoc. U. K.* **80,** 329–332 (2000).
230. Chen, Y., Cheung, S. G. & Shin, P. K. S. The diet of amphioxus in subtropical Hong Kong as indicated by fatty acid and stable isotopic analyses. *J. Mar. Biol. Assoc. U. K.* **88,** 1487–1491 (2008).
231. Pan, M., Yuan, D., Chen, S. & Xu, A. Diversity and composition of the bacterial community in amphioxus feces: amphioxus microbiome. *J. Basic Microbiol.* **55,** 1336–1342 (2015).
232. Carvalho, J. E. & Schubert, M. in *Vitamin-binding proteins: functional consequences* (eds. Dakshinamurti, K. & Dakshinamurti, S.) 1–30 (CRC Press, 2013).
233. Niederreither, K. & Dollé, P. Retinoic acid in development: towards an integrated view. *Nat. Rev. Genet.* **9,** 541–553 (2008).
234. Collins, M. D. & Mao, G. E. Teratology of retinoids. *Annu. Rev. Pharmacol. Toxicol.* **39,** 399–430 (1999).
235. Blomhoff, R. & Blomhoff, H. K. Overview of retinoid metabolism and function. *J. Neurobiol.* **66,** 606–630 (2006).
236. Glover, J. C., Renaud, J.-S. & Rijli, F. M. Retinoic acid and hindbrain patterning. *J. Neurobiol.* **66,** 705–725 (2006).
237. Mark, M., Ghyselinck, N. B. & Chambon, P. Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. *Annu. Rev. Pharmacol. Toxicol.* **46,** 451–480 (2006).
238. Duester, G. Retinoic acid synthesis and signaling during early organogenesis. *Cell* **134,** 921–931 (2008).
239. Theodosiou, M., Laudet, V. & Schubert, M. From carrot to clinic: an overview of the retinoic acid signaling pathway. *Cell. Mol. Life Sci.* **67,** 1423–1445 (2010).
240. Hall, J. A., Grainger, J. R., Spencer, S. P. & Belkaid, Y. The role of retinoic acid in tolerance and immunity. *Immunity* **35,** 13–22 (2011).
241. Noy, N. Between death and survival: retinoic acid in regulation of apoptosis. *Annu. Rev. Nutr.* **30,** 201–217 (2010).
242. Das, B. C. *et al.* Retinoic acid signaling pathways in development and diseases. *Bioorg. Med. Chem.* **22,** 673–683 (2014).
243. Hale, F. Pigs born without eye balls. *J. Hered.* **24,** 105–106 (1933).
244. Wilson, J. G. & Warkany, J. Aortic-arch and cardiac anomalies in the offspring of vitamin A deficient rats. *Am. J. Anat.* **85,** 113–155 (1949).
245. Wilson, J. G., Roth, C. B. & Warkany, J. An analysis of the syndrome of malformations induced by maternal vitamin a deficiency. Effects of restoration of vitamin a at various times during gestation. *Am. J. Anat.* **92,** 189–217 (1953).

246. Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. A human retinoic acid receptor which belongs to the family of nuclear receptors. *Nature* **330,** 444–450 (1987).
247. Giguere, V., Ong, E. S., Segui, P. & Evans, R. M. Identification of a receptor for the morphogen retinoic acid. *Nature* **330,** 624–629 (1987).
248. Kane, M. A. Analysis, occurrence, and function of 9-*cis*-retinoic acid. *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids* **1821,** 10–20 (2012).
249. Gutierrez-Mazariegos, J. *et al.* A mollusk retinoic acid receptor (RAR) ortholog sheds light on the evolution of ligand binding. *Endocrinology* **155,** 4275–4286 (2014).
250. Chambon, P. A decade of molecular biology of retinoic acid receptors. *FASEB J.* **10,** 940–954 (1996).
251. Fraser, P. D. & Bramley, P. M. The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid Res.* **43,** 228–265 (2004).
252. von Lintig, J. *et al.* Towards a better understanding of carotenoid metabolism in animals. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1740,** 122–131 (2005).
253. Hessel, S. *et al.* CMO1 deficiency abolishes vitamin A production from β-carotene and alters lipid metabolism in mice. *J. Biol. Chem.* **282,** 33553–33561 (2007).
254. Lampert, J. M. *et al.* Provitamin A conversion to retinal via theβ ,β-carotene-15,15′-oxygenase (*bcox*) is essential for pattern formation and differentiation during zebrafish embryogenesis. *Development* **130,** 2173–2186 (2003).
255. Kiefer, C. *et al.* Identification and characterization of a mammalian enzyme catalyzing the asymmetric oxidative cleavage of provitamin A. *J. Biol. Chem.* **276,** 14110–14116 (2001).
256. Lobo, G. P., Amengual, J., Palczewski, G., Babino, D. & von Lintig, J. Mammalian carotenoid-oxygenases: key players for carotenoid function and homeostasis. *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids* **1821,** 78–87 (2012).
257. Simões-costa, M. S., Azambuja, A. P. & Xavier-Neto, J. The search for non-chordate retinoic acid signaling: lessons from chordates. *J. Exp. Zoolog. B Mol. Dev. Evol.* **310B,** 54–72 (2008).
258. Spiegler, E., Kim, Y.-K., Wassef, L., Shete, V. & Quadro, L. Maternal–fetal transfer and metabolism of vitamin A and its precursor β-carotene in the developing tissues. *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids* **1821,** 88–98 (2012).
259. Wu, L. *et al.* Molecular aspects of β, β-carotene-9', 10'-oxygenase 2 in carotenoid metabolism and diseases. *Exp. Biol. Med. Maywood NJ* (2016). doi:10.1177/1535370216657900
260. Ross, A. C. Retinoid production and catabolism: role of diet in regulating retinol esterification and retinoic acid oxidation. *J. Nutr.* **133,** 291S–296S (2003).
261. D'Ambrosio, D. N., Clugston, R. D. & Blaner, W. S. Vitamin A metabolism: an update. *Nutrients* **3,** 63–103 (2011).
262. Noy, N. Signaling by retinol and its serum binding protein. *Prostaglandins Leukot. Essent. Fat. Acids PLEFA* **93,** 3–7 (2015).
263. Richardson, S. J. Evolutionary changes to transthyretin: evolution of transthyretin biosynthesis. *FEBS J.* **276,** 5342–5356 (2009).
264. Yamauchi, K. & Ishihara, A. Evolutionary changes to transthyretin: developmentally regulated and tissue-specific gene expression. *FEBS J.* **276,** 5357–5366 (2009).
265. Quadro, L. *et al.* Transplacental delivery of retinoid: the role of retinol-binding protein and lipoprotein retinyl ester. *Am. J. Physiol. - Endocrinol. Metab.* **286,** E844–E851 (2004).
266. Li, Z., Korzh, V. & Gong, Z. Localized *rbp4* expression in the yolk syncytial layer plays a role in yolk cell extension and early liver development. *BMC Dev. Biol.* **7,** 117 (2007).
267. Quadro, L. *et al.* Pathways of vitamin A delivery to the embryo: insights from a new tunable model of embryonic vitamin A deficiency. *Endocrinology* **146,** 4479–4490 (2005).
268. Barron, M., McAllister, D., Smith, S. M. & Lough, J. Expression of retinol binding protein and transthyretin during early embryogenesis. *Dev. Dyn.* **212,** 413–422 (1998).
269. Kim, Y.-K. *et al.* β-Carotene and its cleavage enzyme β-carotene-15,15′-oxygenase (CMOI) affect retinoid metabolism in developing tissues. *FASEB J.* **25,** 1641–1652 (2011).
270. Kawaguchi, R. *et al.* A membrane receptor for retinol binding protein mediates cellular uptake of vitamin A. *Science* **315,** 820–825 (2007).
271. Kawaguchi, R., Zhong, M., Kassai, M., Ter-Stepanian, M. & Sun, H. Vitamin A transport mechanism of the multitransmembrane cell-surface receptor STRA6. *Membranes* **5,** 425–453 (2015).
272. Chen, Y. *et al.* Structure of the STRA6 receptor for retinol uptake. *Science* **353,** aad8266-aad8266 (2016).
273. Zhong, M., Kawaguchi, R., Ter-Stepanian, M., Kassai, M. & Sun, H. Vitamin A transport and the transmembrane pore in the cell-surface receptor for plasma retinol binding protein. *PLOS ONE* **8,** e73838 (2013).
274. Parés, X., Farrés, J., Kedishvili, N. & Duester, G. Medium- and short-chain dehydrogenase/reductase gene and protein families. *Cell. Mol. Life Sci.* **65,** 3936–3949 (2008).

275. Farjo, K. M. *et al.* RDH10 is the primary enzyme responsible for the first step of embryonic Vitamin A metabolism and retinoic acid synthesis. *Dev. Biol.* **357,** 347–355 (2011).
276. Sandell, L. L. *et al.* RDH10 is essential for synthesis of embryonic retinoic acid and is required for limb, craniofacial, and organ development. *Genes Dev.* **21,** 1113–1124 (2007).
277. Feng, L., Hernandez, R. E., Waxman, J. S., Yelon, D. & Moens, C. B. Dhrs3a regulates retinoic acid biosynthesis through a feedback inhibition mechanism. *Dev. Biol.* **338,** 1–14 (2010).
278. Niederreither, K., Subbarayan, V., Dollé, P. & Chambon, P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat. Genet.* **21,** 444–448 (1999).
279. Mic, F. A., Haselbeck, R. J., Cuenca, A. E. & Duester, G. Novel retinoic acid generating activities in the neural tube and heart identified by conditional rescue of *Raldh2* null mutant mice. *Development* **129,** 2271–2282 (2002).
280. Begemann, G., Schilling, T. F., Rauch, G.-J., Geisler, R. & Ingham, P. W. The zebrafish *neckless* mutation reveals a requirement for *raldh2* in mesodermal signals that pattern the hindbrain. *Development* **128,** 3081–3094 (2001).
281. Chen, Y., Pollet, N., Niehrs, C. & Pieler, T. Increased XRALDH2 activity has a posteriorizing effect on the central nervous system of *Xenopus* embryos. *Mech. Dev.* **101,** 91–103 (2001).
282. Duester, G., Mic, F. A. & Molotkov, A. Cytosolic retinoid dehydrogenases govern ubiquitous metabolism of retinol to retinaldehyde followed by tissue-specific metabolism to retinoic acid. *Chem. Biol. Interact.* **143–144,** 201–210 (2003).
283. Rhinn, M. & Dollé, P. Retinoic acid signalling during development. *Development* **139,** 843–858 (2012).
284. Liang, D. *et al.* Expressions of Raldh3 and Raldh4 during zebrafish early development. *Gene Expr. Patterns* **8,** 248–253 (2008).
285. Lin, M., Zhang, M., Abraham, M., Smith, S. M. & Napoli, J. L. Mouse retinal dehydrogenase 4 (RALDH4), molecular cloning, cellular expression, and activity in 9-*cis*-retinoic acid biosynthesis in intact cells. *J. Biol. Chem.* **278,** 9856–9861 (2003).
286. Chambers, D., Wilson, L., Maden, M. & Lumsden, A. RALDH-independent generation of retinoic acid during vertebrate embryogenesis by CYP1B1. *Development* **134,** 1369–1383 (2007).
287. Faiq, M. A., Dada, R., Sharma, R., Saluja, D. & Dada, T. CYP1B1: a unique gene with unique characteristics. *Curr. Drug Metab.* **15,** 893–914 (2014).
288. Bempong, D. K., Honigberg, I. L. & Meltzer, N. M. Normal phase LC-MS determination of retinoic acid degradation products. *J. Pharm. Biomed. Anal.* **13,** 285–291 (1995).
289. Reijntjes, S., Gale, E. & Maden, M. Generating gradients of retinoic acid in the chick embryo: Cyp26C1 expression and a comparative analysis of the Cyp26 enzymes. *Dev. Dyn.* **230,** 509–517 (2004).
290. Ross, A. C. & Zolfaghari, R. Cytochrome P450s in the regulation of cellular retinoic acid metabolism. *Annu. Rev. Nutr.* **31,** 65–87 (2011).
291. White, R. J. & Schilling, T. F. How degrading: Cyp26s in hindbrain development. *Dev. Dyn.* **237,** 2775–2790 (2008).
292. Abu-Abed, S. *et al.* The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. *Genes Dev.* **15,** 226–240 (2001).
293. Yashiro, K. *et al.* Regulation of retinoic acid distribution is required for proximodistal patterning and outgrowth of the developing mouse limb. *Dev. Cell* **6,** 411–422 (2004).
294. MacLean, G., Li, H., Metzger, D., Chambon, P. & Petkovich, M. Apoptotic extinction of germ cells in testes of *Cyp26b1* knockout mice. *Endocrinology* **148,** 4560–4567 (2007).
295. Uehara, M. *et al.* CYP26A1 and CYP26C1 cooperatively regulate anterior–posterior patterning of the developing brain and the production of migratory cranial neural crest cells in the mouse. *Dev. Biol.* **302,** 399–411 (2007).
296. Noy, N. Retinoid-binding proteins: mediators of retinoid action. *Biochem. J.* **348,** 481–495 (2000).
297. Napoli, J. L. A gene knockout corroborates the integral function of cellular retinol-binding protein in retinoid metabolism. *Nutr. Rev.* **58,** 230–236 (2000).
298. Ghyselinck, N. B. *et al.* Cellular retinol-binding protein I is essential for vitamin A homeostasis. *EMBO J.* **18,** 4903–4914 (1999).
299. Matt, N. *et al.* Retinoic acid-dependent eye morphogenesis is orchestrated by neural crest cells. *Development* **132,** 4789–4800 (2005).
300. Suruga, K. *et al.* Cloning of chick cellular retinol-binding protein, type II and comparison to that of some mammals: expression of the gene at different developmental stages, and possible involvement of RXRs and PPAR. *Comp. Biochem. Physiol. A Physiol.* **118,** 859–869 (1997).
301. Napoli, J. L. Interactions of retinoid binding proteins and enzymes in retinoid metabolism. *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids* **1440,** 139–162 (1999).
302. Nelson, C. H. *et al.* Direct protein–protein interactions and substrate channeling between cellular retinoic acid binding proteins and CYP26B1. *FEBS Lett.* **590,** 2527–2535 (2016).

303. Budhu, A. S. & Noy, N. Direct channeling of retinoic acid between cellular retinoic acid-binding protein II and retinoic acid receptor sensitizes mammary carcinoma cells to retinoic acid-induced growth arrest. *Mol. Cell. Biol.* **22,** 2632–2641 (2002).
304. Dong, D., Ruuska, S. E., Levinthal, D. J. & Noy, N. Distinct roles for cellular retinoic acid-binding proteins I and II in regulating signaling by retinoic acid. *J. Biol. Chem.* **274,** 23695–23698 (1999).
305. Cai, A. Q. *et al.* Cellular retinoic acid-binding proteins are essential for hindbrain patterning and signal robustness in zebrafish. *Development* **139,** 2150–2155 (2012).
306. Gronemeyer, H., Gustafsson, J.-Å. & Laudet, V. Principles for modulation of the nuclear receptor superfamily. *Nat. Rev. Drug Discov.* **3,** 950–964 (2004).
307. Linney, E., Donerly, S., Mackey, L. & Dobbs-McAuliffe, B. The negative side of retinoic acid receptors. *Neurotoxicol. Teratol.* **33,** 631–640 (2011).
308. Waxman, J. S. & Yelon, D. Comparison of the expression patterns of newly identified zebrafish retinoic acid and retinoid X receptors. *Dev. Dyn.* **236,** 587–595 (2007).
309. Mic, F. A., Molotkov, A., Benbrook, D. M. & Duester, G. Retinoid activation of retinoic acid receptor but not retinoid X receptor is sufficient to rescue lethal defect in retinoic acid synthesis. *Proc. Natl. Acad. Sci.* **100,** 7135–7140 (2003).
310. Balmer, J. E. & Blomhoff, R. A robust characterization of retinoic acid response elements based on a comparison of sites in three species. *J. Steroid Biochem. Mol. Biol.* **96,** 347–354 (2005).
311. Ross, S. A., McCaffery, P. J., Drager, U. C. & Luca, L. M. D. Retinoids in embryonal development. *Physiol. Rev.* **80,** 1021–1054 (2000).
312. Chatagnon, A. *et al.* RAR/RXR binding dynamics distinguish pluripotency from differentiation associated cis-regulatory elements. *Nucleic Acids Res.* **43,** 4833–4854 (2015).
313. Vilhais-Neto, G. C. & Pourquié, O. Retinoic acid. *Curr. Biol.* **18,** R191–R192 (2008).
314. Matt, N., Ghyselinck, N. B., Wendling, O., Chambon, P. & Mark, M. Retinoic acid-induced developmental defects are mediated by RARβ/RXR heterodimers in the pharyngeal endoderm. *Development* **130,** 2083–2093 (2003).
315. Mark, M., Ghyselinck, N. B. & Chambon, P. Function of retinoic acid receptors during embryonic development. *Nucl. Recept. Signal.* **7,** (2009).
316. Krezel, W. *et al.* RXRγ null mice are apparently normal and compound $RXR\alpha^{+/-}$ /$RXR\beta^{-/-}$ /$RXR\gamma^{-/-}$ mutant mice are viable. *Proc. Natl. Acad. Sci.* **93,** 9010–9014 (1996).
317. Samarut, E. & Rochette-Egly, C. Nuclear retinoic acid receptors: conductors of the retinoic acid symphony during development. *Mol. Cell. Endocrinol.* **348,** 348–360 (2012).
318. Samarut, E., Fraher, D., Laudet, V. & Gibert, Y. ZebRA: an overview of retinoic acid signaling during zebrafish development. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1849,** 73–83 (2015).
319. Shaw, N., Elholm, M. & Noy, N. Retinoic acid is a high affinity selective ligand for the peroxisome proliferator-activated receptor β/δ. *J. Biol. Chem.* **278,** 41589–41592 (2003).
320. Tan, N.-S. *et al.* Selective cooperation between fatty acid binding proteins and peroxisome proliferator-activated receptors in regulating transcription. *Mol. Cell. Biol.* **22,** 5114–5127 (2002).
321. Schug, T. T., Berry, D. C., Shaw, N. S., Travis, S. N. & Noy, N. Opposing effects of retinoic acid on cell growth result from alternate activation of two different nuclear receptors. *Cell* **129,** 723–733 (2007).
322. Berry, D. C. & Noy, N. All-*trans*-retinoic acid represses obesity and insulin resistance by activating both peroxisome proliferation-activated receptor β/δ and retinoic acid receptor. *Mol. Cell. Biol.* **29,** 3286–3296 (2009).
323. Zhou, X. E. *et al.* The orphan nuclear receptor TR4 is a vitamin A-activated nuclear receptor. *J. Biol. Chem.* **286,** 2877–2885 (2011).
324. Kruse, S. W. *et al.* Identification of COUP-TFII orphan nuclear receptor as a retinoic acid–activated receptor. *PLOS Biol* **6,** e227 (2008).
325. Stehlin-Gaon, C. *et al.* All-trans retinoic acid is a ligand for the orphan nuclear receptor RORβ. *Nat. Struct. Mol. Biol.* **10,** 820–825 (2003).
326. McNamara, P. *et al.* Regulation of CLOCK and MOP4 by nuclear hormone receptors in the vasculature: a humoral mechanism to reset a peripheral clock. *Cell* **105,** 877–889 (2001).
327. Shirai, H., Oishi, K. & Ishida, N. Bidirectional CLOCK/BMAL1-dependent circadian gene regulation by retinoic acid *in vitro*. *Biochem. Biophys. Res. Commun.* **351,** 387–391 (2006).
328. Ochoa, W. F. *et al.* Retinoic acid binds to the C2-domain of protein kinase Cα. *Biochemistry (Mosc.)* **42,** 8774–8779 (2003).
329. Aggarwal, S. *et al.* Nonclassical action of retinoic acid on the activation of the cAMP response element-binding protein in normal human bronchial epithelial cells. *Mol. Biol. Cell* **17,** 566–575 (2006).
330. Zanotto-Filho, A. *et al.* Retinoic acid induces apoptosis by a non-classical mechanism of ERK1/2 activation. *Toxicol. In Vitro* **22,** 1205–1212 (2008).

331. López-Carballo, G., Moreno, L., Masiá, S., Pérez, P. & Barettino, D. Activation of the phosphatidylinositol 3-kinase/Akt signaling pathway by retinoic acid is required for neural differentiation of SH-SY5Y human neuroblastoma cells. *J. Biol. Chem.* **277,** 25297–25304 (2002).
332. Nagl, F. *et al.* Retinoic acid-induced nNOS expression depends on a novel PI3K/Akt/DAX1 pathway in human TGW-nu-I neuroblastoma cells. *Am. J. Physiol. - Cell Physiol.* **297,** C1146–C1156 (2009).
333. Piskunov, A. & Rochette-Egly, C. A retinoic acid receptor RARα pool present in membrane lipid rafts forms complexes with G protein αQ to activate p38MAPK. *Oncogene* **31,** 3333–3345 (2012).
334. Duong, V. & Rochette-Egly, C. The molecular physiology of nuclear retinoic acid receptors. From health to disease. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1812,** 1023–1031 (2011).
335. Gutierrez-Mazariegos, J., Theodosiou, M., Campo-Paysaa, F. & Schubert, M. Vitamin A: a multifunctional tool for development. *Semin. Cell Dev. Biol.* **22,** 603–610 (2011).
336. Yasuda, Y. *et al.* Induction of avascular yolk sac due to reduction of basic fibroblast growth factor by retinoic acid in mice. *Dev. Biol.* **150,** 397–413 (1992).
337. Ribes, V., Roux, I. L., Rhinn, M., Schuhbaur, B. & Dollé, P. Early mouse caudal development relies on crosstalk between retinoic acid, Shh and Fgf signalling pathways. *Development* **136,** 665–676 (2009).
338. Engberg, N., Kahn, M., Petersen, D. R., Hansson, M. & Serup, P. Retinoic acid synthesis promotes development of neural progenitors from mouse embryonic stem cells by suppressing endogenous, Wnt-dependent nodal signaling. *STEM CELLS* **28,** 1498–1509 (2010).
339. Sheng, N. *et al.* Retinoic acid regulates bone morphogenic protein signal duration by promoting the degradation of phosphorylated Smad1. *Proc. Natl. Acad. Sci.* **107,** 18886–18891 (2010).
340. Pourquié, O. The segmentation clock: converting embryonic time into spatial pattern. *Science* **301,** 328–330 (2003).
341. Zhang, Z., Balmer, J. E., Løvlie, A., Fromm, S. H. & Blomhoff, R. Specific teratogenic effects of different retinoic acid isomers and analogs in the developing anterior central nervous system of zebrafish. *Dev. Dyn.* **206,** 73–86 (1996).
342. Avantaggiato, V., Acampora, D., Tuorto, F. & Simeone, A. Retinoic acid induces stage-specific repatterning of the rostral central nervous system. *Dev. Biol.* **175,** 347–357 (1996).
343. White, J. C., Highland, M., Kaiser, M. & Clagett-Dame, M. Vitamin A deficiency results in the dose-dependent acquisition of anterior character and shortening of the caudal hindbrain of the rat embryo. *Dev. Biol.* **220,** 263–284 (2000).
344. Durston, A. J. *et al.* Retinoic acid causes an anteroposterior transformation in the developing central nervous system. *Nature* **340,** 140–144 (1989).
345. Hernandez, R. E., Putzke, A. P., Myers, J. P., Margaretha, L. & Moens, C. B. Cyp26 enzymes generate the retinoic acid response pattern necessary for hindbrain development. *Development* **134,** 177–187 (2007).
346. Sirbu, I. O., Gresh, L., Barra, J. & Duester, G. Shifting boundaries of retinoic acid activity control hindbrain segmental gene expression. *Development* **132,** 2611–2622 (2005).
347. Rijli, F. M., Gavalas, A. & Chambon, P. Segmentation and specification in the branchial region of the head: the role of the *Hox* selector genes. *Int. J. Dev. Biol.* **42,** 393–401 (1998).
348. Stoney, P. N. *et al.* Expression of the retinoic acid catabolic enzyme CYP26B1 in the human brain to maintain signaling homeostasis. *Brain Struct. Funct.* **221,** 3315–3326 (2015).
349. Marshall, H., Morrison, A., Studer, M., Pöpperl, H. & Krumlauf, R. Retinoids and Hox genes. *FASEB J.* **10,** 969–978 (1996).
350. del Corral, R. D. & Storey, K. G. Opposing FGF and retinoid pathways: a signalling switch that controls differentiation and patterning onset in the extending vertebrate body axis. *BioEssays* **26,** 857–869 (2004).
351. Irving, C. & Mason, I. Signalling by FGF8 from the isthmus patterns anterior hindbrain and establishes the anterior limit of Hox gene expression. *Development* **127,** 177–186 (2000).
352. Liu, J.-P., Laufer, E. & Jessell, T. M. Assigning the positional identity of spinal motor neurons: rostrocaudal patterning of Hox-c expression by FGFs, Gdf11, and retinoids. *Neuron* **32,** 997–1012 (2001).
353. Kin Ting Kam, R., Deng, Y., Chen, Y. & Zhao, H. Retinoic acid synthesis and functions in early embryonic development. *Cell Biosci.* **2,** 11 (2012).
354. Maden, M. Retinoid signalling in the development of the central nervous system. *Nat. Rev. Neurosci.* **3,** 843–853 (2002).
355. Novitch, B. G., Wichterle, H., Jessell, T. M. & Sockanathan, S. A requirement for retinoic acid-mediated transcriptional activation in ventral neural patterning and motor neuron specification. *Neuron* **40,** 81–95 (2003).
356. Molotkova, N., Molotkov, A., Sirbu, I. O. & Duester, G. Requirement of mesodermal retinoic acid generated by *Raldh2* for posterior neural transformation. *Mech. Dev.* **122,** 145–155 (2005).
357. Corcoran, J., Shroot, B., Pizzey, J. & Maden, M. The role of retinoic acid receptors in neurite outgrowth from different populations of embryonic mouse dorsal root ganglia. *J Cell Sci* **113,** 2567–2574 (2000).

358. Martínez-Morales, P. L. *et al.* FGF and retinoic acid activity gradients control the timing of neural crest cell emigration in the trunk. *J. Cell Biol.* **194,** 489–503 (2011).
359. Krispin, S., Nitzan, E., Kassem, Y. & Kalcheim, C. Evidence for a dynamic spatiotemporal fate map and early fate restrictions of premigratory avian neural crest. *Development* **137,** 585–595 (2010).
360. Hall, B. K. *The neural crest and neural crest cells in vertebrate development and evolution*. (Springer US, 2009).
361. Minoux, M. & Rijli, F. M. Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. *Development* **137,** 2605–2621 (2010).
362. Trainor, P. A. & Krumlauf, R. *Hox* genes, neural crest cells and branchial arch patterning. *Curr. Opin. Cell Biol.* **13,** 698–705 (2001).
363. Plant, M. R., MacDonald, M. P., Grad, L. I., Ritchie, S. J. & Richman, J. M. Locally released retinoic acid repatterns the first branchial arch cartilages *in vivo*. *Dev. Biol.* **222,** 12–26 (2000).
364. Usami, M., Mitsunaga, K., Irie, T., Miyajima, A. & Doi, O. Simple *in vitro* migration assay for neural crest cells and the opposite effects of all-*trans*-retinoic acid on cephalic- and trunk-derived cells. *Congenit. Anom.* **54,** 184–188 (2014).
365. Chawla, B., Schley, E., Williams, A. L. & Bohnsack, B. L. Retinoic acid and Pitx2 regulate early neural crest survival and migration in craniofacial and ocular development. *Birth Defects Res. B. Dev. Reprod. Toxicol.* **107,** 126–135 (2016).
366. Jimenez, L. *et al.* Phenotypic chemical screening using a zebrafish neural crest EMT reporter identifies retinoic acid as an inhibitor of epithelial morphogenesis. *Dis. Model. Mech.* **9,** 389–400 (2016).
367. Schlosser, G. in *International Review of Cell and Molecular Biology* (ed. Jeon, K.) **283,** 129–234 (Academic Press, 2010).
368. Dupé, V., Ghyselinck, N. B., Wendling, O., Chambon, P. & Mark, M. Key roles of retinoic acid receptors alpha and beta in the patterning of the caudal hindbrain, pharyngeal arches and otocyst in the mouse. *Development* **126,** 5051–5059 (1999).
369. Song, Y., Hui, J. N., Fu, K. K. & Richman, J. M. Control of retinoic acid synthesis and FGF expression in the nasal pit is required to pattern the craniofacial skeleton. *Dev. Biol.* **276,** 313–329 (2004).
370. Gibbs, M. A. & Northcutt, R. G. Retinoic acid repatterns axolotl lateral line receptors. *Int. J. Dev. Biol.* **48,** 63–66 (2004).
371. McCaffery, P., Wagner, E., O'Neil, J., Petkovich, M. & Dräger, U. C. Dorsal and ventral retinal territories defined by retinoic acid synthesis, break-down and nuclear receptor expression. *Mech. Dev.* **82,** 119–130 (1999).
372. Wasiak, S. & Lohnes, D. Retinoic acid affects left–right patterning. *Dev. Biol.* **215,** 332–342 (1999).
373. Tanaka, Y., Okada, Y. & Hirokawa, N. FGF-induced vesicular release of Sonic hedgehog and retinoic acid in leftward nodal flow is critical for left–right determination. *Nature* **435,** 172–177 (2005).
374. Sirbu, I. O. & Duester, G. Retinoic-acid signalling in node ectoderm and posterior neural plate directs left–right patterning of somitic mesoderm. *Nat. Cell Biol.* **8,** 271–277 (2006).
375. Vermot, J. & Pourquié, O. Retinoic acid coordinates somitogenesis and left–right patterning in vertebrate embryos. *Nature* **435,** 215–220 (2005).
376. Brent, A. E. & Tabin, C. J. Developmental regulation of somite derivatives: muscle, cartilage and tendon. *Curr. Opin. Genet. Dev.* **12,** 548–557 (2002).
377. Dubrulle, J. & Pourquié, O. Coupling segmentation to axis formation. *Development* **131,** 5783–5793 (2004).
378. Mallo, M. Revisiting the involvement of signaling gradients in somitogenesis. *FEBS J.* **283,** 1430–1437 (2016).
379. Duester, G. Retinoic acid regulation of the somitogenesis clock. *Birth Defects Res. Part C Embryo Today Rev.* **81,** 84–92 (2007).
380. Moreno, T. A. & Kintner, C. Regulation of segmental patterning by retinoic acid signaling during *Xenopus* somitogenesis. *Dev. Cell* **6,** 205–218 (2004).
381. Kumar, S., Cunningham, T. J. & Duester, G. Nuclear receptor corepressors *Ncor1* and *Ncor2* (*Smrt*) are required for retinoic acid-dependent repression of *Fgf8* during somitogenesis. *Dev. Biol.* **418,** 204–215 (2016).
382. Xavier-Neto, J. *et al.* Retinoid signaling and cardiac anteroposterior segmentation. *genesis* **31,** 97–104 (2001).
383. Rosenthal, N. & Xavier-Neto, J. From the bottom of the heart: anteroposterior decisions in cardiac muscle differentiation. *Curr. Opin. Cell Biol.* **12,** 742–746 (2000).
384. Chazaud, C., Chambon, P. & Dolle, P. Retinoic acid is required in the mouse embryo for left-right asymmetry determination and heart morphogenesis. *Development* **126,** 2589–2596 (1999).
385. El Robrini, N. *et al.* Cardiac outflow morphogenesis depends on effects of retinoic acid signaling on multiple cell lineages. *Dev. Dyn.* **245,** 388–401 (2016).

386. Narematsu, M., Kamimura, T., Yamagishi, T., Fukui, M. & Nakajima, Y. Impaired development of left anterior heart field by ectopic retinoic acid causes transposition of the great arteries. *J. Am. Heart Assoc.* **4,** e001889 (2015).
387. D'Aniello, E. & Waxman, J. S. Input overload: contributions of retinoic acid signaling feedback mechanisms to heart development and teratogenesis. *Dev. Dyn.* **244,** 513–523 (2015).
388. Xavier-Neto, J. *et al.* Signaling through retinoic acid receptors in cardiac development: doing the right things at the right times. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1849,** 94–111 (2015).
389. Keegan, B. R., Feldman, J. L., Begemann, G., Ingham, P. W. & Yelon, D. Retinoic acid signaling restricts the cardiac progenitor pool. *Science* **307,** 247–249 (2005).
390. Waxman, J. S., Keegan, B. R., Roberts, R. W., Poss, K. D. & Yelon, D. Hoxb5b acts downstream of retinoic acid signaling in the forelimb field to restrict heart field potential in zebrafish. *Dev. Cell* **15,** 923–934 (2008).
391. Lin, S.-C. *et al.* Endogenous retinoic acid regulates cardiac progenitor differentiation. *Proc. Natl. Acad. Sci.* **107,** 9234–9239 (2010).
392. Hochgreb, T. *et al.* A caudorostral wave of RALDH2 conveys anteroposterior information to the cardiac field. *Development* **130,** 5363–5374 (2003).
393. Stainier, D. Y. R. & Fishman, M. C. Patterning the zebrafish heart tube: acquisition of anteroposterior polarity. *Dev. Biol.* **153,** 91–101 (1992).
394. Sirbu, I. O., Zhao, X. & Duester, G. Retinoic acid controls heart anteroposterior patterning by down-regulating *Isl1* through the *Fgf8* pathway. *Dev. Dyn.* **237,** 1627–1635 (2008).
395. Serluca, F. C. & Fishman, M. C. Pre-pattern in the pronephric kidney field of zebrafish. *Development* **128,** 2233–2241 (2001).
396. Cartry, J. *et al.* Retinoic acid signalling is required for specification of pronephric cell fate. *Dev. Biol.* **299,** 35–51 (2006).
397. Wingert, R. A. *et al.* The *cdx* genes and retinoic acid control the positioning and segmentation of the zebrafish pronephros. *PLOS Genet* **3,** e189 (2007).
398. Bollig, F. *et al.* A highly conserved retinoic acid responsive element controls *wt1a* expression in the zebrafish pronephros. *Development* **136,** 2883–2892 (2009).
399. Lee, S. J., Kim, S., Choi, S.-C. & Han, J.-K. *XPteg* (*Xenopus proximal tubules-expressed gene*) is essential for pronephric mesoderm specification and tubulogenesis. *Mech. Dev.* **127,** 49–61 (2010).
400. Rosselot, C. *et al.* Non-cell-autonomous retinoid signaling is crucial for renal development. *Development* **137,** 283–292 (2010).
401. Tickle, C., Alberts, B., Wolpert, L. & Lee, J. Local application of retinoic acid to the limb bond mimics the action of the polarizing region. *Nature* **296,** 564–566 (1982).
402. Campo-Paysaa, F., Marlétaz, F., Laudet, V. & Schubert, M. Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. *Genesis* **46,** 640–656 (2008).
403. Cunningham, T. J. & Duester, G. Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat. Rev. Mol. Cell Biol.* **16,** 110–123 (2015).
404. Zhao, X. *et al.* Retinoic acid promotes limb induction through effects on body axis extension but is unnecessary for limb patterning. *Curr. Biol.* **19,** 1050–1057 (2009).
405. Gibert, Y., Gajewski, A., Meyer, A. & Begemann, G. Induction and prepatterning of the zebrafish pectoral fin bud requires axial retinoic acid signaling. *Development* **133,** 2649–2659 (2006).
406. Grandel, H. & Brand, M. Zebrafish limb development is triggered by a retinoic acid signal during gastrulation. *Dev. Dyn.* **240,** 1116–1126 (2011).
407. Zhao, X., Brade, T., Cunningham, T. J. & Duester, G. Retinoic acid controls expression of tissue remodeling genes *Hmgn1* and *Fgf18* at the digit–interdigit junction. *Dev. Dyn.* **239,** 665–671 (2010).
408. Dupé, V. *et al.* Essential roles of retinoic acid signaling in interdigital apoptosis and control of BMP-7 expression in mouse autopods. *Dev. Biol.* **208,** 30–43 (1999).
409. Cunningham, T. J., Zhao, X. & Duester, G. Uncoupling of retinoic acid signaling from tailbud development before termination of body axis extension. *genesis* **49,** 776–783 (2011).
410. Bayha, E., Jørgensen, M. C., Serup, P. & Grapin-Botton, A. Retinoic acid signaling organizes endodermal organ specification along the entire antero-posterior axis. *PLOS ONE* **4,** e5845 (2009).
411. Niederreither, K. *et al.* The regional pattern of retinoic acid synthesis by RALDH2 is essential for the development of posterior pharyngeal arches and the enteric nervous system. *Development* **130,** 2525–2534 (2003).
412. Kopinke, D., Sasine, J., Swift, J., Stephens, W. Z. & Piotrowski, T. Retinoic acid is required for endodermal pouch morphogenesis and not for pharyngeal endoderm specification. *Dev. Dyn.* **235,** 2695–2709 (2006).
413. Wang, Z., Dollé, P., Cardoso, W. V. & Niederreither, K. Retinoic acid regulates morphogenesis and patterning of posterior foregut derivatives. *Dev. Biol.* **297,** 433–445 (2006).

414. Stafford, D., Hornbruch, A., Mueller, P. R. & Prince, V. E. A conserved role for retinoid signaling in vertebrate pancreas development. *Dev. Genes Evol.* **214,** 432–441 (2004).
415. Molotkov, A., Molotkova, N. & Duester, G. Retinoic acid generated by *Raldh2* in mesoderm is required for mouse dorsal endodermal pancreas development. *Dev. Dyn.* **232,** 950–957 (2005).
416. Cañestro, C., Postlethwait, J. H., Gonzàlez-Duarte, R. & Albalat, R. Is retinoic acid genetic machinery a chordate innovation? *Evol. Dev.* **8,** 394–406 (2006).
417. Marlétaz, F., Holland, L. Z., Laudet, V. & Schubert, M. Retinoic acid signaling and the evolution of chordates. *Int. J. Biol. Sci.* **2,** 38–47 (2006).
418. Escriva, H., Holland, N. D., Gronemeyer, H., Laudet, V. & Holland, L. Z. The retinoic acid signaling pathway regulates anterior/posterior patterning in the nerve cord and pharynx of amphioxus, a chordate lacking neural crest. *Development* **129,** 2905–2916 (2002).
419. Escriva, H. *et al.* Neofunctionalization in vertebrates: the example of retinoic acid receptors. *PLOS Genet.* **2,** (2006).
420. Gutierrez-Mazariegos, J. *et al.* Evolutionary diversification of retinoic acid receptor ligand-binding pocket structure by molecular tinkering. *R. Soc. Open Sci.* **3,** 150484 (2016).
421. Sobreira, T. J. P. *et al.* Structural shifts of aldehyde dehydrogenase enzymes were instrumental for the early evolution of retinoid-dependent axial patterning in metazoans. *Proc. Natl. Acad. Sci.* **108,** 226–231 (2011).
422. Cañestro, C., Godoy, L., Gonzàlez-Duarte, R. & Albalat, R. Comparative expression analysis of *Adh3* during arthropod, urochordate, cephalochordate, and vertebrate development challenges its predicted housekeeping role. *Evol. Dev.* **5,** 157–162 (2003).
423. Dalfó, D., Marqués, N. & Albalat, R. Analysis of the NADH-dependent retinaldehyde reductase activity of amphioxus retinol dehydrogenase enzymes enhances our understanding of the evolution of the retinol dehydrogenase family. *FEBS J.* **274,** 3739–3752 (2007).
424. Cañestro, C. *et al.* Amphioxus alcohol dehydrogenase is a class 3 form of single type and of structural conservation but with unique developmental expression. *Eur. J. Biochem.* **267,** 6511–6518 (2000).
425. Jackman, W. R., Mougey, J. M., Panopoulou, G. D. & Kimmel, C. B. *crabp* and *maf* highlight the novelty of the amphioxus club-shaped gland. *Acta Zool.* **85,** 91–99 (2004).
426. Holland, L. Z. & Holland, N. D. Expression of *AmphiHox-1* and *AmphiPax-1* in amphioxus embryos treated with retinoic acid: insights into evolution and patterning of the chordate nerve cord and pharynx. *Development* **122,** 1829–1838 (1996).
427. Schubert, M., Holland, N. D., Laudet, V. & Holland, L. Z. A retinoic acid-*Hox* hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. *Dev. Biol.* **296,** 190–202 (2006).
428. Schubert, M., Holland, N. D., Escriva, H., Holland, L. Z. & Laudet, V. Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and their gene expression in a developing chordate (amphioxus). *Proc. Natl. Acad. Sci. U. S. A.* **101,** 10320–10325 (2004).
429. Koop, D., Holland, L. Z., Setiamarga, D., Schubert, M. & Holland, N. D. Tail regression induced by elevated retinoic acid signaling in amphioxus larvae occurs by tissue remodeling, not cell death. *Evol. Dev.* **13,** 427–435 (2011).
430. Schubert, M. *et al.* Retinoic acid signaling acts via Hox1 to establish the posterior limit of the pharynx in the chordate amphioxus. *Development* **132,** 61–73 (2005).
431. Osborne, P. W., Benoit, G., Laudet, V., Schubert, M. & Ferrier, D. E. K. Differential regulation of ParaHox genes by retinoic acid in the invertebrate chordate amphioxus (*Branchiostoma floridae*). *Dev. Biol.* **327,** 252–262 (2009).
432. Koop, D. *et al.* Retinoic acid signaling targets *Hox* genes during the amphioxus gastrula stage: insights into early anterior–posterior patterning of the chordate body plan. *Dev. Biol.* **338,** 98–106 (2010).
433. Martí-Solans, J. *et al.* Coelimination and survival in gene network evolution: dismantling the RA-signaling in a chordate. *Mol. Biol. Evol.* **33,** 2401–2416 (2016).
434. Cañestro, C., Yokoi, H. & Postlethwait, J. H. Evolutionary developmental biology and genomics. *Nat. Rev. Genet.* **8,** 932–942 (2007).
435. Nagatomo, K., Ishibashi, T., Satou, Y., Satoh, N. & Fujiwara, S. Retinoic acid affects gene expression and morphogenesis without upregulating the retinoic acid receptor in the ascidian *Ciona intestinalis*. *Mech. Dev.* **120,** 363–372 (2003).
436. Cañestro, C. & Postlethwait, J. H. Development of a chordate anterior–posterior axis without classical retinoic acid signaling. *Dev. Biol.* **305,** 522–538 (2007).
437. De Bernardi, F., Sotgia, C. & Ortolani, G. Retinoic acid treatment of ascidian embryos: effects on larvae and metamorphosis. *Anim Biol* **3,** 75–81 (1994).
438. Nagatomo, K. & Fujiwara, S. Expression of *Raldh2*, *Cyp26* and *Hox-1* in normal and retinoic acid-treated *Ciona intestinalis* embryos. *Gene Expr. Patterns* **3,** 273–277 (2003).

439. Natale, A. *et al.* Evolution of anterior *Hox* regulatory elements among chordates. *BMC Evol. Biol.* **11,** 330 (2011).
440. Wada, H., Escriva, H., Zhang, S. & Laudet, V. Conserved RARE localization in amphioxus *Hox* clusters and implications for *Hox* code evolution in the vertebrate neural crest. *Dev. Dyn.* **235,** 1522–1531 (2006).
441. Kanda, M., Ikeda, T. & Fujiwara, S. Identification of a retinoic acid-responsive neural enhancer in the *Ciona intestinalis Hox1* gene. *Dev. Growth Differ.* **55,** 260–269 (2013).
442. Kanda, M., Wada, H. & Fujiwara, S. Epidermal expression of *Hox1* is directly activated by retinoic acid in the *Ciona intestinalis* embryo. *Dev. Biol.* **335,** 454–463 (2009).
443. Ikuta, T., Yoshida, N., Satoh, N. & Saiga, H. *Ciona intestinalis* Hox gene cluster: its dispersed structure and residual colinear expression in development. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 15118–15123 (2004).
444. Seo, H.-C. *et al. Hox* cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* **431,** 67–71 (2004).
445. Katsuyama, Y. & Saiga, H. Retinoic acid affects patterning along the anterior–posterior axis of the ascidian embryo. *Dev. Growth Differ.* **40,** 413–422 (1998).
446. Katsuyama, Y., Wada, S., Yasugi, S. & Saiga, H. Expression of the *labial* group Hox gene *HrHox-1* and its alteration induced by retinoic acid in development of the ascidian *Halocynthia roretzi*. *Development* **121,** 3197–3205 (1995).
447. Sasakura, Y. *et al.* Retinoic acid-driven *Hox1* is required in the epidermis for forming the otic/atrial placodes during ascidian metamorphosis. *Development* **139,** 2156–2160 (2012).
448. Hinman, V. F. & Degnan, B. M. Retinoic acid disrupts anterior ectodermal and endodermal development in ascidian larvae and postlarvae. *Dev. Genes Evol.* **208,** 336–345 (1998).
449. Hinman, V. F. & Degnan, B. M. Retinoic acid perturbs *Otx* gene expression in the ascidian pharynx. *Dev. Genes Evol.* **210,** 129–139 (2000).
450. Ogasawara, M., Wada, H., Peters, H. & Satoh, N. Developmental expression of *Pax1/9* genes in urochordate and hemichordate gills: insight into function and evolution of the pharyngeal epithelium. *Development* **126,** 2539–2550 (1999).
451. Albalat, R. & Cañestro, C. Identification of Aldh1a, Cyp26 and RAR orthologs in protostomes pushes back the retinoic acid genetic machinery in evolutionary time to the bilaterian ancestor. *Chem. Biol. Interact.* **178,** 188–196 (2009).
452. Farrar, N. R., Dmetrichuk, J. M., Carlone, R. L. & Spencer, G. E. A novel, nongenomic mechanism underlies retinoic acid-induced growth cone turning. *J. Neurosci.* **29,** 14136–14142 (2009).
453. Sukiban, J., Bräunig, P., Mey, J. & Bui-Göbbels, K. Retinoic acid as a survival factor in neuronal development of the grasshopper, Locusta migratoria. *Cell Tissue Res.* **358,** 303–312 (2014).
454. Dmetrichuk, J. M., Carlone, R. L. & Spencer, G. E. Retinoic acid induces neurite outgrowth and growth cone turning in invertebrate neurons. *Dev. Biol.* **294,** 39–49 (2006).
455. Dmetrichuk, J. M., Carlone, R. L., Jones, T. R. B., Vesprini, N. D. & Spencer, G. E. Detection of endogenous retinoids in the molluscan CNS and characterization of the trophic and tropic actions of 9-*cis* retinoic acid on isolated neurons. *J. Neurosci.* **28,** 13014–13024 (2008).
456. Carter, C. J., Farrar, N., Carlone, R. L. & Spencer, G. E. Developmental expression of a molluscan RXR and evidence for its novel, nongenomic role in growth cone guidance. *Dev. Biol.* **343,** 124–137 (2010).
457. Urushitani, H. *et al.* Cloning and characterization of the retinoic acid receptor-like protein in the rock shell, *Thais clavigera*. *Aquat. Toxicol.* **142–143,** 403–413 (2013).
458. Créton, R., Zwaan, G. & Dohmen, R. Specific developmental defects in molluscs after treatment with retinoic acid during gastrulation. *Dev. Growth Differ.* **35,** 357–364 (1993).
459. Halme, A., Cheng, M. & Hariharan, I. K. Retinoids regulate a developmental checkpoint for tissue regeneration in *Drosophila*. *Curr. Biol.* **20,** 458–463 (2010).
460. Hopkins, P. M. Limb regeneration in the fiddler crab, *Uca pugilator*: hormonal and growth factor control. *Am. Zool.* **41,** 389–398 (2001).
461. Bui-Göbbels, K., Quintela, R. M., Bräunig, P. & Mey, J. Is retinoic acid a signal for nerve regeneration in insects? *Neural Regen. Res.* **10,** 901–903 (2015).
462. Müller, W. A. Retinoids and pattern formation in a hydroid. *Development* **81,** 253–271 (1984).
463. Estephane, D. & Anctil, M. Retinoic acid and nitric oxide promote cell proliferation and differentially induce neuronal differentiation *in vitro* in the cnidarian *Renilla koellikeri*. *Dev. Neurobiol.* **70,** 842–852 (2010).
464. Wiens, M., Batel, R., Korzhev, M. & Müller, W. E. G. Retinoid X receptor and retinoic acid response in the marine sponge *Suberites domuncula*. *J. Exp. Biol.* **206,** 3261–3271 (2003).
465. Dalfó, D., Albalat, R., Molotkov, A., Duester, G. & Gonzàlez-Duarte, R. Retinoic acid synthesis in the prevertebrate amphioxus involves retinol oxidation. *Dev. Genes Evol.* **212,** 388–393 (2002).
466. Onai, T. *et al.* Retinoic acid and Wnt/β-catenin have complementary roles in anterior/posterior patterning embryos of the basal chordate amphioxus. *Dev. Biol.* **332,** 223–233 (2009).

467. Koop, D. *et al.* Roles of retinoic acid and Tbx1/10 in pharyngeal segmentation: amphioxus and the ancestral chordate condition. *EvoDevo* **5,** 36 (2014).
468. Bertrand, S. *et al.* Evolution of the role of RA and FGF signals in the control of somitogenesis in chordates. *PLOS ONE* **10,** e0136587 (2015).
469. Hirsinger, E. *et al.* Expression of fluorescent proteins in *Branchiostoma lanceolatum* by mRNA injection into unfertilized oocytes. *J. Vis. Exp.* (2015). doi:10.3791/52042
470. Weber, T. & Köster, R. Genetic tools for multicolor imaging in zebrafish larvae. *Methods* **62,** 279–291 (2013).
471. Weil, T. T., Parton, R. M. & Davis, I. Making the message clear: visualizing mRNA localization. *Trends Cell Biol.* **20,** 380–390 (2010).
472. Zhang, Y. & Yu, L.-C. Microinjection as a tool of mechanical delivery. *Curr. Opin. Biotechnol.* **19,** 506–510 (2008).
473. Stepicheva, N. & Song, J. High throughput microinjections of sea urchin zygotes. *J. Vis. Exp. JoVE J. Vis. Exp. JoVE* e50841–e50841 (2014). doi:10.3791/50841, 10.3791/50841
474. Holland, L. Z. Evolution of new characters after whole genome duplications: insights from amphioxus. *Semin. Cell Dev. Biol.* **24,** 101–109 (2013).
475. Holland, L. Z. & Onai, T. Analyses of gene function in amphioxus embryos by microinjection of mRNAs and morpholino oligonucleotides. *Methods Mol. Biol. Clifton NJ* **770,** 423–438 (2011).
476. Liu, X., Li, G., Feng, J., Yang, X. & Wang, Y.-Q. An efficient microinjection method for unfertilized eggs of Asian amphioxus *Branchiostoma belcheri*. *Dev. Genes Evol.* **223,** 269–278 (2013).
477. Harvey, K. J., Lukovic, D. & Ucker, D. S. Membrane-targeted green fluorescent protein reliably and uniquely marks cells through apoptotic death. *Cytometry* **43,** 273–278 (2001).
478. Maruyama, J., Nakajima, H. & Kitamoto, K. Visualization of nuclei in *Aspergillus oryzae* with EGFP and analysis of the number of nuclei in each conidium by FACS. *Biosci. Biotechnol. Biochem.* **65,** 1504–1510 (2001).
479. Rupp, R. A., Snider, L. & Weintraub, H. *Xenopus* embryos regulate the nuclear localization of *XMyoD*. *Genes Dev.* **8,** 1311–1323 (1994).
480. Turner, D. L. & Weintraub, H. Expression of achaete-scute homolog 3 in *Xenopus* embryos converts ectodermal cells to a neural fate. *Genes Dev.* **8,** 1434–1447 (1994).
481. Nakagawa, S., Niimura, Y., Gojobori, T., Tanaka, H. & Miura, K. Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. *Nucleic Acids Res.* **36,** 861–871 (2008).
482. Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28,** 292–292 (2000).
483. Deheyn, D. D. *et al.* Endogenous green fluorescent protein (GFP) in amphioxus. *Biol. Bull.* **213,** 95–100 (2007).
484. Yu, J.-K., Holland, N. D. & Holland, L. Z. Tissue-specific expression of *FoxD* reporter constructs in amphioxus embryos. *Dev. Biol.* **274,** 452–461 (2004).
485. Beaster-Jones, L., Schubert, M. & Holland, L. Z. *Cis*-regulation of the amphioxus *engrailed* gene: insights into evolution of a muscle-specific enhancer. *Mech. Dev.* **124,** 532–542 (2007).
486. Urasaki, A., Mito, T., Noji, S., Ueda, R. & Kawakami, K. Transposition of the vertebrate *Tol2* transposable element in *Drosophila melanogaster*. *Gene* **425,** 64–68 (2008).
487. Li, G. *et al.* Mutagenesis at specific genomic loci of amphioxus *Branchiostoma belcheri* using TALEN method. *J. Genet. Genomics Yi Chuan Xue Bao* **41,** 215–219 (2014).
488. Dobbs-McAuliffe, B., Zhao, Q. & Linney, E. Feedback mechanisms regulate retinoic acid production and degradation in the zebrafish embryo. *Mech. Dev.* **121,** 339–350 (2004).
489. Lee, L. M. Y. *et al.* A paradoxical teratogenic mechanism for retinoic acid. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 13668–13673 (2012).
490. Schilling, T. F., Nie, Q. & Lander, A. D. Dynamics and precision in retinoic acid morphogen gradients. *Curr. Opin. Genet. Dev.* **22,** 562–569 (2012).
491. D'Aniello, E., Rydeen, A. B., Anderson, J. L., Mandal, A. & Waxman, J. S. Depletion of retinoic acid receptors initiates a novel positive feedback mechanism that promotes teratogenic increases in retinoic acid. *PLOS Genet.* **9,** e1003689 (2013).
492. Rydeen, A. *et al.* Excessive feedback of Cyp26a1 promotes cell non-autonomous loss of retinoic acid signaling. *Dev. Biol.* **405,** 47–55 (2015).
493. Balmer, J. E. & Blomhoff, R. Gene expression regulation by retinoic acid. *J. Lipid Res.* **43,** 1773–1808 (2002).
494. Umesono, K., Murakami, K. K., Thompson, C. C. & Evans, R. M. Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D_3 receptors. *Cell* **65,** 1255–1266 (1991).

495. Topletz, A. R. *et al.* Induction of CYP26A1 by metabolites of retinoic acid: evidence that CYP26A1 is an important enzyme in the elimination of active retinoids. *Mol. Pharmacol.* **87,** 430–441 (2015).
496. White, R. J. & Schilling, T. F. How degrading: Cyp26s in hindbrain development. *Dev. Dyn.* **237,** 2775–2790 (2008).
497. Kumar, S. & Duester, G. Retinoic acid controls body axis extension by directly repressing *Fgf8* transcription. *Development* **141,** 2972–2977 (2014).
498. Wahl, M. B., Deng, C., Lewandoski, M. & Pourquié, O. FGF signaling acts upstream of the NOTCH and WNT signaling pathways to control segmentation clock oscillations in mouse somitogenesis. *Development* **134,** 4033–4041 (2007).
499. Bothe, I., Tenin, G., Oseni, A. & Dietrich, S. Dynamic control of head mesoderm patterning. *Development* **138,** 2807–2821 (2011).
500. Loudig, O. *et al.* Cytochrome P450RAI(CYP26) promoter: a distinct composite retinoic acid response element underlies the complex regulation of retinoic acid metabolism. *Mol. Endocrinol.* **14,** 1483–1497 (2000).
501. Loudig, O., Maclean, G. A., Dore, N. L., Luu, L. & Petkovich, M. Transcriptional co-operativity between distant retinoic acid response elements in regulation of *Cyp26A1* inducibility. *Biochem. J.* **392,** 241 (2005).
502. Hurles, M. Gene duplication: the genomic trade in spare parts. *PLOS Biol* **2,** e206 (2004).
503. Bertrand, S. *et al.* Unexpected novel relational links uncovered by extensive developmental profiling of nuclear receptor expression. *PLOS Genet* **3,** e188 (2007).
504. Shimeld, S. M. & Holland, N. D. Amphioxus molecular biology: insights into vertebrate evolution and developmental mechanisms. *Can. J. Zool.* **83,** 90–100 (2005).
505. Stoppie, P. *et al.* R115866 inhibits all-trans-retinoic acid metabolism and exerts retinoidal effects in rodents. *J. Pharmacol. Exp. Ther.* **293,** 304–312 (2000).
506. Thatcher, J. E. *et al.* Substrate specificity and ligand interactions of CYP26A1, the human liver retinoic acid hydroxylase. *Mol. Pharmacol.* **80,** 228–239 (2011).
507. Flood, P. R. Ciliary rootlet-fibres as tail fin-rays in larval amphioxus (*Branchiostoma lanceolatum*, pallas). *J. Ultrastruct. Res.* **51,** 218–225 (1975).
508. Mansfield, J. H. & Holland, N. D. Amphioxus tails: source and fate of larval fin rays and the metamorphic transition from an ectodermal to a predominantly mesodermal tail. *Acta Zool.* **96,** 117–125 (2013).
509. Hu, P. *et al.* Retinoid regulation of the zebrafish cyp26a1 promoter. *Dev. Dyn.* **237,** 3798–3808 (2008).
510. Zhang, Y., Zolfaghari, R. & Ross, A. C. Multiple retinoic acid response elements cooperate to enhance the inducibility of CYP26A1 gene expression in liver. *Gene* **464,** 32–43 (2010).
511. Pascual-Anaya, J. *et al.* Broken colinearity of the amphioxus *Hox* cluster. *EvoDevo* **3,** 28 (2012).
512. White, R. J., Nie, Q., Lander, A. D. & Schilling, T. F. Complex regulation of *cyp26a1* creates a robust retinoic acid gradient in the zebrafish embryo. *PLOS Biol* **5,** e304 (2007).
513. Shimozono, S., Iimura, T., Kitaguchi, T., Higashijima, S. & Miyawaki, A. Visualization of an endogenous retinoic acid gradient across embryonic development. *Nature* **496,** 363–366 (2013).
514. Schubert, M., Holland, L. Z., Stokes, M. D. & Holland, N. D. Three amphioxus *Wnt* genes (*AmphiWnt3*, *AmphiWnt5*, and *AmphiWnt6*) associated with the tail bud: the evolution of somitogenesis in chordates. *Dev. Biol.* **240,** 262–273 (2001).
515. Pennimpede, T. *et al.* The role of CYP26 enzymes in defining appropriate retinoic acid exposure during embryogenesis. *Birt. Defects Res. A. Clin. Mol. Teratol.* **88,** 883–894 (2010).
516. Ahn, Y., Mullan, H. E. & Krumlauf, R. Long-range regulation by shared retinoic acid response elements modulates dynamic expression of posterior *Hoxb* genes in CNS development. *Dev. Biol.* **388,** 134–144 (2014).
517. Helsen, C. & Claessens, F. Looking at nuclear receptors from a new angle. *Mol. Cell. Endocrinol.* **382,** 97–106 (2014).
518. Rodríguez-Marí, A. *et al.* Retinoic acid metabolic genes, meiosis, and gonadal sex differentiation in zebrafish. *PLOS ONE* **8,** e73951 (2013).
519. Dehal, P. & Boore, J. L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLOS Biol* **3,** e314 (2005).
520. Amores, A. *et al.* Zebrafish *hox* clusters and vertebrate genome evolution. *Science* **282,** 1711–1714 (1998).
521. Barash, M. S. Mass extinction of the marine biota at the Ordovician-Silurian transition due to environmental changes. *Oceanology* **54,** 780–787 (2014).
522. Berner, R. A., VandenBrooks, J. M. & Ward, P. D. Oxygen and evolution. *Science* **316,** 557–558 (2007).
523. Kremer, B. & Kaźmierczak, J. Cyanobacterial mats from Silurian black radiolarian cherts: phototrophic life at the edge of darkness? *J. Sediment. Res.* **75,** 897–906 (2005).
524. Castle, J. W. & Rodgers, J. H. Hypothesis for the role of toxin-producing algae in Phanerozoic mass extinctions based on evidence from the geologic record and modern environments. *Environ. Geosci.* **16,** 1–23 (2009).

525. Dudley, R. Atmospheric oxygen, giant Paleozoic insects and the evolution of aerial locomotor performance. *J. Exp. Biol.* **201,** 1043–1050 (1998).
526. Alroy, J. The shifting balance of diversity among major marine animal groups. *Science* **329,** 1191–1194 (2010).
527. Jia, C. *et al.* Microbial response to limited nutrients in shallow water immediately after the end-Permian mass extinction. *Geobiology* **10,** 60–71 (2012).
528. Berman-Frank, I., Lundgren, P. & Falkowski, P. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Res. Microbiol.* **154,** 157–164 (2003).
529. Kelman, D., Ben-Amotz, A. & Berman-Frank, I. Carotenoids provide the major antioxidant defence in the globally significant N_2-fixing marine cyanobacterium *Trichodesmium*. *Environ. Microbiol.* **11,** 1897–1908 (2009).
530. Wu, X., Jiang, J., Wan, Y., Giesy, J. P. & Hu, J. Cyanobacteria blooms produce teratogenic retinoic acids. *Proc. Natl. Acad. Sci.* **109,** 9477–9482 (2012).
531. Jonas, A. *et al.* Retinoid-like activity and teratogenic effects of cyanobacterial exudates. *Aquat. Toxicol. Amst. Neth.* **155,** 283–290 (2014).
532. Yu, J. K. S. & Holland, L. Z. Extraction of RNA from amphioxus embryos or adult amphioxus tissue. *Cold Spring Harb. Protoc.* **2009,** pdb.prot5288 (2009).
533. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27,** 221–224 (2010).
534. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27,** 1164–1165 (2011).
535. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9,** 772–772 (2012).
536. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).
537. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* **17,** 754–755 (2001).
538. Rambaut, A., Suchard, M. A., Xie, D. & Drummond, A. J. *Tracer v1.6*. (2014).
539. Rambaut, A. FigTree. (2012). Available at: http://tree.bio.ed.ac.uk/software/figtree/. (Accessed: 7th October 2016)
540. Bouckaert, R. *et al.* BEAST 2: a software platform for bayesian evolutionary analysis. *PLOS Comput Biol* **10,** e1003537 (2014).
541. Benton, M. J., Donoghue, P. C. J. & Asher, R. J. in *The timetree of life* 35–86 (Oxford University Press, 2009).
542. dos Reis, M. *et al.* Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr. Biol.* **25,** 2939–2950 (2015).
543. Bailey, T. L. *et al.* MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* **37,** W202–W208 (2009).
544. Rupp, R. A., Snider, L. & Weintraub, H. *Xenopus* embryos regulate the nuclear localization of XMyoD. *Genes Dev.* **8,** 1311–1323 (1994).
545. Yu, J. K. S. & Holland, L. Z. Amphioxus whole-mount in situ hybridization. *Cold Spring Harb. Protoc.* **2009,** pdb.prot5286 (2009).
546. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).
547. Berry, D. C. & Noy, N. Is PPARβ/δ a retinoid receptor? *PPAR Res.* **2007,** e73256 (2008).
548. Wolf, G. Retinoic acid activation of peroxisome proliferation-activated receptor δ represses obesity and insulin resistance. *Nutr. Rev.* **68,** 67–70 (2010).
549. Huang, P., Chandra, V. & Rastinejad, F. Retinoic acid actions through mammalian nuclear receptors. *Chem. Rev.* **114,** 233–254 (2014).
550. Rochette-Egly, C. & Germain, P. Dynamic and combinatorial control of gene expression by nuclear retinoic acid receptors (RARs). *Nucl. Recept. Signal.* **7,** e005 (2009).
551. Chang, Y. S. *et al.* 9-cis retinoic acid induces insulin-like growth factor binding protein-3 through DR-8 retinoic acid responsive elements. *Cancer Biol. Ther.* **5,** 586–592 (2006).
552. Lalevée, S. *et al.* Genome-wide *in silico* identification of new conserved and functional retinoic acid receptor response elements (direct repeats separated by 5 bp). *J. Biol. Chem.* **286,** 33322–33334 (2011).
553. Perissi, V., Jepsen, K., Glass, C. K. & Rosenfeld, M. G. Deconstructing repression: evolving models of co-repressor action. *Nat. Rev. Genet.* **11,** 109–123 (2010).
554. Kashyap, V. & Gudas, L. J. Epigenetic regulatory mechanisms distinguish retinoic acid-mediated transcriptional responses in stem cells and fibroblasts. *J. Biol. Chem.* **285,** 14534–14548 (2010).

555. Mahony, S. *et al.* Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biol.* **12,** R2 (2011).
556. Paschaki, M. *et al.* Transcriptomic analysis of murine embryos lacking endogenous retinoic acid signaling. *PLOS ONE* **8,** e62274 (2013).
557. Jolma, A. *et al.* DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527,** 384–388 (2015).
558. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15,** 272–286 (2014).
559. Moorman, C. *et al.* Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **103,** 12027–12032 (2006).
560. Li, X. *et al.* Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol.* **6,** e27 (2008).
561. Ouyang, Z., Zhou, Q. & Wong, W. H. ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. *Proc. Natl. Acad. Sci.* **106,** 21521–21526 (2009).
562. Hurwitz, J. L. *et al.* Hotspots for vitamin-steroid-thyroid hormone response elements within switch regions of immunoglobulin heavy chain loci predict a direct influence of vitamins and hormones on B cell class switch recombination. *Viral Immunol.* **29,** 132–136 (2016).
563. Takitani, K. *et al.* Expression of retinoic acid receptor–target genes during retinoic acid therapy for acute promyelocytic leukemia. *Leukemia* **17,** 646–648 (2003).
564. Savory, J. G. A., Edey, C., Hess, B., Mears, A. J. & Lohnes, D. Identification of novel retinoic acid target genes. *Dev. Biol.* **395,** 199–208 (2014).
565. Niewiadomska-Cimicka, A. *et al.* Genome-wide analysis of RARβ transcriptional targets in mouse striatum links retinoic acid signaling with Huntington's disease and other neurodegenerative disorders. *Mol. Neurobiol.* (2016). doi:10.1007/s12035-016-0010-4
566. Moutier, E. *et al.* Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology. *J. Biol. Chem.* **287,** 26328–26341 (2012).
567. Hua, S., Kittler, R. & White, K. P. Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. *Cell* **137,** 1259–1271 (2009).
568. Farnham, P. J. Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.* **10,** 605–616 (2009).
569. Parvin, M. S. *et al.* Autoregulatory loop and retinoic acid repression regulate *pou2/pou5f1* gene expression in the zebrafish embryonic brain. *Dev. Dyn.* **237,** 1373–1388 (2008).
570. Studer, M., Pöpperl, H., Marshall, H., Kuroiwa, A. & Krumlauf, R. Role of a conserved retinoic acid response element in rhombomere restriction of Hoxb-1. *Science* **265,** 1728–1732 (1994).
571. Bastian, H. & Gruss, P. A murine *even-skipped* homologue, <i>Evx 1</>, is expressed during early embryogenesis and neurogenesis in a biphasic manner. *EMBO J.* **9,** 1839–1852 (1990).
572. Yokouchi, Y., Ohsugi, K., Sasaki, H. & Kuroiwa, A. Chicken homeobox gene *Msx-1*: structure, expression in limb buds and effect of retinoic acid. *Dev. Camb. Engl.* **113,** 431–444 (1991).
573. Turner, E. E., Jenne, K. J. & Rosenfeld, M. G. Brn-3.2: A Brn-3-related transcription factor with distinctive central nervous system expression and regulation by retinoic acid. *Neuron* **12,** 205–218 (1994).
574. Mistri, T. K. *et al.* Selective influence of Sox2 on POU transcription factor binding in embryonic and neural stem cells. *EMBO Rep.* **16,** 1177–1191 (2015).
575. Chang, Y. K. *et al.* Quantitative profiling of selective Sox/POU pairing on hundreds of sequences in parallel by Coop-seq. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw1198
576. Liu, X., Huang, X. & Sigmund, C. D. Identification of a nuclear orphan receptor (Ear2) as a negative regulator of renin gene transcription. *Circ. Res.* **92,** 1033–1040 (2003).
577. Weatherford, E. T., Liu, X. & Sigmund, C. D. Regulation of renin expression by the orphan nuclear receptors Nr2f2 and Nr2f6. *Am. J. Physiol. - Ren. Physiol.* **302,** F1025–F1033 (2012).
578. Massari, M. E. & Murre, C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol.* **20,** 429–440 (2000).
579. Fritzenwanker, J. H., Gerhart, J., Freeman, R. M. & Lowe, C. J. The Fox/Forkhead transcription factor family of the hemichordate *Saccoglossus kowalevskii*. *EvoDevo* **5,** 17 (2014).
580. Hannenhalli, S. & Kaestner, K. H. The evolution of Fox genes and their role in development and disease. *Nat. Rev. Genet.* **10,** 233–240 (2009).
581. Lam, E. W.-F., Brosens, J. J., Gomes, A. R. & Koo, C.-Y. Forkhead box proteins: tuning forks for transcriptional harmony. *Nat. Rev. Cancer* **13,** 482–495 (2013).
582. Lecroisey, C., Laudet, V. & Schubert, M. The cephalochordate amphioxus: a key to reveal the secrets of nuclear receptor evolution. *Brief. Funct. Genomics* **11,** 156–166 (2012).
583. Yu, J.-K. *et al.* The Fox genes of *Branchiostoma floridae*. *Dev. Genes Evol.* **218,** 629–638 (2008).

584. Aldea, D., Leon, A., Bertrand, S. & Escriva, H. Expression of Fox genes in the cephalochordate *Branchiostoma lanceolatum*. *Evol. Popul. Genet.* 80 (2015). doi:10.3389/fevo.2015.00080
585. Kaestner, K. H. The FoxA factors in organogenesis and differentiation. *Curr. Opin. Genet. Dev.* **20,** 527–532 (2010).
586. Wang, R. *et al.* The ratio of FoxA1 to FoxA2 in lung adenocarcinoma is regulated by LncRNA HOTAIR and chromatin remodeling factor LSH. *Sci. Rep.* **5,** 17826 (2015).
587. Santisteban, P., Recacha, P., Metzger, D. E. & Zaret, K. S. Dynamic expression of *Groucho*-related genes *Grg1* and *Grg3* in foregut endoderm and antagonism of differentiation. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **239,** 980–986 (2010).
588. Martens, J. H. A., Rao, N. A. S. & Stunnenberg, H. G. Genome-wide interplay of nuclear receptors with the epigenome. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1812,** 818–823 (2011).
589. Carroll, J. S. *et al.* Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122,** 33–43 (2005).
590. Lupien, M. *et al.* FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132,** 958–970 (2008).
591. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10,** 1213–1218 (2013).
592. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109,** 21.29.1-21.29.9 (2015).
593. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
594. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 (2014).
595. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** (2014).
596. Alexa, A. & Rahnenfuhrer, J. *topGO: enrichment analysis for gene ontology*. (2016).
597. Cline, M. S. *et al.* Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* **2,** 2366–2382 (2007).
598. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
599. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
600. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–589 (2010).
601. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
602. Thomas-Chollier, M. *et al.* A complete workflow for the analysis of full-size ChIP-seq (and similar) data sets using peak-motifs. *Nat. Protoc.* **7,** 1551–1568 (2012).
603. Thomas-Chollier, M. *et al.* RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res.* **40,** e31–e31 (2012).

Table of illustrations

Appendixes

Appendix 1: Supporting material of "R1: Expression of fluorescent proteins in *Branchiostoma lanceolatum* by mRNA injection into unfertilized oocytes"

Appendix 2: Supporting material of "R2: Lineage-specific duplication of amphioxus retinoic acid degrading enzymes (CYP26) resulted in sub-functionalization of patterning and homeostatic roles"

Appendix 3: Supporting material of "R3:"

Appendix 4: Published book chapter: "Retinoic acid: metabolism, developmental functions, and evolution."

Appendix 5: Published article: "Evolution of bilaterian central nervous systems: a single origin?"

Appendix 6: Published article: "Roles of retinoic acid and Tbx1/10 in pharyngeal segmentation: amphioxus and the ancestral chordate condition."

Appendix 7: Published article: "Expression of fluorescent proteins in *Branchiostoma lanceolatum* by mRNA injection into unfertilized oocytes."

Appendix 8: Published article: "The human Mixed Lineage Leukemia 5 (MLL5), a sequentially and structurally divergent SET domain-containing protein with no intrinsic catalytic activity."

Appendix 9: *Curriculum vitae*

Appendix 1

Name of Material/ Equipment	Company	Catalog Number	Comments/Description
Consumables			
35 mm Petri dishes	Falcon	353001	culture-treated
Filtration unit (Stericup 1L)	Fisher	W21719	0.22 micron filtration
Spin-X tubes	Costar	8160	0.22 micron filtration tube
Needle storage jar for 1.2 mm diameter capillaries	WPI	E212	
Pasteur Pipettes			230 mm long
Aspiration tube	Dutscher	75056	
Capillaries for injection needles	Sutter	BF 120-94-10	Borosilicate glass with filament, OD 1.20 mm, ID 0.94 mm, length 10 mm
Reagents			
Low-melting agarose	SIGMA	A9414	
Phenol Red	SIGMA	114537	
Glycerol	SIGMA	G2025	
Poly-L-Lysine hydrobromide	SIGMA	P9155	
H2O Dnase, Rnase-free	Gibco	10977-035	
mMessage mMachine SP6 Transcription kit	Ambion	AM1340	mRNA synthesis kit
Phenol pH8	SIGMA	P4557	
24:1 chloroform:isoamylic alcohol	SIGMA	C0549	
5:1 phenol pH4.7:chloroform	SIGMA	P1944	
Reef Crystal salts (200 kg)	Europrix		Commercial salts
Equipment			
Fluorescent dissecting scope with 200x magnification	Leica	MZ16F	25x oculars, DSR and GFP2 filters
Micromanipulator	Marzhauzer	M-33	
Injector	Picospritzer	model II or III	
Needle puller	Sutter	P97	heating-filament needle puller

A

eGFP:CAAX box
eGFP: from pEGFP-1 vector (GenBank U55761)
CAAX box: membrane localization signal from human HRAS (GenBank BT019421)
Mutated sites are indicated in bold and underlined.

```
1     CGCCATTCTGCCTGGGGACGTCGGAGCAAGCTTGATTTAGGTGACACTAT
51    AGAATACAAGCTACTTGTTCTTTTTGCAACCGTCAACATGGTGAGCAAGG
101   GCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC
151   GACGTGAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGC
201   CACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC
251   CCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGC
301   TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGC
351   CATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACG
401   GCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTG
451   AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCT
501   GGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGG
551   CCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAAC
601   ATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCC
651   CATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCC
701   AGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTG
751   CTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTA
801   CAAGGGAGGAGGAAGGTCTAAGCTGAACCCTCCTGATGAGAGTGGCCCCG
851   GCTGCATGAGCTGCAAGTGTGTGCTCTCCTGATCTAGAACTATAGTGAGT
901   CGTATTACGTAGATCCAGACATGATAAGATACATTGATGAGTTTGGACAA
951   ACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGA
1001  TGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACA
1051  ACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAG
1101  GTGTGGGAGGTTTTTTAATTCGCGGCCGCGGCGCCAATGCATTGGGCCCG
```

B

H2B:eGFP
H2B: zebrafish histone 2B (hist2h2l) exon (Ensembl ENSDARG00000068996.2)
eGFP: from pEGFP-1 vector (GenBank U55761)
Mutated sites are indicated in bold and underlined.

```
1     CGCCATTCTGCCTGGGGACGTCGGAGCAAGCTTGATTTAGGTGACACTAT
51    AGAATACAAGCTACTTGTTCTTTTTGCAGGAACCGTCAACATGCCCGAAC
101   CTGCGAAGTCAGCGCCCGCTCCCAAAAAAGGCTCTAAAAAAGCTGTCGCC
151   AAGACCCAGAAGAAGGGGGATAAGAAAAGGCGTAAGACCAGGAAAGAGAG
201   TTACGCCATTTACGTGTACAAAGTGCTGAAACAAGTCCACCCGGACACTG
251   GCATCTCCTCAAAGGCGATGGGCATTATGAACTCATTTGTGAACGACATC
301   TTCGAGCGCATCGCCGGAGAAGCGTCGCGCCTGGCGCATTACAACAAGCG
351   CTCCACTATCACATCCCGGGAGATCCAGACGGCCGTGCGCCTGCTCCTGC
401   CCGGAGAACTGGCCAAACACGCTGTGTCTGAGGGCACAAAGGCCGTGACC
451   AAGTACACCAGCTCCAAGGTGCCGCGGGCCCGGGATCCACCGGTCGCCAC
501   CATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
551   TCGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGTCCGGCGAG
601   GGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGG
651   CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCG
701   TGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTC
751   AAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
801   GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACA
851   CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGC
901   AACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTA
951   TATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCC
1001  GCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAG
1051  AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCT
1101  GAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACA
1151  TGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
1201  GAGCTGTACAAGTAAAGCGGCCGCGACTCTAGAACTATAGTGAGTCGTAT
1251  TACGTAGATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCAC
1301  AACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTA
1351  TTGCTTTATTTGTAACCATTTATAAGCTGCAATAAACAAGTTAACAACAA
1401  CAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTT
1451  TTTAATTCGCGGCCGCGGCGCCAATGCATTGGGCCCGGTACCCAGCTTTT
```

C

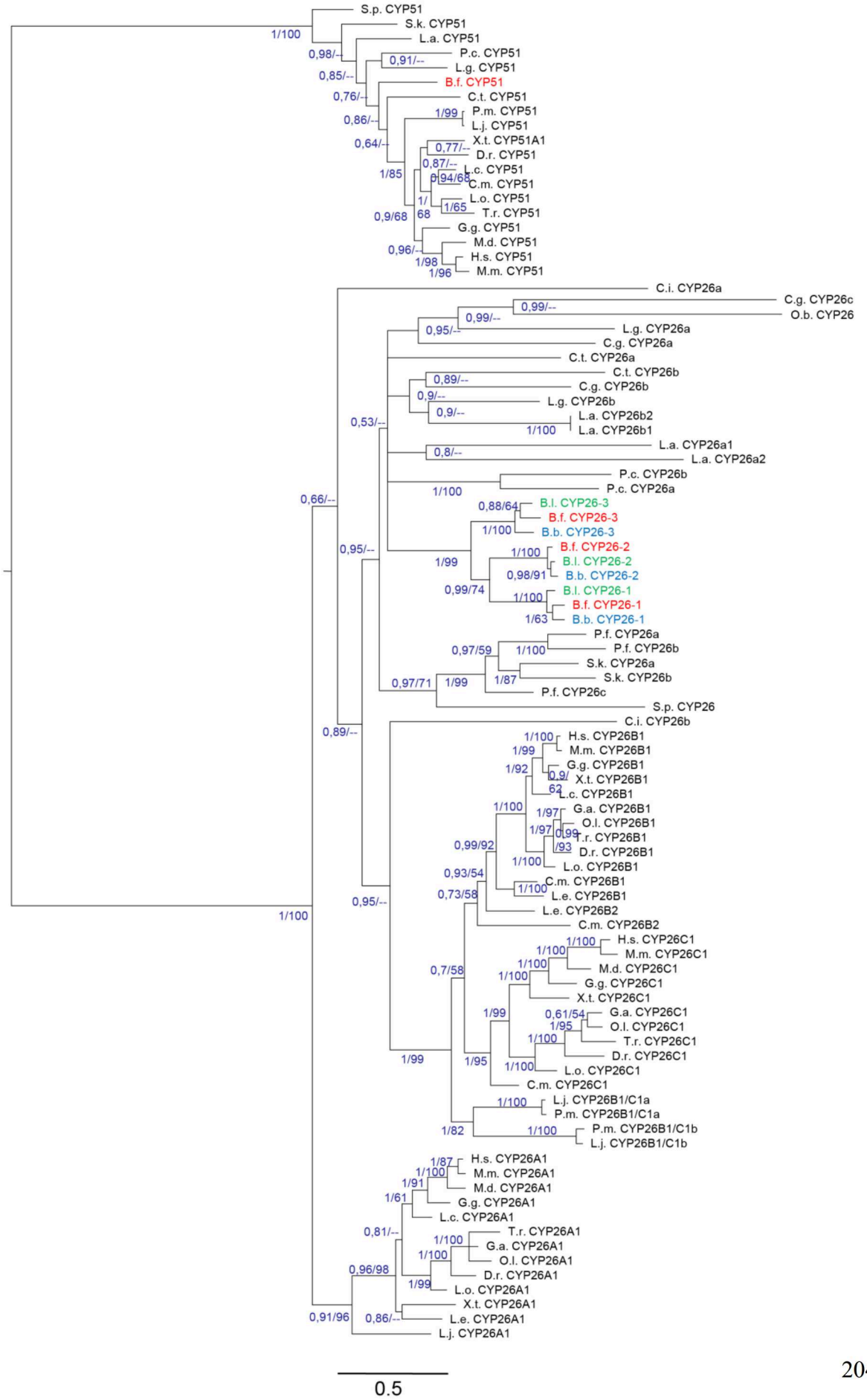
H2B:mCherry
H2B: zebrafish histone 2B (hist2h2l) exon (Ensembl ENSDARG00000068996.2)
mCherry: from pRSET-B-mCherry plasmid (the Forchheimer Center plasmid collection)
Mutated sites are indicated in bold and underlined.

```
1     CGCCATTCTGCCTGGGGACGTCGGAGCAAGCTTGATTTAGGTGACACTAT
51    AGAATACAAGCTACTTGTTCTTTTTGCAGGAACCGTCAACATGCCCGAAC
101   CTGCGAAGTCAGCGCCCGCTCCCAAAAAAGGCTCTAAAAAAGCTGTCGCC
151   AAGACCCAGAAGAAGGGGGGATAAGAAAAGGCGTAAGACCAGGAAAGAGAG
201   TTACGCCATTTACGTGTACAAAGTGCTGAAACAAGTCCACCCGGACACTG
251   GCATCTCCTCAAAGGCGATGGGCATTATGAACTCATTTGTGAACGACATC
301   TTCGAGCGCATCGCCGGAGAAGCGTCGCGCCTGGCGCATTACAACAAGCG
351   CTCCACTATCACATCCCGGGAGATCCAGACGGCCGTGCGCCTGCTCCTGC
401   CCGGAGAACTGGCCAAACACGCTGTGTCTGAGGGCACAAAGGCCGTGACC
451   AAGTACACCAGCTCCAAGGTGCCGCGGGCCCGGGATCCCGCCACCATGGT
501   GAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCT
551   TCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAG
601   GGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAA
651   GGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTC
701   AGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC
751   GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGAT
801   GAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGC
851   AGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCC
901   TCCGACGGCCCCGTGATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTC
951   CGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGA
1001  GGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACC
1051  TACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACAT
1101  CAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGT
1151  ACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTAC
1201  AAGTAAGAATTCAAGGCCTCTCGAGCCTCTAGAACTATAGTGAGTCGTAT
1251  TACGTAGATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCAC
1301  AACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTA
1351  TTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAAC
1401  AATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTT
1451  TTAATTCGCGGCCGCGGCGCCAATGCATTGGGCCCGGTACCCAGCTTTTG
```

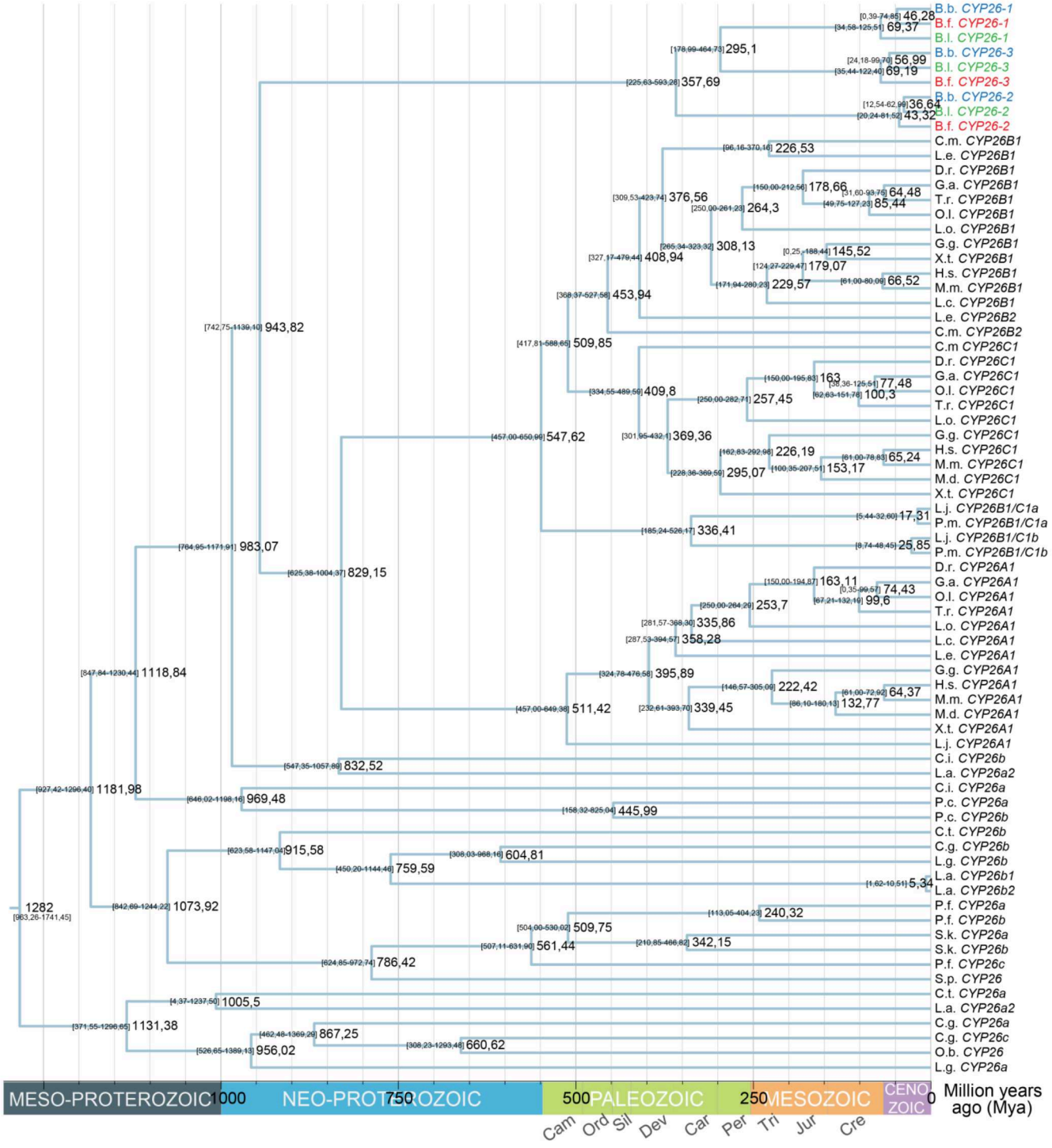
Appendix 2

Species name	CYP26			CYP51	Database information
Octopus bimaculoides	Ocbimv22014306m				Ensembl Metazoa
Strongylocentrotus purpuratus	SPU_020876-tr			NM_001001906	Ensembl Metazoa + NCBI
	CYP26a	***CYP26b***			
Capitella teleta	CapteT212322	CapteT150007		CapteT173561	Ensembl Metazoa
Ciona intestinalis	ENSCING00000004832	ENSCING00000004829			Ensembl
Lingula anatina	g12897 (**1**) & g3653 (**2**)	g32063 (**1**) & g9815 (**2**)		g19581	http://marinegenomics.oist.jp/lingula/viewer/info?project_id=47
Lottia gigantea	LotgiT189041	LotgiT111029		LotgiT108695	Ensembl Metazoa
Priapulus caudatus	XM_014805928	XM_014806206		XM_014805882	NCBI
Saccoglossus kowalevskii	NM_001168052	XM_002734838		XM_002741540	NCBI
	CYP26a	***CYP26b***	***CYP26c***		
Crassostrea gigas	EKC33614	XM_011438536	EKC37998		Ensembl Metazoa + NCBI
Ptychodera flava	pfl_40v0_9_20150316_1g13469	pfl_40v0_9_20150316_1g3888-1g3889	pfl_40v0_9_20150316_1g33013-1g3891		http://marinegenomics.oist.jp/acornworm/viewer/info?project_id=33
	CYP26-1	***CYP26-2***	***CYP26-3***		
Branchiostoma belcheri	200590F	200570F	200560F		http://mosas.sysu.edu.cn/genome/B.belcheri_v18h27.r3_ref_cds
Branchiostoma floridae	87629	87630	124944	115377	JGI
Branchiostoma lanceolatum	KX118106	KX118108	KX118107		NCBI
	CYP26A1	***CYP26B1***	***CYP26C1***		
Callorhinchus milii		SINCAMT00000006531 (**1**) & SINCAMT00000015587 (**2**)	SINCAMT00000020684	SINCAMT00000007547	Ensembl
Danio rerio	ENSDART00000041728	ENSDART00000110347	ENSDART00000077809	ENSDART00000062551	Ensembl
Gallus gallus	ENSGALT00000010885	ENSGALT00000025943	ENSGALT00000010892	NM_001048077	Ensembl + NCBI
Gasterosteus aculeatus	ENSGACT00000020316	ENSGACT00000024919	ENSGACT00000014688		Ensembl
Homo sapiens	ENST00000224356	ENST00000001146	ENST00000285949	ENST00000003100	Ensembl
Latimeria chalumnae	ENSLACT00000018096	ENSLACT00000017367		XM_014495166	Ensembl + NCBI
Lepisosteus oculatus	ENSLOCT00000008440	ENSLOCT00000001423	ENSLOCT00000008459	XM_006636012	Ensembl + NCBI
Lethenteron japonicum	JL2346	JL2347 (**a**) & JL2570 (**b**)		JL5543	http://jlampreygenome.imcb.a-star.edu.sg/
Leucoraja erinacea	ctg91131	ctg64746 (**1**) & ctg97427-91018-26897 (**2**)			http://skatebase.org/
Monodelphis domestica	ENSMODT00000008033		XM_007478797	XM_001378153	Ensembl + NCBI
Mus musculus	ENSMUST00000025946	ENSMUST00000077705	ENSMUST00000073391	ENSMUST00000001507	Ensembl
Oryzias latipes	ENSORLT00000018195	ENSORLT00000004309	ENSORLT00000002542		Ensembl
Petromyzon marinus		ENSPMAT00000005233 (**a**) & ENSPMAT00000000847 (**b**)		ENSPMAT00000008420	Ensembl
Takifugu rubripes	ENSTRUT00000004636	ENSTRUT00000032048	ENSTRUT00000032509	XM_003969360	Ensembl + NCBI
Xenopus tropicalis	ENSXETT00000039436	ENSXETT00000046568	ENSXETT00000039433	ENSXETT00000007437	Ensembl

Additional file 1. Accession numbers of the *CYP26* and *CYP51* subfamily sequences used to calculate the phylogenetic trees and to carry out the dating analyses. Species names and database information are provided for each sequence.

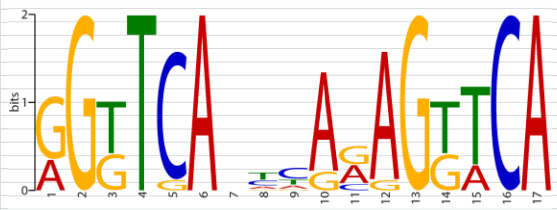

S.p. CYP51
S.k. CYP51
L.a. CYP51
P.c. CYP51
L.g. CYP51
B.f. CYP51
C.t. CYP51
P.m. CYP51
L.j. CYP51
X.t. CYP51A1
D.r. CYP51
L.c. CYP51
C.m. CYP51
L.o. CYP51
T.r. CYP51
G.g. CYP51
M.d. CYP51
H.s. CYP51
M.m. CYP51
C.i. CYP26a
C.g. CYP26c
O.b. CYP26
L.g. CYP26a
C.g. CYP26a
C.t. CYP26a
C.t. CYP26b
C.g. CYP26b
L.g. CYP26b
L.a. CYP26b2
L.a. CYP26b1
L.a. CYP26a1
L.a. CYP26a2
P.c. CYP26b
P.c. CYP26a
B.l. CYP26-3
B.f. CYP26-3
B.b. CYP26-3
B.f. CYP26-2
B.l. CYP26-2
B.b. CYP26-2
B.l. CYP26-1
B.f. CYP26-1
B.b. CYP26-1
P.f. CYP26a
P.f. CYP26b
S.k. CYP26a
S.k. CYP26b
P.f. CYP26c
S.p. CYP26
C.i. CYP26b
H.s. CYP26B1
M.m. CYP26B1
G.g. CYP26B1
X.t. CYP26B1
L.c. CYP26B1
G.a. CYP26B1
O.l. CYP26B1
T.r. CYP26B1
D.r. CYP26B1
L.o. CYP26B1
C.m. CYP26B1
L.e. CYP26B1
L.e. CYP26B2
C.m. CYP26B2
H.s. CYP26C1
M.m. CYP26C1
M.d. CYP26C1
G.g. CYP26C1
X.t. CYP26C1
G.a. CYP26C1
O.l. CYP26C1
T.r. CYP26C1
D.r. CYP26C1
L.o. CYP26C1
C.m. CYP26C1
L.j. CYP26B1/C1a
P.m. CYP26B1/C1a
P.m. CYP26B1/C1b
L.j. CYP26B1/C1b
H.s. CYP26A1
M.m. CYP26A1
M.d. CYP26A1
G.g. CYP26A1
L.c. CYP26A1
T.r. CYP26A1
G.a. CYP26A1
O.l. CYP26A1
D.r. CYP26A1
L.o. CYP26A1
X.t. CYP26A1
L.e. CYP26A1
L.j. CYP26A1
0.5

Additional file 2. Phylogenetic tree of the *CYP26* subfamily, with *CYP51* as outgroup. Tree calculated with the Bayesian Inference (BI) and Maximum Likelihood (ML) methods based on 15 vertebrate species (42 *CYP26* sequences), 1 tunicate species (2 *CYP26* sequences), 3 cephalochordate species (9 *CYP26* sequences), 3 ambulacrarian species (6 *CYP26* sequences), 5 lophotrochozoan species (12 *CYP26* sequences), and 1 ecdysozoan species (2 *CYP26* sequences) for details see Additional file 1. The robustness of each node was assessed by posterior probability (for BI phylogeny) and bootstrap (for the ML phylogeny) analyses, which are indicated at each node: posterior probabilities (ranging from 0 to 1)/bootstrap percentages (ranging from 0 to 100). "--" indicates nodes where the ML tree did not recover the branching pattern of the BI tree, or with bootstrap inferior to 50%. Branch lengths are representative of the amino acid substitution rate. Species name abbreviations: B.b., *Branchiostoma belcheri* (in blue); B.f., *Branchiostoma floridae* (in red); B.l., *Branchiostoma lanceolatum* (in green); C.g., *Crassostrea gigas*; C.i., *Ciona intestinalis*; C.m., *Callorhinchus milii;* C.t., *Capitella teleta*; D.r., *Danio rerio*; G.a., *Gasterosteus aculeatus*; G.g., *Gallus gallus*; H.s., *Homo sapiens*; L.a. *Lingula anatina*; L.c., *Latimeria chalumnae*; L.e., *Leucoraja erinacea*; L.g., *Lottia gigantea*; L.j., *Lethenteron japonicum*; L.o., *Lepisosteus oculatus*; M.d., *Monodelphis domestica*; M.m., *Mus musculus*; O.b., *Octopus bimaculoides*; O.l., *Oryzias latipes*; P.c., *Priapulus caudatus*; P.f., *Ptychodera flava*; P.m., *Petromyzon marinus*; S.k., *Saccoglossus kowalevskii*; S.p., *Strongylocentrotus purpuratus*; T.r., *Takifugu rubripes*; X.t., *Xenopus tropicalis*.


B.b. CYP26-1
B.f. CYP26-1
B.l. CYP26-1
B.b. CYP26-3
B.l. CYP26-3
B.f. CYP26-3
B.b. CYP26-2
B.l. CYP26-2
B.f. CYP26-2
C.m. CYP26B1
L.e. CYP26B1
D.r. CYP26B1
G.a. CYP26B1
T.r. CYP26B1
O.l. CYP26B1
L.o. CYP26B1
G.g. CYP26B1
X.t. CYP26B1
H.s. CYP26B1
M.m. CYP26B1
L.c. CYP26B1
L.e. CYP26B2
C.m. CYP26B2
C.m CYP26C1
D.r. CYP26C1
G.a. CYP26C1
O.l. CYP26C1
T.r. CYP26C1
L.o. CYP26C1
G.g. CYP26C1
H.s. CYP26C1
M.m. CYP26C1
M.d. CYP26C1
X.t. CYP26C1
L.j. CYP26B1/C1a
P.m. CYP26B1/C1a
L.j. CYP26B1/C1b
P.m. CYP26B1/C1b
D.r. CYP26A1
G.a. CYP26A1
O.l. CYP26A1
T.r. CYP26A1
L.o. CYP26A1
L.c. CYP26A1
L.e. CYP26A1
G.g. CYP26A1
H.s. CYP26A1
M.m. CYP26A1
M.d. CYP26A1
X.t. CYP26A1
L.j. CYP26A1
C.i. CYP26b
L.a. CYP26a2
C.i. CYP26a
P.c. CYP26a
P.c. CYP26b
C.t. CYP26b
C.g. CYP26b
L.g. CYP26b
L.a. CYP26b1
L.a. CYP26b2
P.f. CYP26a
P.f. CYP26b
S.k. CYP26a
S.k. CYP26b
P.f. CYP26c
S.p. CYP26
C.t. CYP26a
L.a. CYP26a2
C.g. CYP26a
C.g. CYP26c
O.b. CYP26
L.g. CYP26a
MESO-PROTEROZOIC
NEO-PROTEROZOIC
PALEOZOIC
MESOZOIC
CENO-ZOIC
1000
750
500
250
0
Cam
Ord
Sil
Dev
Car
Per
Tri
Jur
Cre
Million years ago (Mya)

Additional file 3. Comparative dating tree for the *CYP26* subfamily. Divergence dates of the *CYP26* sequences were estimated using fossil-based calibration intervals. Cladogram node dates represent estimation of the mean node height and the highest posterior density intervals fixed at 95% are shown in parentheses. Geological era abbreviations: Cam, Cambrian; Car, Carboniferous; Cre, Cretaceous; Dev, Devonian; Jur, Jurassic; Ord, Ordovician; Per, Permian; Sil, Silurian; Tri, Triassic. Species name abbreviations: B.b., *Branchiostoma belcheri* (in blue); B.f., *Branchiostoma floridae* (in red); B.l., *Branchiostoma lanceolatum* (in green); C.g., *Crassostrea gigas*; C.i., *Ciona intestinalis*; C.m., *Callorhinchus milii;* C.t., *Capitella teleta*; D.r., *Danio rerio*; G.a., *Gasterosteus aculeatus*; G.g., *Gallus gallus*; H.s., *Homo sapiens*; L.a. *Lingula anatina*; L.c., *Latimeria chalumnae*; L.e., *Leucoraja erinacea*; L.g., *Lottia gigantea*; L.j., *Lethenteron japonicum*; L.o., *Lepisosteus oculatus*; M.d., *Monodelphis domestica*; M.m., *Mus musculus*; O.b., *Octopus bimaculoides*; O.l., *Oryzias latipes*; P.c., *Priapulus caudatus*; P.f., *Ptychodera flava*; P.m., *Petromyzon marinus*; S.k., *Saccoglossus kowalevskii*; S.p., *Strongylocentrotus purpuratus*; T.r., *Takifugu rubripes*; X.t., *Xenopus tropicalis*.

RARE name (orientation)	*B. lanceolatum* (5'–3')	*B. belcheri* (5'–3')	*B. floridae* (5'–3')
DR5-1 (-)	ATGTC**TGACCT**TTGGG**TGAACT**CTGAA	ATGCC**TGACCT**TTTGG**TGAACT**CTAAA	ATGCC**TGACCT**TTGGG**TGAACT**CTAAA
DR5-2 (+)	GCCAG**GGTTCA**TTCAA**AGGACA**CGGAA	GCCAG**GGTTCA**TCCAA**AGGACA**CGGAA	GCCAG**GGTTCA**TCCAA**AGGACA**CGGAA
DR5-3 (-)	CTCAC**TGTCCT**CTGAC**TGAACC**CAATA	CTCAC**TGTCCT**CTGAA**TGAACC**CGATA	CTCAT**TGTCCT**CTGGA**TGAACC**CAATA
DR5-4 (-)	GGAAC**TGTCCT**TTTGG**TGAACC**CACTC	GGAGC**TGTCCT**TTTGT**TGAACC**CACTC	GGAAC**TGTCCT**TTTGT**TGAACC**CACTC
DR5-5 (-), DR5-6 (-), DR5-7 (-)	CTTTA**TGAACT**CCAGG**TGAACT**TTTGT**TGAACT**TTGTC**TGACCC**CCCCG	CTTTA**TGAACT**CCAGG**TGAACT**TTTGT**TGAACT**TTGTC**TGACCC**CCCCG	CTTTA**TGAACT**CCAGG**TGAACT**TTTGT**TGAACT**TTGCC**TGACCC**CCCCG
DR5-8 (-)	GAGAA**TGAACC**CTGTT**TGACCC**GTCAG	GAGAG**TGAACC**CTGTT**TGACCC**GTCAG	GAGAG**TGAACC**CTGTT**TGACCC**AGCAG
DR2-1 (+)	CGGCC**GGGTCA**AG**AGTTCA**GATGG	GTCCA**GGGTCA**AG**AGTTCA**GATGG	GGCCG**GGGTCA**AG**AGTTCA**GATGG
DR5-9 (-)	GGGCC**TGAACT**CTGCC**TGAACT**CAGCC	GGGCC**TGAACT**CTGCC**TGAACT**CAGCC	CTGCC**TGAACT**CAGTC**TGAACT**TTCGA
DR5-10 (-), DR3-1 (-), DR5-11 (-), DR3-2 (-)	TGTAG**TGAACT**GCAAA**TGACCC**TCG**TGAACT**CTACC**TGACCT**TCCGTTCCA**TGACCC**CCT**TGACCC**ATATC	TGTGG**TGAACT**CCAAA**TGACCC**TCC**TGAACT**CTACC**TGACCT**TCCGTTCCA**TGACCC**CCT**TGACCC**ATATC	TATGG**TGAACT**CTAAA**TGACCC**TCC**TGAACT**CTCCC**TGACCT**TCCGATCCA**TGACCC**CCT**TGACCC**ATATC
DR5-12 (-)	CGAGC**TGAACT**CTGAA**TGAACC**ACATA	CGACC**TGAACT**CTGAA**TGAACC**ACTTA	CGAGC**TGAACT**CTGAA**TGAACC**ACTTA
DR5-13 (+)	AGGTG**GGGTGA**ATGAC**AGGTCA**GTTGC	AGGTG**GGGTGA**ATGAC**AGGTCA**ACCGG	AGGTG**GGGTGA**ATGAC**AGGTCA**ACCGA



Additional file 4. Conserved retinoic acid response elements (RAREs) found in the *Branchiostoma lanceolatum*, *Branchiostoma belcheri*, and *Branchiostoma floridae CYP26* clusters. *B. lanceolatum* RAREs recognized and bound by the *B. lanceolatum* RAR/RXR heterodimer *in vitro* (Fig. 8) are indicated in black, RAREs that do not associate with the *B. lanceolatum* RAR/RXR heterodimer *in vitro* are shown in grey. The *in vitro* validated *B. lanceolatum* RAREs composed of two direct repeats (DRs) and separated by a 5 nucleotide spacer (i.e. the validated *B. lanceolatum* DR5 type RAREs) were used to calculate a consensus *B. lanceolatum* DR5 RARE by multiple expectation maximization algorithm for motif elicitation (MEME) analysis.

Species name	Database information and genome version	Scaffold
Capitella teleta	JGI v1.0 (Capca1)	251
		827
Lottia gigantea	JGI v1.0 (Lotgi1)	12
		26
Strongylocentrotus purpuratus	EchinoBase v3.1 (Spur_v3.1)	2064
Saccoglossus kowalevskii	Metazome v3.0 (Sakowv3)	1401
Branchiostoma floridae	JGI v1.0 (Brafl1)	164
Branchiostoma belcheri	Mosas (B. belcheri_v18h27.r3_ref_genome)	44
Homo sapiens	Ensembl (GRCh38.p3)	Chrom 10
Mus musculus	Ensembl (GRCm38.p4)	Chrom 19
Takifugu rubripes	Ensembl (FUGU 4.0)	320

Additional file 5. Species names, corresponding genome access, and scaffold information related to the sequences used as input for the *in silico* identification of conserved putative retinoic acid response elements (RAREs).

Gene	Forward primer	Reverse primer
18S	CCTACATGGATAACTGTGGTAAT	GTGCGATCGGCTGGGTTATC
RAR	GTCTGCCACCGGGATAAGAA	ATAACCACAATCAGAGAGGC
CYP26-1	ATCCCATTCGGAGGACTTTC	TTCTTTTGATGTGGCCTTCC
CYP26-2	CAACACCTCACTTTCCTCTTCAC	TCTTCCTCTGAATGTGGTTCATC
CYP26-3	CAAGACGAGGACGAGATCAGTAG	CTTCTCGTGGATGTGACGTTTTA

Additional file 6. Sequences of the oligonucleotides used as primers for quantitative real time PCR (qPCR) analyses on the indicated *B. lanceolatum* genes.

Appendix 3

Additional file 1. RNAseq final dataset. (12 pages)

Additonal file2. ATACseq final dataset. . (8 pages)

Additional file 3. Raw dataset showing all Fox genes and genes described as regulated by RA that were not included in the final dataset. (2 page)

Additional file 4. *In situ* hybridization for some genes present in the RNAseq final dataset (4pages)

		N0				N3				N4						
Cluster	Gene	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS4934	l2FC R1158667	l2FC RA	Auto nomenclature	N.º of RAREs	ATACseq peaks	RARE hubs
1	BL06879	NA	NA	NA	NA	1131,06	NA	0,62	4,13	3251,82	-1,84	3,60	3,57	CYP2J2	2	Yes	No
1	BL19425	1063,34	NA	NA	1,24	6715,76	-0,65	1,53	3,97	8271,85	-1,29	3,57	3,82	CYP26B1	7	Yes	Yes
1	BL19609	14,86	NA	NA	0,29	155,70	-0,53	1,78	3,54	180,87	-0,93	3,51	3,67	CYP26A1	6	Yes	No
2	BL00743	78,71	NA	NA	0,42	479,53	-0,61	0,62	1,99	1220,72	-1,87	1,74	1,86	HMX3	1	No	No
2	BL00808	13,92	NA	NA	0,23	515,54	-1,28	0,77	2,26	1370,23	-1,51	0,87	1,24	ACSF2	1	Yes	No
2	BL05994	37,28	NA	NA	0,81	252,94	-0,88	0,63	2,28	570,43	-1,85	1,21	1,56	THNSL2	1	No	No
2	BL06458	NA	NA	NA	NA	144,52	-0,49	0,50	2,12	370,12	-1,67	1,11	1,70	SLC22A5	5	No	No
2	BL08252	9,38	NA	NA	0,29	157,43	-0,84	0,87	2,68	527,73	-2,14	1,31	1,59	FRRS1	3	No	No
2	BL10029	NA	NA	NA	NA	61,52	-0,65	0,50	1,77	77,05	-1,70	1,32	1,88	SULT1C2	0	Yes	No
2	BL10628	149,70	NA	NA	0,42	319,38	-1,62	0,71	1,94	460,65	-2,37	1,20	1,28	Hypp15387	0	Yes	No
2	BL13668	NA	NA	NA	NA	106,26	NA	NA	2,90	3975,26	-1,26	0,59	1,36	NLGN4X	1	No	No
2	BL14407	17,98	NA	NA	0,41	81,42	-0,78	0,49	1,75	136,98	-2,02	1,13	1,28	DAO	2	No	Yes
2	BL15783	NA	NA	NA	NA	111,98	NA	1,05	3,36	646,82	-1,66	1,78	1,10	ANKRD10	0	Yes	No
2	BL16767	17,34	NA	NA	0,17	1037,20	-1,84	0,51	1,48	2331,90	-2,59	0,77	0,70	HNF1B	0	Yes	No
2	BL16830	NA	NA	NA	NA	188,77	-0,81	0,76	2,82	583,43	-2,36	1,59	2,36	Hypp4760	0	No	No
2	BL18624	87,01	NA	NA	0,48	1881,16	-2,00	0,69	2,47	2202,83	-2,00	1,06	1,48	ITGB2	2	Yes	No
2	BL26793	15,38	NA	NA	0,34	155,62	-0,53	0,75	3,39	512,90	-2,96	1,39	2,46	Hypp217	2	No	No
2	BL39466	NA	NA	NA	NA	80,08	-0,70	0,51	2,14	96,03	-1,87	1,32	1,77	SULT1C4	0	Yes	No
2	BL91555	107,19	NA	NA	1,12	955,72	-2,11	0,34	1,13	857,27	-2,14	0,68	1,02	Hypnc52099	4	Yes	No
2	BL95634	20,09	NA	NA	0,32	213,92	-0,86	0,98	2,79	738,28	-1,67	1,56	1,86	ABCC8	2	Yes	No
2	BL95885	NA	NA	NA	NA	111,67	-0,61	0,58	2,04	248,31	-1,74	1,17	1,64	MOSC1	3	No	No
3	BL00688	315,44	NA	NA	0,57	1202,37	-0,75	0,41	1,60	2063,89	-1,23	0,80	0,88	ACCS	2	No	No
3	BL00973	35,85	NA	NA	0,33	191,46	-1,17	NA	1,18	231,89	-1,46	0,65	0,72	DHRS4	4	No	No
3	BL01008	95,64	NA	NA	0,50	284,53	-0,70	NA	0,86	369,41	-0,82	0,75	0,73	PGP	1	No	No
3	BL01409	240,49	-0,35	NA	0,95	159,18	-0,76	NA	1,08	112,47	-0,72	0,65	0,94	HOXB2	4	Yes	No
3	BL01466	64,90	NA	NA	0,54	121,81	-1,10	NA	1,08	115,23	-1,53	0,83	0,93	PEX7	4	No	No
3	BL02461	112,82	NA	0,67	1,04	271,72	-0,93	0,77	1,31	311,44	-1,24	1,20	1,18	SLC22A18	5	No	Yes
3	BL04198	666,87	-0,35	NA	0,97	875,71	-1,02	0,39	1,38	919,03	-1,24	0,72	0,93	HRSP12	2	No	No
3	BL05510	97,78	NA	NA	0,30	1361,67	-1,03	0,43	1,26	1425,68	-1,41	0,48	0,85	BHMT	5	No	No
3	BL07788	NA	NA	NA	NA	171,95	-0,64	0,42	1,28	441,49	-1,17	0,80	0,69	Hypp14981 (NRF-6 like)	5	Yes	No
3	BL07802	21,03	NA	NA	0,23	332,91	-0,88	NA	1,32	341,45	-1,31	0,46	0,81	Hypp10695	5	No	No
3	BL07881	97,31	NA	NA	0,44	437,07	-1,17	0,41	1,20	562,24	-1,34	0,74	0,84	SDS	1	No	No
3	BL08837	52,50	NA	NA	0,36	167,40	-0,60	NA	0,89	278,67	-0,92	0,65	0,87	HSD17B6	1	No	No
3	BL10401	6313,52	-0,58	NA	0,67	8824,68	-1,00	0,38	1,08	8002,48	-0,83	1,04	0,93	RARB	22	Yes	No
3	BL10663	NA	NA	NA	NA	76,47	-0,86	0,59	1,34	124,18	-1,25	0,87	0,96	Hypp6551	1	No	No
3	BL11528	NA	NA	NA	NA	219,39	-1,31	NA	1,33	267,84	-1,03	0,47	0,85	SLC25A15	1	No	No
3	BL11673	32,72	NA	NA	0,46	163,69	-0,73	NA	0,85	208,42	-0,78	0,76	0,81	ENSG00000262563	2	No	No
3	BL11746	67,34	NA	NA	0,47	70,34	-0,50	NA	0,76	87,49	-0,70	0,68	1,16	ADD1	2	No	No
3	BL12380	NA	NA	NA	NA	102,89	-0,54	0,56	1,45	201,81	-1,20	0,81	0,62	AC000120.8	6	No	No
3	BL12864	500,24	-0,35	NA	0,73	1632,93	-0,95	0,28	1,20	2608,65	-1,39	0,52	1,02	GSTM3	3	No	No
3	BL13011	830,44	-0,29	NA	0,51	1075,74	-0,54	NA	1,04	577,10	-1,10	0,47	0,93	PRSS33	0	No	No
3	BL13050	492,01	-0,29	0,55	1,10	1018,09	-0,61	1,00	1,38	2518,86	-0,75	1,37	0,75	ABCA5	1	No	No
3	BL13238	735,75	-0,39	NA	0,47	2473,93	-0,86	0,26	1,25	2533,63	-0,97	0,64	1,15	Hypp16164	2	Yes	No
3	BL13375	NA	NA	NA	NA	1711,70	-0,91	0,39	0,90	1023,54	-1,30	1,32	1,45	Hypp15523	1	No	No
3	BL13989	1083,50	NA	NA	0,26	216,94	-0,90	0,53	0,91	141,65	-1,08	0,98	1,06	ST8SIA2	0	Yes	No
3	BL14085	973,55	NA	NA	0,41	1855,84	-0,75	0,20	0,77	2063,43	-1,01	0,45	0,57	ALDH6A1	4	No	No
3	BL14546	1103,33	-0,66	0,43	1,67	5343,06	-0,79	0,27	0,94	5540,68	-1,20	0,59	0,88	HOX3	2	Yes	No
3	BL14845	15,88	NA	NA	0,21	402,91	-1,23	0,71	1,66	546,51	-1,93	0,64	0,97	EDIL3	7	No	No
3	BL15073	NA	NA	NA	NA	90,72	-0,59	0,57	2,03	149,30	-1,64	0,73	0,90	LAMP1	1	No	No
3	BL15643	246,52	NA	NA	0,84	342,76	-0,71	0,54	1,07	347,84	-0,91	0,83	0,77	ACOX1	1	No	No
3	BL15734	354,43	-0,28	NA	0,47	604,76	-1,00	0,32	1,23	888,60	-1,30	1,20	1,31	SQRDL	5	No	No
3	BL17288	693,89	-0,43	NA	0,53	898,40	-1,09	NA	0,57	885,51	-1,37	0,45	0,42	FOXRED1	2	No	No
3	BL17840	233,32	NA	NA	0,86	720,93	-0,90	0,28	1,21	686,70	-1,05	0,80	0,98	ADH5	1	No	No
3	BL17851	1353,33	-0,49	NA	0,49	666,42	-1,22	0,46	0,96	803,09	-1,62	0,67	0,61	ESR2	0	No	No
3	BL18915	309,71	-0,29	NA	0,39	294,67	-1,15	NA	1,11	276,97	-1,48	0,59	0,72	C9orf64	3	No	No
3	BL19115	42,75	NA	NA	0,45	280,34	-0,91	NA	0,93	223,32	-0,92	0,53	0,60	GRHPR	2	No	No
3	BL19627	458,99	-0,48	NA	0,66	2495,37	-0,74	0,22	0,92	3973,37	-1,24	NA	NA	A1CF	4	Yes	No
3	BL20098	487,62	-0,29	NA	0,64	391,43	-0,88	0,32	1,02	336,06	-1,26	0,71	0,78	RDH10	2	Yes	No
3	BL20190	61,47	NA	NA	0,44	184,76	-0,71	NA	1,00	257,92	-1,01	0,58	0,80	IDO2	0	No	No
3	BL20667	355,33	NA	NA	0,40	478,93	-0,91	NA	0,93	466,77	-0,75	0,71	0,79	SFXN1	4	No	No
3	BL22295	NA	NA	NA	NA	267,37	-1,15	0,38	1,70	1096,71	-1,49	0,52	0,95	CD9	1	Yes	No
3	BL22740	NA	NA	NA	NA	455,57	-0,87	0,42	1,31	829,60	-1,33	1,01	1,12	CYP46A1	1	No	No
3	BL23299	1916,41	-0,43	NA	0,51	1866,61	-0,54	0,35	0,79	1457,33	-0,93	1,04	0,76	ABCG2	4	No	No
3	BL57081	160,49	NA	NA	0,71	130,38	-0,63	NA	0,82	98,89	-0,85	0,83	1,12	Hypnc21869	4	No	No
3	BL86317	NA	NA	NA	NA	122,98	-0,54	0,78	1,41	270,13	-1,38	0,95	0,80	Hypnc47474	2	No	No
3	BL95117	NA	NA	NA	NA	144,56	-1,00	0,43	1,17	197,60	-1,51	0,73	0,90	Hypp8245	3	No	No
3	BL95460	NA	NA	NA	NA	233,30	-0,98	0,55	1,71	445,44	-1,88	0,75	0,45	OSR2	4	Yes	No
3	BL95525	18,77	NA	NA	0,32	74,46	-0,90	NA	1,17	98,74	-1,24	0,82	0,80	SLC16A12	2	No	No
4	BL00388	921,92	-0,44	NA	1,26	1219,05	NA	0,31	1,34	943,60	-0,54	1,20	1,15	CYP26C1	1	No	No
4	BL01539	388,47	NA	0,59	0,89	455,32	NA	0,64	0,67	468,53	NA	0,99	0,76	CYP2U1	1	No	No
4	BL02122	NA	NA	NA	NA	155,07	NA	1,06	1,37	520,64	NA	1,74	1,20	KCNAB1	0	No	No
4	BL03125	16,35	NA	NA	0,21	60,03	NA	0,69	1,20	148,31	NA	0,99	1,34	RP11-369J21.6	2	No	No
4	BL03556	252,06	NA	NA	0,35	321,34	NA	0,44	0,59	369,00	NA	1,01	0,80	EPHX4	2	No	No
4	BL03787	164,14	NA	0,58	0,75	199,10	NA	0,61	0,32	NA	NA	NA	NA	CYP3A4	2	No	No
4	BL03985	377,51	NA	0,93	0,93	1663,05	0,31	0,74	0,60	2230,86	NA	0,91	NA	CYP4F12	4	No	No
4	BL04579	NA	NA	NA	NA	132,53	NA	0,49	0,57	323,09	NA	0,76	0,48	ABCA3	1	No	No
4	BL05041	88,97	NA	0,53	0,82	156,25	NA	0,79	1,20	169,21	NA	1,33	0,96	EPHX4	2	No	No
4	BL05204	74,34	NA	NA	0,51	239,94	NA	0,49	1,25	252,47	NA	0,74	0,83	STS	5	Yes	No
4	BL05579	4454,57	NA	0,39	0,35	5583,33	NA	0,38	0,47	5836,41	NA	0,61	0,50	MGST3	1	No	No
4	BL06167	133,68	NA	0,33	0,59	336,02	-0,53	0,63	0,79	437,67	-0,66	0,87	0,69	ABHD3	1	No	No
4	BL06312	9,81	NA	NA	0,19	97,08	NA	0,88	1,35	147,79	NA	0,90	0,72	CYP2U1	2	No	No
4	BL06352	NA	NA	NA	NA	188,37	NA	0,55	0,46	171,15	NA	0,85	0,62	Hypp9140	1	No	No
4	BL06617	201,08	NA	0,31	0,46	195,03	NA	0,70	0,78	223,38	NA	0,80	NA	HACL1	4	No	No
4	BL06835	169,23	NA	0,81	1,08	436,81	NA	0,85	0,92	443,84	NA	0,53	NA	ECHDC3	3	No	No
4	BL07039	96,17	NA	0,35	0,56	361,25	NA	0,57	0,50	492,33	NA	0,61	NA	CYP24A1	3	No	Yes
4	BL07211	276,71	NA	1,06	1,04	411,09	NA	0,85	0,79	491,38	NA	0,78	0,43	C9orf46	1	No	No
4	BL10832	65,13	NA	0,65	0,94	89,33	NA	NA	0,42	101,66	NA	1,02	0,85	CYP2D6	0	No	No
4	BL10967	187,89	NA	0,98	1,28	208,76	NA	1,51	2,07	161,64	NA	1,17	0,83	UGT2A1	3	No	No
4	BL11540	237,67	NA	0,47	0,63	392,40	NA	0,44	0,80	430,12	NA	0,43	0,53	PXMP2	0	No	No
4	BL11591	746,06	NA	1,03	1,54	1362,18	NA	0,54	1,09	1495,23	NA	1,52	1,46	CRABP2	1	No	No
4	BL12419	NA	NA	NA	NA	528,14	NA	0,49	1,42	1849,17	-0,42	1,15	1,01	RHAG	3	No	No
4	BL12883	408,47	NA	1,33	1,59	386,18	0,45	1,42	1,54	NA	NA	NA	NA	CYP2U1	0	No	No
4	BL13182	217,11	NA	0,60	0,96	810,37	-0,28	0,53	1,10	946,29	-0,59	0,54	0,67	IGFALS	0	No	No
4	BL14060	172,95	NA	0,80	0,71	306,28	0,42	2,26	2,46	340,28	NA	1,49	1,09	SULT1C4	2	No	No
4	BL14416	38,97	NA	NA	0,28	91,06	NA	0,48	0,61	108,65	NA	0,78	0,55	RDH13	1	No	No
4	BL15433	107,92	NA	0,43	0,75	221,90	NA	0,78	0,88	258,20	NA	0,72	0,47	GSTT1	1	No	No
4	BL15644	106,63	NA	0,41	0,56	169,39	NA	0,89	0,70	167,85	NA	0,84	0,54	LACTB2	0	No	No
4	BL16564	522,12	NA	0,58	0,67	929,60	NA	0,26	0,28	905,55	NA	0,48	NA	OSGIN2	0	No	No
4	BL17661	203,55	NA	0,46	0,69	315,54	-0,53	0,51	0,67	410,40	-0,55	0,75	0,60	ACOT7	0	No	No
4	BL17714	39,14	NA	0,40	0,93	93,05	NA	1,39	3,16	55,89	NA	1,39	2,16	Hypp15083	0	No	No
4	BL18519	71,32	NA	0,56	0,84	232,37	NA	1,14	1,63	318,65	NA	1,17	1,06	CYP2U1	3	No	No
4	BL20594	18,47	NA	NA	0,30	117,26	NA	1,22	1,23	89,38	NA	1,17	0,91	Hypp12399	1	No	No
4	BL20884	1185,53	NA	0,36	0,50	952,85	NA	0,40	0,52	670,80	NA	0,66	0,50	NR1H4	2	No	Yes
4	BL21592	173,93	NA	0,50	0,78	80,29	NA	NA	0,88	NA	NA	NA	NA	AC069257.9	0	No	No
4	BL22293	360,17	NA	0,79	1,14	762,43	NA	0,56	0,56	733,77	NA	0,60	NA	CYP27A1	3	No	No
4	BL22377	312,42	NA	NA	0,70	921,14	-0,39	0,47	0,98	NA	NA	NA	NA	CCDC30	1	No	No
4	BL23419	79,65	NA	0,59	0,79	126,74	NA	0,98	1,21	138,68	NA	0,95	0,74	CYP2J2	2	No	No
4	BL23676	54,47	NA	NA	0,40	303,32	-0,65	0,53	1,09	501,05	NA	1,07	0,72	NR1H4	2	No	No
4	BL24655	59,72	NA	0,70	1,03	79,67	NA	NA	0,34	112,36	NA	1,13	0,71	CYP2U1	1	No	No
4	BL40355	44,40	NA	NA	0,42	124,36	NA	0,47	0,51	118,58	NA	0,73	0,75	UGT1A6	0	Yes	No

		N0				N3				N4							
Cluster	Gene	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS4934	l2FC R1158667	l2FC RA	Auto nomenclature	N.º of RAREs	ATACseq peaks	RARE hubs
4	BL42534	86,37	NA	0,67	0,74	84,34	NA	1,11	1,03	67,34	NA	1,04	0,81	Hypnc9043	0	No	No
4	BL47446	107,34	NA	NA	0,94	37,28	NA	NA	1,14	43,60	NA	0,80	0,67	WFIKKN2	2	No	No
4	BL70565	51,70	NA	NA	0,43	62,37	NA	0,93	1,72	50,40	NA	1,90	1,57	Hypnc33708	6	Yes	No
4	BL75757	1894,52	NA	NA	0,23	3048,30	NA	0,79	1,52	3673,84	NA	1,13	1,19	ABCC5	0	Yes	No
4	BL77285	667,68	NA	NA	1,14	961,87	NA	0,45	1,15	854,27	NA	1,48	1,50	Hypnc39566	0	No	No
4	BL78634	56,91	NA	NA	0,53	180,73	-0,67	0,41	1,12	262,82	NA	0,74	0,53	C10orf11	8	Yes	Yes
4	BL82040	47,21	NA	0,29	0,44	85,81	NA	0,67	0,66	131,57	NA	1,00	NA	Hypp60	1	No	No
4	BL86797	220,99	NA	0,47	0,67	272,18	NA	0,77	0,77	264,25	NA	1,27	1,04	Hypp15216	0	No	No
4	BL95162	325,36	NA	0,58	0,74	893,54	NA	0,40	0,64	NA	NA	NA	NA	ABCC9	2	No	No
4	BL95215	57,29	NA	0,39	0,63	157,99	NA	0,94	1,96	188,66	NA	1,01	1,33	ARSB	1	No	No
4	BL95353	847,38	NA	0,43	0,83	1186,99	NA	0,51	1,10	NA	NA	NA	NA	SLCO3A1	1	No	No
4	BL95363	NA	NA	NA	NA	120,97	NA	0,63	0,58	284,16	NA	1,02	0,47	EPHX1	0	No	No
4	BL95508	65,31	NA	NA	0,38	206,50	NA	0,53	0,68	277,20	NA	0,59	0,47	ACOX1	2	No	No
4	BL96029	409,72	NA	0,84	1,07	513,91	-0,41	0,65	0,94	512,64	NA	0,82	0,65	ALDH3A2	3	No	No
4	BL97374	72,60	NA	0,55	0,75	39,03	NA	NA	0,75	32,38	NA	1,27	0,76	UGT1A1	3	No	No
5	BL00787	109,92	NA	NA	0,36	500,84	-0,57	0,27	0,68	664,24	-0,52	0,51	0,53	VAC14	2	No	No
5	BL01238	NA	NA	NA	NA	221,00	-0,57	NA	1,20	310,57	-1,08	0,62	0,94	ADHFE1	2	No	No
5	BL01262	170,08	NA	NA	0,29	9013,51	-1,00	NA	0,29	5097,90	-1,62	NA	0,52	Hypp1279	2	No	No
5	BL01271	2399,69	NA	NA	0,16	2714,93	-0,51	NA	0,32	2643,92	-0,62	0,29	NA	RXRA	6	No	Yes
5	BL01875	NA	NA	NA	NA	3317,72	-0,43	NA	0,51	2868,71	-0,68	0,21	0,28	Hypp5878	4	No	No
5	BL02328	2723,39	NA	NA	0,31	4236,39	-0,60	NA	0,55	3225,63	-0,73	NA	0,29	PCCA	1	No	No
5	BL02371	NA	NA	NA	NA	197,45	-0,57	NA	0,95	390,60	-1,08	0,35	0,43	Hypp13139	3	No	No
5	BL02374	NA	NA	NA	NA	374,88	-0,57	NA	0,83	275,15	-0,85	0,47	0,72	ACSF2	1	No	No
5	BL02387	624,15	NA	NA	0,55	841,23	-0,59	NA	0,62	771,88	-0,55	0,38	0,48	BRP44	3	No	No
5	BL02628	NA	NA	NA	NA	127,46	NA	NA	1,22	542,60	-0,80	0,35	0,76	Hypp9095	0	No	No
5	BL02762	8918,87	-0,22	NA	0,20	8754,66	-0,66	0,21	0,66	8620,12	-0,74	0,33	0,28	IVNS1ABP	0	No	No
5	BL03040	NA	NA	NA	NA	473,88	-0,58	NA	0,47	458,10	-0,81	0,53	0,44	OXCT1	4	No	No
5	BL03583	NA	NA	NA	NA	1024,21	-0,71	0,22	0,79	1192,09	-1,05	0,45	0,59	ACSS3	5	No	No
5	BL03903	NA	NA	NA	NA	204,97	NA	NA	1,69	2271,39	-0,76	0,38	0,73	CADM4	1	No	No
5	BL04138	269,73	NA	NA	0,54	456,15	-0,49	0,46	0,91	477,53	-0,56	0,66	0,55	MATN2	1	No	No
5	BL05173	NA	NA	NA	NA	5012,37	-0,31	NA	0,27	1733,05	-0,69	NA	0,38	HSD11B1	0	No	No
5	BL05210	NA	NA	NA	NA	451,33	-0,65	NA	0,85	516,17	-0,89	NA	0,63	TBC1D24	1	No	No
5	BL05214	NA	NA	NA	NA	470,99	-0,59	NA	0,56	514,19	-0,64	NA	0,63	TBC1D24	0	No	No
5	BL05311	NA	NA	NA	NA	554,44	-0,75	0,28	0,91	429,15	-1,42	0,50	0,77	NPC1	0	No	No
5	BL05940	NA	NA	NA	NA	5280,31	-0,34	NA	0,36	8894,11	-0,70	NA	0,48	Hypp12161	2	No	Yes
5	BL06303	NA	NA	NA	NA	1162,52	-0,63	NA	0,20	970,93	-0,49	0,28	0,34	UBFD1	1	No	No
5	BL06377	NA	NA	NA	NA	311,30	-0,42	NA	0,72	1747,04	-0,88	NA	0,72	Hypp14155	1	No	No
5	BL06743	NA	NA	NA	NA	138,09	NA	0,62	1,49	234,26	-0,73	0,54	0,52	SKAP2	2	Yes	Yes
5	BL06754	NA	NA	NA	NA	56,71	-0,58	NA	0,84	102,57	-1,17	NA	0,56	CYP2U1	3	Yes	No
5	BL06976	2265,46	NA	NA	0,25	3343,46	-0,21	0,18	0,88	8145,02	-0,35	0,47	0,36	LOXL2	2	No	No
5	BL07183	NA	NA	NA	NA	985,95	-0,41	0,26	0,46	899,03	-0,61	NA	0,35	UNCX	0	No	No
5	BL07260	12780,50	NA	NA	0,23	8007,21	-0,27	0,26	0,76	6792,66	-0,36	NA	0,69	HLF	0	No	No
5	BL07799	NA	NA	NA	NA	437,93	-0,78	NA	0,80	426,71	-1,00	0,47	0,82	AIFM2	10	No	No
5	BL07968	123,56	NA	NA	0,33	338,04	-0,50	NA	0,76	280,97	-1,19	0,51	0,71	Hypp4350	1	No	No
5	BL08301	143,59	NA	NA	0,30	236,13	NA	NA	1,27	413,75	-1,16	0,63	0,95	ALDH1A1	2	No	No
5	BL08310	1227,29	NA	NA	0,21	600,16	-0,78	NA	0,69	366,88	-1,40	0,46	0,84	Hypp13872	6	No	No
5	BL08519	74,33	NA	NA	0,35	420,40	-0,53	0,39	0,98	561,26	-1,00	NA	0,72	Hypp369	3	Yes	No
5	BL08550	NA	NA	NA	NA	2043,99	-0,48	NA	0,47	2016,89	-1,06	0,43	0,62	PVRL4	0	No	No
5	BL08717	NA	NA	NA	NA	456,84	-0,44	NA	0,69	1305,20	-0,66	0,48	0,57	MFSD5	0	No	No
5	BL08793	NA	NA	NA	NA	31,75	NA	NA	1,01	123,07	-1,12	0,68	0,76	IRF2	2	No	No
5	BL08810	1245,04	-0,26	NA	NA	1886,53	-0,38	NA	0,41	1639,81	-0,41	NA	0,69	Hypp12843	3	No	Yes
5	BL08896	5474,28	NA	NA	0,67	4454,25	-0,35	0,46	0,69	5735,18	-0,41	0,44	0,42	ZNF503	1	Yes	No
5	BL09276	NA	NA	NA	NA	317,55	-0,60	NA	1,19	361,94	-0,93	0,55	1,04	ST8SIA2	0	No	No
5	BL09545	NA	NA	NA	NA	335,13	-0,43	NA	1,17	1225,45	-0,68	0,41	0,58	KIAA1161	0	No	No
5	BL09562	39244,02	NA	NA	0,15	7041,20	-0,54	NA	0,39	1305,84	-1,19	0,40	0,75	DDAH1	1	No	No
5	BL10061	473,21	NA	NA	0,36	751,39	-0,62	0,36	0,75	669,79	-0,61	0,65	0,47	HGD	0	No	No
5	BL10232	1231,30	NA	NA	0,28	1841,50	-0,93	NA	0,48	2051,99	-0,88	0,46	0,53	C17orf37	5	No	No
5	BL10835	NA	NA	NA	NA	312,70	-0,63	NA	0,33	258,75	-0,82	0,38	NA	C16orf7	1	No	No
5	BL11574	596,87	NA	NA	0,24	2064,02	-0,33	0,19	0,62	2189,36	-0,46	0,45	0,49	PTCHD1	2	No	No
5	BL11698	NA	NA	NA	NA	1630,97	-0,55	0,27	0,83	2371,18	-0,71	0,45	0,60	SUOX	4	No	No
5	BL11992	3278,82	-0,25	NA	0,60	3569,86	-0,47	0,26	0,69	2536,05	-0,51	0,27	0,27	PPHLN1	3	No	No
5	BL12289	4926,31	-0,15	NA	0,28	3672,76	-0,39	0,32	0,54	2567,62	-0,89	0,58	0,58	HOXA1	4	No	No
5	BL12308	NA	NA	NA	NA	117,38	NA	NA	1,48	1553,42	-1,24	0,70	0,91	SLC6A17	4	No	No
5	BL12457	NA	NA	NA	NA	169,44	-0,67	NA	0,72	180,29	-1,28	0,50	0,58	Hypp7515	1	No	No
5	BL12467	NA	NA	NA	NA	194,37	-0,68	NA	0,86	270,21	-0,96	0,53	0,57	PYROXD2	4	No	No
5	BL12588	71,64	NA	NA	0,37	333,42	-0,79	0,38	0,86	350,80	-0,91	NA	0,73	BHLHA15	1	No	No
5	BL14013	NA	NA	NA	NA	11290,57	-0,53	NA	0,39	2229,15	-0,80	NA	0,39	HAO1	1	No	No
5	BL14433	NA	NA	NA	NA	56,34	-0,68	NA	0,68	101,23	-1,06	NA	0,87	TMEM22	2	No	No
5	BL14477	2643,28	NA	NA	0,26	4147,29	-0,49	NA	0,37	4499,84	-0,62	0,28	0,38	Hypp4979	2	No	No
5	BL14642	111,43	NA	NA	0,48	566,70	NA	0,51	1,03	970,42	-0,73	0,64	0,48	SDR16C5	0	No	No
5	BL14890	NA	NA	NA	NA	421,78	-0,41	NA	0,76	460,55	-0,52	0,59	0,55	DBT	2	No	Yes
5	BL15016	16153,85	NA	NA	0,15	14344,87	-0,44	NA	0,58	13418,80	-0,96	0,30	0,70	CA12	0	No	No
5	BL15112	10077,76	NA	NA	0,19	14627,93	-0,46	0,41	0,98	13731,36	-0,88	0,85	0,65	PBX3	6	Yes	No
5	BL15639	187,89	NA	NA	0,29	276,36	-0,45	NA	0,63	352,56	-0,66	NA	0,43	Hypp15527	0	No	No
5	BL15857	1661,71	NA	NA	0,61	1136,08	-0,55	0,34	1,04	1489,33	-0,74	0,71	0,65	CHST15	2	No	No
5	BL16002	NA	NA	NA	NA	279,79	-0,55	NA	0,50	229,48	-0,53	0,44	0,63	B3GALT1	1	No	No
5	BL16053	3554,81	-0,19	NA	0,24	2436,79	-0,53	NA	0,62	3021,86	-0,73	0,39	0,43	ASNS	2	No	No
5	BL16091	2871,23	NA	NA	0,24	3385,74	-0,44	0,25	0,59	4027,92	-0,52	0,22	0,31	SOX7	2	No	No
5	BL16849	NA	NA	NA	NA	395,52	-0,80	0,39	0,83	418,87	-1,43	0,57	0,66	Hypp7634	0	No	No
5	BL17200	NA	NA	NA	NA	168,96	-0,41	NA	0,72	244,69	-1,04	NA	0,45	DYNLL2	1	No	No
5	BL17472	NA	NA	NA	NA	1794,13	-0,42	NA	1,12	2767,54	-1,19	0,58	0,88	ACE	1	No	No
5	BL17536	331,11	NA	NA	0,36	425,79	-0,79	NA	0,44	358,80	-0,84	0,47	0,51	SMOX	4	No	No
5	BL17760	156,48	NA	NA	0,55	814,60	-0,33	0,31	0,75	907,99	-0,51	0,33	0,43	ACAD11	1	No	No
5	BL17845	NA	NA	NA	NA	4989,97	-0,38	NA	0,55	5494,75	-0,87	NA	0,47	NEUROG2	2	Yes	No
5	BL17962	NA	NA	NA	NA	347,87	-0,59	NA	1,10	366,13	-1,06	0,88	0,98	SLCO4A1	1	No	No
5	BL18450	NA	NA	NA	NA	348,22	-0,57	NA	0,81	463,58	-0,79	0,43	0,56	Hypp10679	2	No	No
5	BL18603	NA	NA	NA	NA	320,68	-0,34	NA	0,99	663,17	-0,59	0,51	0,65	Hypp1663	1	No	No
5	BL19397	528,23	NA	NA	0,59	1634,77	-0,60	NA	0,67	1114,81	-1,23	NA	0,72	ENSG00000262499	1	No	No
5	BL19847	45,99	NA	NA	0,47	186,33	NA	0,52	1,17	247,31	-0,74	0,75	0,78	SERHL2	1	No	No
5	BL19951	NA	NA	NA	NA	518,93	-0,59	NA	0,78	482,15	-1,02	NA	0,47	H6PD	2	No	No
5	BL20218	801,28	NA	NA	0,21	1141,82	-0,48	NA	0,52	1127,43	-0,66	NA	0,43	MFSD2A	1	No	No
5	BL20403	NA	NA	NA	NA	382,22	-0,73	0,32	1,32	727,76	-0,97	0,78	1,06	AGRN	0	No	No
5	BL20462	106,31	NA	NA	0,35	98,98	-0,61	NA	0,74	90,25	-1,18	0,66	0,70	DHRS3	0	No	No
5	BL20505	NA	NA	NA	NA	471,61	-0,65	NA	0,32	513,94	-0,56	0,45	0,42	C3orf23	0	No	No
5	BL20565	NA	NA	NA	NA	399,55	-0,50	NA	1,05	430,01	-1,01	0,85	1,15	DAO	2	No	No
5	BL20566	1214,19	NA	NA	0,26	1422,12	-0,39	NA	0,68	1754,40	-0,75	NA	0,60	NDUFC2	0	No	No
5	BL20743	NA	NA	NA	NA	246,87	-0,50	NA	0,79	279,39	-0,67	NA	0,50	Hypp11837	3	No	No
5	BL20991	NA	NA	NA	NA	51,32	NA	NA	1,16	131,02	-1,18	0,63	0,68	AKR1A1	1	No	No
5	BL21057	NA	NA	NA	NA	3401,67	-0,17	0,30	0,55	3616,86	-0,75	NA	0,34	ENSG00000262217	3	No	Yes
5	BL21398	NA	NA	NA	NA	74,31	-0,51	NA	1,34	238,81	-1,10	0,56	0,50	CYP3A4	2	No	No
5	BL21655	12794,80	NA	NA	0,14	2074,02	-0,61	NA	0,33	247,24	-1,31	0,45	0,85	DDAH1	1	No	No
5	BL21663	NA	NA	NA	NA	789,06	-0,43	NA	0,90	533,27	-1,00	0,35	0,67	CYGB	1	No	No
5	BL21684	NA	NA	NA	NA	293,64	-0,42	NA	0,64	395,91	-0,53	0,33	0,38	ECH1	2	No	No
5	BL21856	NA	NA	NA	NA	3275,74	-0,36	NA	0,69	3089,54	-0,99	NA	0,59	PTPRG	1	No	No
5	BL21983	NA	NA	NA	NA	1408,27	-0,23	NA	0,36	1821,25	-0,59	NA	0,31	SLC24A5	0	No	No
5	BL22432	1643,52	-0,28	NA	0,34	4273,30	-0,49	NA	0,43	2737,55	-0,54	0,51	0,62	SLC25A10	5	Yes	No
5	BL22534	1212,34	-0,26	0,25	0,61	2458,71	-0,77	0,22	0,67	2593,73	-0,74	0,53	0,56	PYCR2	1	No	No
5	BL22552	NA	NA	NA	NA	579,09	-0,82	NA	1,21	800,00	-0,92	NA	0,61	VIM	0	Yes	No
5	BL22745	3788,35	-0,18	NA	0,26	3386,29	-0,55	0,19	0,59	2649,95	-0,82	0,52	0,61	BCKDK	4	No	No
5	BL23237	NA	NA	NA	NA	64,40	-0,65	NA	1,02	49,57	-1,09	NA	1,10	CRYZ	1	No	No

		N0				N3				N4							
Cluster	Gene	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS4934	l2FC R1158667	l2FC RA	Auto nomenclature	N.º of RAREs	ATACseq peaks	RARE hubs
5	BL23369	145,22	NA	NA	0,35	126,58	-0,57	NA	0,38	104,11	-0,86	NA	0,48	MFSD6	0	No	No
5	BL23538	NA	NA	NA	NA	273,27	-0,53	NA	0,94	289,40	-0,95	NA	0,54	Hypp326	3	Yes	No
5	BL23612	110,96	NA	NA	0,50	270,95	-0,59	NA	0,65	223,70	-0,83	NA	0,59	ASRGL1	3	No	No
5	BL23835	NA	NA	NA	NA	63,03	-0,57	NA	0,94	117,58	-1,02	0,61	0,71	SLC22A5	0	No	No
5	BL24121	945,59	NA	NA	0,34	544,04	-0,27	NA	0,65	499,95	-0,49	NA	0,36	GGT1	1	No	No
5	BL24926	NA	NA	NA	NA	243,59	-0,40	NA	1,10	320,19	-1,16	NA	0,56	Hypp3587	2	No	No
5	BL38167	NA	NA	NA	NA	513,03	NA	NA	1,28	2158,49	-0,72	0,42	0,89	CNTN1	0	No	No
5	BL45159	629,33	NA	NA	0,29	854,36	-0,71	NA	0,55	1020,14	-0,79	0,31	0,47	ACAT1	0	No	No
5	BL48149	NA	NA	NA	NA	455,25	-0,63	0,43	0,79	545,81	-1,25	0,60	0,87	Hypnc13982	1	No	No
5	BL52172	157,88	NA	NA	0,54	259,25	-0,32	0,61	0,69	284,28	-0,57	0,75	0,49	TYSND1	0	No	No
5	BL54853	NA	NA	NA	NA	2200,71	-0,37	NA	0,50	2257,08	-0,63	0,27	0,47	BCKDHA	1	No	No
5	BL58405	NA	NA	NA	NA	352,08	-0,39	0,46	1,12	1447,17	-0,70	0,57	0,94	Hypnc23064	0	No	No
5	BL65097	NA	NA	NA	NA	404,19	-0,92	0,34	1,16	602,72	-0,62	0,86	0,56	ENSG00000261899	6	Yes	No
5	BL69071	NA	NA	NA	NA	1347,32	-0,47	NA	0,92	2381,00	-1,10	0,39	0,41	ANGPT2	2	No	No
5	BL69319	NA	NA	NA	NA	314,01	-0,62	NA	1,20	1041,36	-0,67	NA	0,45	Hypnc32603	1	No	No
5	BL71558	NA	NA	NA	NA	57,41	-0,68	NA	0,87	123,65	-1,14	0,72	0,71	Hypp8613	0	Yes	No
5	BL72107	404,19	NA	NA	0,46	1179,77	-0,54	0,34	0,74	1165,88	-0,62	0,46	0,67	Hypnc35066	1	No	No
5	BL76585	NA	NA	NA	NA	243,43	-1,05	NA	0,48	333,93	-1,14	0,42	0,44	Hypp5787	1	No	No
5	BL87339	NA	NA	NA	NA	1839,84	-0,43	0,35	1,39	6936,60	-0,77	0,39	0,63	KIAA1161	0	No	No
5	BL88341	NA	NA	NA	NA	2918,83	-0,39	NA	0,58	7495,75	-0,62	0,29	0,59	FABP7	5	Yes	Yes
5	BL90924	NA	NA	NA	NA	318,72	-0,31	NA	0,54	337,38	-0,67	0,57	0,44	Hypnc51544	0	No	No
5	BL91554	14,01	NA	NA	0,33	75,55	-0,56	0,48	0,85	162,24	-0,65	0,48	0,40	Hypnc52098	0	No	No
5	BL95116	43,42	NA	NA	0,28	122,20	-1,01	NA	0,82	123,16	-1,41	0,54	0,82	Hypp8244	5	No	No
5	BL95144	NA	NA	NA	NA	748,10	-0,58	NA	0,60	897,50	-0,56	0,37	0,39	ENSG00000262057	3	No	No
5	BL95250	198,88	NA	NA	0,44	389,54	-0,33	NA	0,87	195,36	-0,67	NA	0,70	CYP24A1	4	No	No
5	BL95251	1771,76	NA	NA	0,32	687,67	-0,29	NA	0,75	296,63	-0,73	NA	0,65	GGT1	3	No	No
5	BL95310	NA	NA	NA	NA	175,78	-0,53	NA	0,88	158,74	-1,11	NA	0,92	Hypp11786	1	No	No
5	BL95311	NA	NA	NA	NA	567,78	-0,58	NA	0,59	418,01	-1,12	NA	0,52	Hypp11787	1	No	No
5	BL95834	NA	NA	NA	NA	198,12	-0,84	NA	1,28	164,86	-0,88	0,57	0,73	TRIM3	2	No	No
5	BL96819	NA	NA	NA	NA	412,97	-0,34	NA	0,48	799,24	-0,68	0,43	0,57	TLL1	1	No	No
6	BL00590	NA	NA	NA	NA	303,14	-0,51	NA	0,56	636,68	-0,90	NA	NA	TRIM45	1	No	No
6	BL00603	NA	NA	NA	NA	NA	NA	NA	NA	77,89	-0,69	NA	NA	KLHL21	3	No	No
6	BL00640	NA	NA	NA	NA	66,84	NA	NA	0,89	131,44	-0,58	NA	NA	CUBN	4	No	No
6	BL00828	NA	NA	NA	NA	4668,78	-0,45	NA	0,37	2747,31	-0,98	NA	NA	NOS1	2	No	No
6	BL02040	NA	NA	NA	NA	82,41	-0,59	NA	NA	72,31	-0,64	NA	NA	MFSD6	3	No	No
6	BL02065	NA	NA	NA	NA	67,90	NA	NA	0,47	101,15	-0,81	NA	NA	Hypp8346	0	No	No
6	BL02183	64,67	NA	NA	0,46	133,66	-0,57	NA	0,60	115,02	-0,59	NA	NA	IPP	5	No	Yes
6	BL02919	NA	NA	NA	NA	1415,68	-0,43	NA	0,21	1399,65	-0,86	NA	NA	LGMN	4	No	No
6	BL03138	NA	NA	NA	NA	NA	NA	NA	NA	568,78	-0,61	NA	NA	KDM1B	1	No	No
6	BL03237	NA	NA	NA	NA	NA	NA	NA	NA	60,72	-0,61	NA	NA	Hypp12861	4	Yes	No
6	BL03358	NA	NA	NA	NA	NA	NA	NA	NA	1111,43	-0,76	NA	NA	ENSG00000262499	2	No	No
6	BL03434	NA	NA	NA	NA	NA	NA	NA	NA	62,72	-0,70	NA	NA	CALM1	5	No	No
6	BL03933	NA	NA	NA	NA	163,96	-0,42	NA	0,35	145,21	-0,59	NA	NA	HEATR3	6	No	No
6	BL03989	NA	NA	NA	NA	106,95	NA	NA	0,93	118,81	-1,39	NA	NA	TRIM3	0	No	No
6	BL04237	NA	NA	NA	NA	232,43	NA	0,50	0,84	149,22	-1,34	NA	NA	Hypp12272	1	No	No
6	BL04340	NA	NA	NA	NA	NA	NA	NA	NA	152,95	-0,93	NA	NA	Hypp1037	0	No	No
6	BL04434	NA	NA	NA	NA	21,95	NA	NA	0,80	60,21	-1,21	0,79	NA	KY	1	No	No
6	BL04694	NA	NA	NA	NA	65,78	NA	NA	1,14	207,77	-0,55	0,56	NA	ABP1	1	No	No
6	BL04996	NA	NA	NA	NA	130,44	-0,52	NA	0,48	151,71	-0,75	NA	NA	GSTK1	0	No	No
6	BL05191	NA	NA	NA	NA	NA	NA	NA	NA	78,25	-0,70	NA	NA	SLC5A9	0	No	No
6	BL06224	NA	NA	NA	NA	1813,76	-0,33	NA	NA	749,76	-0,69	NA	NA	CALM1	0	No	No
6	BL06226	NA	NA	NA	NA	146,08	-0,57	NA	0,64	108,61	-0,88	NA	NA	Hypp9046	0	No	No
6	BL06710	150,84	NA	NA	0,29	120,53	-0,49	NA	0,39	99,81	-0,75	NA	NA	PIGV	1	No	No
6	BL07347	36,76	NA	NA	0,29	167,35	-0,44	NA	0,68	233,64	-0,67	NA	NA	GAL3ST1	4	No	No
6	BL07943	NA	NA	NA	NA	NA	NA	NA	NA	1643,57	-0,88	NA	NA	FOSL1	0	No	No
6	BL08073	NA	NA	NA	NA	NA	NA	NA	NA	100,53	-0,75	NA	NA	SLC22A5	2	No	No
6	BL08390	NA	NA	NA	NA	NA	NA	NA	NA	56,14	-0,59	NA	NA	GALNS	1	No	No
6	BL08632	NA	NA	NA	NA	19,44	NA	NA	0,53	98,04	-1,06	NA	NA	KCNK16	1	No	No
6	BL09017	NA	NA	NA	NA	4430,79	-0,29	NA	NA	3151,31	-0,66	NA	NA	SOD1	1	No	No
6	BL09144	104,37	NA	NA	0,37	NA	NA	NA	NA	98,71	-0,71	NA	NA	Hypp3592	4	No	No
6	BL09675	NA	NA	NA	NA	232,86	-0,34	NA	0,72	191,73	-0,88	NA	NA	AKR1B10	2	No	No
6	BL09721	NA	NA	NA	NA	1362,12	-0,27	0,28	0,79	1415,37	-0,73	0,49	NA	MEIS1	0	No	No
6	BL09866	NA	NA	NA	NA	1357,55	-0,20	NA	NA	612,28	-0,59	NA	NA	Hypp10182	1	No	No
6	BL10351	NA	NA	NA	NA	NA	NA	NA	NA	50,16	-0,81	NA	NA	Hypp1894	2	No	No
6	BL10400	NA	NA	NA	NA	1439,87	NA	NA	0,69	8325,65	-0,45	0,34	NA	COL4A6	2	No	No
6	BL10581	NA	NA	NA	NA	38,28	NA	NA	1,14	68,10	-0,94	NA	NA	ATOH1	0	Yes	No
6	BL10736	NA	NA	NA	NA	277,29	NA	NA	0,38	244,79	-0,60	NA	NA	CRYL1	2	No	No
6	BL10780	NA	NA	NA	NA	NA	NA	NA	NA	811,57	-0,79	NA	NA	TGM1	3	No	No
6	BL10841	NA	NA	NA	NA	13,55	NA	NA	0,62	67,73	-1,23	0,82	NA	GUCY1B3	1	No	No
6	BL10881	NA	NA	NA	NA	49,80	NA	NA	0,79	68,99	-0,81	NA	NA	TRPC5	1	No	No
6	BL11110	NA	NA	NA	NA	127,88	-0,49	NA	0,37	144,08	-0,80	NA	NA	ATG10	4	No	No
6	BL11229	NA	NA	NA	NA	60,12	-0,51	NA	NA	69,79	-0,72	NA	NA	Hypp1316	0	No	No
6	BL11395	NA	NA	NA	NA	84,94	-0,80	NA	0,42	60,96	-0,98	NA	NA	SLC7A1	2	No	No
6	BL11629	NA	NA	NA	NA	558,13	-0,30	NA	0,23	476,96	-0,63	NA	NA	MPST	2	No	No
6	BL11734	NA	NA	NA	NA	31,37	NA	NA	0,76	98,96	-1,17	NA	NA	CYP2J2	4	No	No
6	BL11912	NA	NA	NA	NA	NA	NA	NA	NA	66,79	-0,65	NA	NA	NPSR1	4	No	No
6	BL12193	NA	NA	NA	NA	118,45	NA	NA	0,56	252,39	-1,10	NA	NA	EMX1	1	No	No
6	BL12229	NA	NA	NA	NA	NA	NA	NA	NA	1419,52	-0,64	NA	NA	Hypp12954	4	No	No
6	BL12546	NA	NA	NA	NA	73,15	NA	NA	0,51	50,99	-1,11	NA	NA	Hypp10036	0	No	No
6	BL13532	NA	NA	NA	NA	NA	NA	NA	NA	385,40	-0,60	NA	NA	Hypp11111	1	No	No
6	BL13559	NA	NA	NA	NA	156,21	-0,53	NA	0,44	139,60	-0,68	NA	NA	RNF160	0	No	No
6	BL13568	NA	NA	NA	NA	593,04	NA	NA	0,48	488,74	-0,62	NA	NA	SLC6A9	4	No	No
6	BL13583	NA	NA	NA	NA	NA	NA	NA	NA	309,60	-0,59	NA	NA	Hypp10044	0	No	No
6	BL13732	795,44	NA	NA	0,42	673,73	NA	0,36	0,60	341,40	-0,49	NA	NA	ASCL1	1	No	No
6	BL13938	NA	NA	NA	NA	881,55	NA	NA	0,48	946,46	-0,59	NA	NA	ASNS	1	No	No
6	BL14117	NA	NA	NA	NA	NA	NA	NA	NA	71,84	-0,65	NA	NA	PNMT	0	No	No
6	BL14410	NA	NA	NA	NA	NA	NA	NA	NA	77,78	-0,66	NA	NA	SAT1	0	No	No
6	BL14697	NA	NA	NA	NA	NA	NA	NA	NA	70,71	-0,64	NA	NA	LPGAT1	0	No	No
6	BL15472	NA	NA	NA	NA	74,10	-0,48	NA	0,64	116,61	-0,81	NA	NA	SYT15	2	No	No
6	BL15525	NA	NA	NA	NA	193,56	NA	NA	0,56	216,65	-0,66	NA	NA	CEPT1	2	No	Yes
6	BL15611	NA	NA	NA	NA	207,45	-0,52	NA	0,37	163,51	-0,59	NA	NA	Hypp2363	0	No	No
6	BL15671	142,96	NA	NA	0,41	326,08	-0,88	NA	0,48	304,03	-0,90	NA	NA	ABHD12	1	No	No
6	BL15691	63,45	NA	NA	0,28	NA	NA	NA	NA	67,62	-0,64	NA	NA	HSD11B1	1	No	No
6	BL15763	NA	NA	NA	NA	341,37	NA	NA	0,32	213,63	-0,63	NA	NA	Hypp14921	3	No	No
6	BL16462	904,94	NA	NA	0,53	2688,97	-0,67	0,24	0,82	2348,89	-0,55	NA	NA	SLC26A11	2	No	No
6	BL16535	NA	NA	NA	NA	NA	NA	NA	NA	96,19	-0,66	NA	NA	SLC2A12	0	Yes	No
6	BL16728	NA	NA	NA	NA	383,05	NA	NA	0,42	210,33	-0,63	NA	NA	TRPM1	0	No	No
6	BL16779	NA	NA	NA	NA	274,95	NA	NA	0,61	210,02	-0,70	NA	NA	SLC24A5	1	No	No
6	BL17162	NA	NA	NA	NA	NA	NA	NA	NA	413,57	-0,66	NA	NA	RFX6	2	Yes	No
6	BL17371	NA	NA	NA	NA	115,27	NA	NA	0,62	135,28	-0,74	NA	NA	ELOVL6	2	No	No
6	BL17481	NA	NA	NA	NA	NA	NA	NA	NA	190,69	-0,72	NA	NA	MFSD2A	2	No	No
6	BL17516	NA	NA	NA	NA	13,93	NA	NA	0,47	65,10	-1,06	NA	NA	CYP2U1	0	No	No
6	BL17719	NA	NA	NA	NA	302,84	-0,62	NA	0,30	391,15	-0,65	NA	NA	SLC43A3	0	No	No
6	BL17785	NA	NA	NA	NA	1622,10	NA	0,27	0,64	2005,75	-0,34	0,34	NA	Hypp5734	0	No	No
6	BL17872	NA	NA	NA	NA	44,66	NA	NA	0,47	56,48	-0,84	NA	NA	C14orf159	1	No	No
6	BL17929	NA	NA	NA	NA	216,54	NA	NA	0,37	143,14	-0,90	NA	NA	PCSK9	4	No	No
6	BL18315	NA	NA	NA	NA	528,25	-0,34	NA	0,26	422,90	-0,59	NA	NA	Hypp2353	0	No	No
6	BL18336	NA	NA	NA	NA	NA	NA	NA	NA	115,98	-0,84	NA	NA	LAMC1	1	No	No
6	BL19675	NA	NA	NA	NA	4316,71	-0,20	NA	0,37	3362,69	-0,67	NA	NA	KRT84	3	No	No
6	BL19936	NA	NA	NA	NA	91,11	-0,54	NA	0,91	74,93	-0,81	NA	NA	KCNK4	0	No	No

		N0				N3				N4							
Cluster	Gene	bM	I2FC BMS493	I2FC R115866	I2FC RA	bM	I2FC BMS493	I2FC R115866	I2FC RA	bM	I2FC BMS4934	I2FC R1158667	I2FC RA	Auto nomenclature	N.º of RAREs	ATACseq peaks	RARE hubs
6	BL19995	NA	NA	NA	NA	927,28	-0,61	NA	0,50	902,83	-0,67	NA	NA	RNF160	2	No	No
6	BL20824	NA	NA	NA	NA	NA	NA	NA	NA	51,03	-0,70	NA	NA	Hypp8452	0	No	No
6	BL20861	NA	NA	NA	NA	NA	NA	NA	NA	111,25	-0,86	NA	NA	TLR6	1	No	No
6	BL21000	NA	NA	NA	NA	NA	NA	NA	NA	50,70	-0,61	NA	NA	ATHL1	2	No	No
6	BL21436	21,56	NA	NA	0,21	2474,86	NA	NA	0,77	9634,17	-0,49	NA	NA	ITGB1	4	No	No
6	BL21686	NA	NA	NA	NA	NA	NA	NA	NA	371,24	-0,60	NA	NA	Hypp12159	0	No	No
6	BL21957	NA	NA	NA	NA	NA	NA	NA	NA	274,46	-0,70	NA	NA	ACPP	1	No	No
6	BL22077	108,48	NA	NA	0,59	497,57	-0,66	NA	0,78	791,36	-0,38	0,44	NA	PTPRA	2	No	No
6	BL22199	NA	NA	NA	NA	333,93	-0,60	NA	0,35	434,12	-0,79	NA	NA	ABCC9	1	Yes	No
6	BL22559	NA	NA	NA	NA	14,08	NA	NA	0,51	138,01	-0,63	NA	NA	FLVCR1	1	No	No
6	BL22668	NA	NA	NA	NA	NA	NA	NA	NA	81,70	-0,69	NA	NA	SAT2	0	No	No
6	BL22731	NA	NA	NA	NA	NA	NA	NA	NA	443,57	-0,86	NA	NA	SRR	2	No	No
6	BL22836	NA	NA	NA	NA	65,83	-0,73	NA	NA	65,25	-0,64	NA	NA	EXD2	0	No	No
6	BL23239	NA	NA	NA	NA	NA	NA	NA	NA	75,50	-0,93	NA	NA	Hypp3774	5	No	No
6	BL23283	NA	NA	NA	NA	95,49	NA	NA	0,50	103,12	-0,83	NA	NA	CEPT1	0	No	No
6	BL23317	NA	NA	NA	NA	27,70	NA	NA	0,50	79,17	-0,83	0,54	NA	DTHD1	0	Yes	No
6	BL23364	NA	NA	NA	NA	NA	NA	NA	NA	598,12	-0,76	NA	NA	SREBF1	1	No	No
6	BL23378	NA	NA	NA	NA	795,53	NA	NA	0,22	399,80	-0,69	NA	NA	CYP24A1	0	No	No
6	BL23403	707,77	NA	NA	0,26	507,98	-0,60	NA	0,42	275,57	-0,66	NA	NA	TRIM3	2	No	No
6	BL23532	NA	NA	NA	NA	850,73	NA	NA	0,31	458,63	-0,59	NA	NA	Hypp6307	1	No	No
6	BL23630	NA	NA	NA	NA	NA	NA	NA	NA	121,78	-0,80	NA	NA	Hypp7183	2	No	No
6	BL24209	NA	NA	NA	NA	NA	NA	NA	NA	229,00	-0,61	NA	NA	REG1A	1	No	No
6	BL24238	NA	NA	NA	NA	163,12	-0,60	NA	0,35	179,73	-0,63	NA	NA	SLC43A3	1	No	No
6	BL24403	NA	NA	NA	NA	153,06	NA	NA	0,41	96,21	-0,69	NA	NA	Hypp16059	1	No	No
6	BL24685	385,75	NA	NA	0,34	714,45	-0,66	0,42	0,54	563,01	-0,47	NA	NA	ZNF292	2	No	No
6	BL24696	NA	NA	NA	NA	124,82	NA	0,46	NA	94,14	-0,61	NA	NA	LPA	0	No	No
6	BL24811	NA	NA	NA	NA	37,82	NA	NA	0,48	63,55	-0,63	NA	NA	GAL3ST1	3	No	No
6	BL31570	NA	NA	NA	NA	124,67	NA	NA	0,55	27,61	-0,99	NA	NA	Hypp8069	0	No	No
6	BL44872	NA	NA	NA	NA	131,30	-0,55	NA	0,66	104,47	-1,00	NA	NA	Hypnc11104	2	No	No
6	BL50074	NA	NA	NA	NA	NA	NA	NA	NA	52,98	-0,70	NA	NA	LRP1B	0	No	No
6	BL50758	NA	NA	NA	NA	NA	NA	NA	NA	65,97	-0,78	NA	NA	Hypnc16293	1	No	No
6	BL52464	NA	NA	NA	NA	228,41	-0,77	NA	0,64	243,93	-1,41	NA	NA	Hypp15310	0	No	No
6	BL68850	NA	NA	NA	NA	NA	NA	NA	NA	65,38	-0,62	NA	NA	Hypnc32187	0	No	No
6	BL80869	NA	NA	NA	NA	41,33	NA	NA	0,79	132,41	-0,66	NA	NA	CYP2U1	2	No	No
6	BL82595	NA	NA	NA	NA	NA	NA	NA	NA	148,51	-0,59	NA	NA	TLCD1	2	No	No
6	BL87757	NA	NA	NA	NA	NA	NA	NA	NA	87,56	-0,63	NA	NA	Hypnc48744	0	No	No
6	BL89509	NA	NA	NA	NA	NA	NA	NA	NA	94,37	-0,77	NA	NA	Hypnc50302	1	No	No
6	BL91557	NA	NA	NA	NA	147,40	-0,44	NA	0,41	136,41	-0,63	NA	NA	Hypnc52102	2	No	No
6	BL95046	NA	NA	NA	NA	201,84	-0,50	NA	0,61	142,76	-0,67	NA	NA	Hypp15468	3	No	No
6	BL95095	NA	NA	NA	NA	14382,37	NA	NA	0,33	7642,28	-0,63	NA	NA	TGM1	1	No	No
6	BL95223	NA	NA	NA	NA	NA	NA	NA	NA	92,30	-0,61	NA	NA	Hypp11037	0	No	No
6	BL95225	NA	NA	NA	NA	424,10	-0,28	NA	NA	177,76	-0,71	NA	NA	SYT3	1	No	No
6	BL95236	NA	NA	NA	NA	240,35	NA	NA	0,34	360,59	-0,63	NA	NA	Hypp11159	0	No	No
6	BL95263	NA	NA	NA	NA	197,89	NA	NA	0,76	227,64	-0,79	NA	NA	Hypp11266	3	No	No
6	BL95286	NA	NA	NA	NA	NA	NA	NA	NA	1545,85	-0,61	NA	NA	SLC22A5	0	No	No
6	BL95455	NA	NA	NA	NA	NA	NA	NA	NA	168,11	-0,59	NA	NA	HSD11B1	13	No	No
6	BL95479	NA	NA	NA	NA	201,97	NA	NA	0,40	178,74	-0,61	NA	NA	RDH12	0	No	No
6	BL95532	NA	NA	NA	NA	88,45	NA	NA	0,81	139,44	-0,82	0,51	NA	Hypp14324	2	No	No
6	BL95657	NA	NA	NA	NA	124,19	NA	NA	0,51	96,56	-0,85	NA	NA	HNMT	4	No	No
6	BL96351	NA	NA	NA	NA	NA	NA	NA	NA	71,81	-0,64	NA	NA	NOTCH1	0	No	No
6	BL96936	NA	NA	NA	NA	NA	NA	NA	NA	1312,20	-0,75	NA	NA	Hypp2417	0	No	No
6	BL96937	NA	NA	NA	NA	NA	NA	NA	NA	1045,21	-0,62	NA	NA	ZNF512B	1	No	No
6	BL97121	NA	NA	NA	NA	1126,58	NA	NA	0,68	2887,63	-0,74	NA	NA	VIM	2	No	No
7	BL00631	1477,60	NA	NA	0,40	5894,45	-0,31	NA	0,63	8020,91	NA	0,50	0,40	CTH	2	No	No
7	BL01248	NA	NA	NA	NA	619,09	NA	NA	0,48	727,98	NA	0,50	0,59	Hypp15822	0	No	No
7	BL01298	NA	NA	NA	NA	540,99	NA	NA	0,63	720,96	NA	0,42	0,44	Hypp10268	1	No	No
7	BL01348	NA	NA	NA	NA	115,73	NA	NA	0,61	NA	NA	NA	NA	COL12A1	2	No	No
7	BL01746	NA	NA	NA	NA	152,68	-0,87	NA	NA	NA	NA	NA	NA	Hypp110	2	No	No
7	BL01954	11,41	NA	NA	0,17	376,79	0,36	NA	1,52	1537,42	NA	NA	0,41	PLA2G3	1	No	No
7	BL02504	416,29	NA	NA	0,29	1138,07	-0,48	NA	0,70	1363,14	NA	0,79	0,62	GALNT11	3	Yes	No
7	BL02978	NA	NA	NA	NA	54,21	-0,48	NA	0,76	117,69	NA	NA	0,81	Hypp763	1	No	No
7	BL04012	NA	NA	NA	NA	387,64	NA	NA	0,74	NA	NA	NA	NA	LAMA3	0	Yes	No
7	BL04225	NA	NA	NA	NA	164,31	NA	NA	0,61	149,50	NA	NA	0,87	Hypp1430	0	No	No
7	BL04332	NA	NA	NA	NA	308,42	NA	NA	0,59	335,18	NA	0,38	0,53	SLC39A11	0	No	No
7	BL04479	NA	NA	NA	NA	1031,78	-0,41	NA	0,61	NA	NA	NA	NA	KIAA0664	4	No	No
7	BL04803	NA	NA	NA	NA	688,36	-0,26	0,28	0,98	653,51	NA	0,68	0,59	Hypp3313	6	No	Yes
7	BL05270	NA	NA	NA	NA	52,87	NA	NA	1,15	556,06	NA	NA	0,39	ERG	2	No	No
7	BL05279	NA	NA	NA	NA	80,43	NA	NA	0,61	NA	NA	NA	NA	SLC5A4	1	No	No
7	BL05319	NA	NA	NA	NA	958,52	NA	NA	0,45	1341,28	NA	0,96	0,72	AACS	0	No	No
7	BL05330	523,84	NA	NA	0,62	2018,96	-0,32	NA	NA	NA	NA	NA	NA	WNT5A	4	No	Yes
7	BL05413	NA	NA	NA	NA	342,05	-0,32	NA	0,28	502,06	NA	0,61	NA	NR1H4	0	No	No
7	BL06127	NA	NA	NA	NA	144,87	NA	NA	0,67	NA	NA	NA	NA	SYT17	1	No	No
7	BL06700	NA	NA	NA	NA	423,59	NA	0,46	0,64	1093,81	-0,23	NA	NA	CCDC48	1	No	No
7	BL06701	NA	NA	NA	NA	83,35	NA	NA	0,81	NA	NA	NA	NA	GAL3ST1	0	No	No
7	BL06719	NA	NA	NA	NA	469,04	NA	NA	1,24	NA	NA	NA	NA	KIRREL3	0	Yes	No
7	BL07482	NA	NA	NA	NA	45,65	NA	NA	0,74	61,93	NA	NA	0,62	RDH12	2	No	No
7	BL07555	NA	NA	NA	NA	257,66	NA	NA	0,43	278,95	NA	0,66	NA	HAL	0	No	No
7	BL07794	NA	NA	NA	NA	441,50	NA	NA	0,78	NA	NA	NA	NA	FASN	0	No	No
7	BL09095	NA	NA	NA	NA	123,76	NA	NA	0,40	173,55	NA	0,79	0,67	ETHE1	2	No	No
7	BL09217	NA	NA	NA	NA	319,62	NA	NA	0,70	NA	NA	NA	NA	CORIN	0	No	No
7	BL09220	NA	NA	NA	NA	53,55	NA	NA	0,72	85,64	NA	1,09	NA	CORIN	0	No	No
7	BL09453	NA	NA	NA	NA	415,18	-0,28	NA	0,32	679,65	NA	0,62	NA	NAALAD2	1	No	No
7	BL09923	NA	NA	NA	NA	260,37	NA	NA	0,59	511,21	NA	0,50	0,58	SCP2	1	No	No
7	BL10022	NA	NA	NA	NA	59,60	NA	NA	0,61	214,94	NA	0,44	0,64	LAMC1	2	No	No
7	BL10266	NA	NA	NA	NA	313,91	NA	NA	0,86	1666,61	NA	0,41	0,43	RUNX1	0	No	No
7	BL10312	NA	NA	NA	NA	62,23	NA	NA	0,70	NA	NA	NA	NA	APLF	1	No	No
7	BL10529	NA	NA	NA	NA	585,92	NA	NA	0,63	NA	NA	NA	NA	LAMA2	5	No	No
7	BL10722	NA	NA	NA	NA	742,41	0,32	0,80	0,30	NA	NA	NA	NA	FBN2	1	No	No
7	BL11231	NA	NA	NA	NA	339,83	NA	NA	0,59	NA	NA	NA	NA	TIPARP	0	No	No
7	BL11326	NA	NA	NA	NA	99,56	NA	NA	0,87	129,34	NA	0,62	NA	STS	3	No	No
7	BL11413	627,34	NA	NA	0,69	1393,23	-0,58	0,22	0,43	NA	NA	NA	NA	PDX1 xlox	1	No	No
7	BL11532	NA	NA	NA	NA	112,69	NA	NA	0,44	202,12	NA	NA	0,64	DDT	2	No	No
7	BL11761	NA	NA	NA	NA	1888,24	NA	0,59	NA	NA	NA	NA	NA	Hypp3017	3	No	No
7	BL12158	NA	NA	NA	NA	673,63	-0,41	0,28	0,50	665,16	NA	0,68	NA	C1orf158	2	No	Yes
7	BL13032	NA	NA	NA	NA	80,13	NA	NA	1,19	157,68	NA	NA	0,66	SLC7A7	0	No	No
7	BL13088	NA	NA	NA	NA	97,90	NA	NA	0,70	NA	NA	NA	NA	FHDC1	1	No	No
7	BL13827	NA	NA	NA	NA	100,20	NA	0,62	0,58	NA	NA	NA	NA	Hypp10528	0	No	No
7	BL14376	NA	NA	NA	NA	68,94	NA	NA	0,62	99,51	NA	NA	0,57	SLC5A4	0	No	No
7	BL14635	NA	NA	NA	NA	90,97	NA	NA	0,80	NA	NA	NA	NA	HPGD	1	No	No
7	BL14843	NA	NA	NA	NA	4192,86	NA	0,60	NA	NA	NA	NA	NA	Hypp6066	2	No	No
7	BL15170	NA	NA	NA	NA	483,10	NA	NA	0,53	640,00	NA	0,87	NA	FADS2	1	No	No
7	BL15401	NA	NA	NA	NA	54,62	NA	NA	1,13	NA	NA	NA	NA	CNTN4	1	No	No
7	BL15612	920,08	NA	NA	0,72	1364,83	-0,24	NA	NA	NA	NA	NA	NA	CCM2	3	No	Yes
7	BL16100	NA	NA	NA	NA	194,71	NA	NA	0,63	NA	NA	NA	NA	C1QTNF5	1	No	No
7	BL16742	320,23	-0,33	NA	0,41	534,85	-0,63	NA	0,47	NA	NA	NA	NA	ST8SIA4	2	No	No
7	BL17153	47,62	NA	NA	0,38	330,58	NA	NA	0,76	486,28	NA	NA	0,42	CYP2U1	1	No	No
7	BL17158	NA	NA	NA	NA	125,45	NA	0,51	0,85	NA	NA	NA	NA	LRP6	1	No	No
7	BL17201	NA	NA	NA	NA	267,49	-0,33	NA	0,72	247,53	NA	NA	0,42	ZFP112	1	No	No
7	BL17224	NA	NA	NA	NA	328,52	NA	0,35	0,94	713,76	NA	NA	0,56	Hypp6119	1	No	No
7	BL17248	NA	NA	NA	NA	61,39	NA	NA	0,95	NA	NA	NA	NA	ENTPD1	2	No	No

		N0				N3				N4						
Cluster	Gene	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS4934	l2FC R1158667	l2FC RA	Auto nomenclature	N.º of RAREs	ATACseq peaks	RARE hubs
7	BL17695	NA	NA	NA	NA	201,96	NA	NA	0,49	266,91	NA	NA	0,60	Hypp11148	5	No	No
7	BL17829	914,54	-0,30	NA	NA	2181,95	-0,68	NA	0,19	2334,30	NA	NA	0,40	TTPA	2	No	Yes
7	BL18028	NA	NA	NA	NA	369,05	-0,43	NA	0,40	669,32	NA	0,80	0,75	SLC26A11	2	No	No
7	BL18031	NA	NA	NA	NA	343,39	NA	NA	0,36	480,27	NA	0,78	NA	TTLL2	1	No	No
7	BL18090	NA	NA	NA	NA	56,01	NA	NA	0,61	NA	NA	NA	NA	NPHP3	2	No	No
7	BL18352	390,35	NA	NA	0,34	313,31	-0,43	NA	0,44	250,56	NA	0,94	0,87	Hypp14671	3	No	No
7	BL18912	NA	NA	NA	NA	15,14	NA	NA	0,48	61,23	NA	0,77	NA	CRYGA	0	No	No
7	BL19685	NA	NA	NA	NA	1465,89	-0,20	0,33	0,62	NA	NA	NA	NA	SLC4A11	0	No	No
7	BL19904	NA	NA	NA	NA	236,41	NA	NA	0,37	343,41	NA	0,74	0,54	Hypp11980	0	No	No
7	BL19972	NA	NA	NA	NA	172,94	NA	0,60	0,40	NA	NA	NA	NA	Hypp10531	0	No	No
7	BL19992	NA	NA	NA	NA	576,85	NA	NA	0,78	NA	NA	NA	NA	TGFBI	6	No	No
7	BL20224	NA	NA	NA	NA	216,28	NA	NA	0,76	NA	NA	NA	NA	ITGB6	2	No	No
7	BL20249	NA	NA	NA	NA	345,26	NA	NA	0,99	NA	NA	NA	NA	SDK2	4	No	No
7	BL20307	NA	NA	NA	NA	133,59	NA	NA	1,50	1364,04	-0,52	NA	NA	Hypp13984	1	No	No
7	BL21005	67,07	NA	NA	0,30	187,76	NA	NA	0,89	213,63	NA	NA	0,59	STS	3	No	No
7	BL21017	NA	NA	NA	NA	74,11	NA	NA	0,61	115,71	NA	0,49	0,56	Hypp2751	2	No	No
7	BL21161	NA	NA	NA	NA	13,03	NA	NA	0,51	108,93	NA	0,49	0,68	Hypp1741	3	No	No
7	BL21234	NA	NA	NA	NA	91,17	NA	NA	0,46	141,29	NA	NA	0,83	CTSL2	1	No	No
7	BL22738	NA	NA	NA	NA	1300,04	NA	NA	0,35	1486,77	NA	NA	0,66	Hypp11034	2	No	No
7	BL23103	NA	NA	NA	NA	82,40	NA	NA	0,62	NA	NA	NA	NA	HPRT1	1	No	No
7	BL23196	NA	NA	NA	NA	48,98	NA	NA	0,59	87,42	NA	NA	0,99	APOB	0	No	No
7	BL23212	NA	NA	NA	NA	217,74	-0,40	NA	0,38	270,37	NA	0,61	0,70	Hypp14261	3	No	Yes
7	BL23257	NA	NA	NA	NA	133,22	NA	NA	0,46	219,82	NA	0,60	NA	PPP1R3D	1	No	No
7	BL23304	NA	NA	NA	NA	402,77	NA	NA	0,33	561,75	NA	NA	0,64	COMTD1	0	No	No
7	BL23801	NA	NA	NA	NA	280,85	NA	NA	0,75	257,58	NA	1,00	NA	ABCA3	9	No	No
7	BL24542	854,99	NA	NA	0,36	2519,07	-0,18	NA	0,53	3123,98	NA	0,84	0,59	ENSG00000269087	3	No	Yes
7	BL24602	NA	NA	NA	NA	50,24	NA	NA	0,59	NA	NA	NA	NA	CASR	4	No	No
7	BL25717	NA	NA	NA	NA	168,77	NA	NA	0,75	324,36	NA	0,39	0,51	C7orf52	0	No	No
7	BL39921	NA	NA	NA	NA	83,71	-0,59	NA	NA	86,20	NA	NA	0,75	RNF217	2	No	Yes
7	BL44051	NA	NA	NA	NA	206,98	NA	NA	0,40	186,51	NA	NA	0,59	CYP2D6	0	No	No
7	BL48111	NA	NA	NA	NA	159,75	NA	NA	0,79	250,04	NA	0,59	0,69	Hypnc13948	0	No	No
7	BL54723	NA	NA	NA	NA	44,42	NA	NA	0,77	164,12	NA	0,57	0,66	ANGPTL7	2	No	No
7	BL56270	NA	NA	NA	NA	79,53	NA	NA	0,48	159,17	NA	0,52	0,59	Hypp6143	2	No	Yes
7	BL56998	NA	NA	NA	NA	733,92	NA	NA	0,28	970,51	NA	0,59	0,68	Hypnc21792	1	No	No
7	BL58899	NA	NA	NA	NA	87,64	-0,59	NA	NA	NA	NA	NA	NA	Hypp16191	1	No	No
7	BL64366	NA	NA	NA	NA	391,59	NA	NA	1,04	1013,32	NA	0,52	0,57	ENSG00000262448	1	No	No
7	BL65803	NA	NA	NA	NA	73,94	-0,61	NA	0,45	NA	NA	NA	NA	Hypp14082	3	Yes	No
7	BL72862	NA	NA	NA	NA	65,80	NA	NA	0,76	105,59	NA	0,77	NA	FZD10	0	No	No
7	BL81245	NA	NA	NA	NA	114,22	-0,61	NA	NA	NA	NA	NA	NA	Hypp7080	0	No	No
7	BL85417	NA	NA	NA	NA	125,88	-0,50	NA	1,19	623,77	-0,41	NA	NA	TMEM22	0	No	No
7	BL95179	NA	NA	NA	NA	41,93	NA	NA	0,75	50,44	NA	0,66	NA	CYP2U1	0	No	No
7	BL95186	499,24	NA	NA	0,62	1526,60	-0,46	0,23	0,59	1393,19	NA	0,40	NA	ABCD3	3	No	No
7	BL95214	40,29	NA	NA	0,25	50,45	NA	NA	1,50	64,83	NA	NA	0,72	ARSB	1	No	No
7	BL95249	NA	NA	NA	NA	75,71	NA	NA	0,76	NA	NA	NA	NA	CYP24A1	4	No	Yes
7	BL95282	NA	NA	NA	NA	11410,77	-0,67	NA	0,49	NA	NA	NA	NA	UXS1	5	No	No
7	BL95391	NA	NA	NA	NA	220,72	-0,34	NA	0,68	185,94	NA	0,57	NA	THRB	0	No	No
7	BL95452	NA	NA	NA	NA	54,40	NA	NA	0,71	137,55	NA	0,58	NA	CES2	2	No	No
7	BL95673	NA	NA	NA	NA	98,88	NA	NA	0,58	254,63	NA	0,88	0,74	RDH13	1	No	No
7	BL95813	NA	NA	NA	NA	270,46	-0,42	NA	0,62	951,07	NA	0,59	0,58	SLC31A1	1	No	No
7	BL95849	NA	NA	NA	NA	58,84	NA	NA	0,80	112,98	NA	NA	0,98	TGM1	1	No	No
7	BL96760	NA	NA	NA	NA	104,47	NA	0,60	NA	NA	NA	NA	NA	FBN3	0	No	No
7	BL97227	NA	NA	NA	NA	282,81	NA	NA	0,37	300,17	NA	0,66	NA	NPHP3	0	No	No
8	BL02114	NA	NA	NA	NA	25,43	NA	NA	1,79	312,32	-2,86	0,76	2,04	Hypp5922	1	No	No
8	BL03591	NA	NA	NA	NA	76,32	NA	NA	1,10	587,82	-3,31	0,92	1,45	CILP	0	No	No
8	BL06315	11,14	NA	NA	0,21	78,41	NA	NA	0,68	155,11	-2,14	1,04	1,54	HSD17B8	2	Yes	No
8	BL07789	NA	NA	NA	NA	122,27	NA	NA	1,45	1358,96	-1,47	0,70	1,27	LRRC70	3	No	No
8	BL09967	NA	NA	NA	NA	12,57	NA	NA	1,06	63,81	-1,77	0,82	1,92	NAALAD2	0	No	No
8	BL10499	NA	NA	NA	NA	876,34	NA	NA	0,57	2318,69	-1,37	0,49	1,58	Hypp1260	1	No	No
8	BL11869	NA	NA	NA	NA	261,86	-1,24	NA	1,72	650,94	-2,20	0,94	1,02	ST6GALNAC2	1	No	No
8	BL12058	NA	NA	NA	NA	30,90	NA	NA	1,40	153,38	-1,55	0,80	1,11	C6orf27	3	No	No
8	BL12294	NA	NA	NA	NA	37,81	NA	NA	0,81	374,92	-1,69	0,62	0,77	EML1	0	No	No
8	BL12569	NA	NA	NA	NA	37,97	NA	NA	0,89	96,18	-1,83	0,72	0,96	BCHE	2	No	No
8	BL12698	NA	NA	NA	NA	255,38	NA	NA	0,90	245,89	-1,67	0,87	2,02	GANAB	2	No	Yes
8	BL12777	47,32	NA	NA	0,30	677,09	-0,82	NA	1,35	1314,58	-2,03	0,59	1,18	SULT1C4	3	No	No
8	BL13081	NA	NA	NA	NA	30,21	NA	NA	1,41	106,98	-2,08	0,61	1,10	Hypp5438	2	Yes	No
8	BL17419	NA	NA	NA	NA	195,39	-0,90	NA	1,83	980,77	-1,93	0,78	1,20	Hypp2495	1	No	No
8	BL17836	NA	NA	NA	NA	12,10	NA	NA	0,87	153,21	-2,60	0,58	1,56	DNASE2	1	No	No
8	BL17910	NA	NA	NA	NA	44,24	NA	NA	0,72	238,43	-1,63	0,46	0,93	ENSG00000262217	0	No	No
8	BL21445	NA	NA	NA	NA	44,06	NA	NA	0,76	59,82	-1,60	0,99	1,52	SLCO3A1	0	No	No
8	BL21607	NA	NA	NA	NA	40,18	NA	NA	1,57	142,89	-2,67	0,94	1,74	HEBP2	2	No	No
8	BL22082	42,73	NA	NA	0,22	379,48	-0,80	NA	1,62	832,59	-2,80	1,21	1,68	HSD17B8	2	Yes	No
8	BL22303	NA	NA	NA	NA	31,44	NA	NA	0,80	406,81	-1,43	0,64	1,22	Hypp6038 (TSPAN family)	3	Yes	No
8	BL22640	NA	NA	NA	NA	28,05	NA	NA	1,08	147,56	-2,13	0,74	1,09	SLC5A9	5	No	No
8	BL24603	NA	NA	NA	NA	34,08	NA	NA	1,67	975,21	-2,13	0,83	1,75	NLGN3	2	No	No
8	BL59603	NA	NA	NA	NA	34,73	NA	NA	0,90	198,12	-2,11	NA	1,26	Hypp10502	1	No	No
8	BL86446	NA	NA	NA	NA	113,22	-1,12	NA	1,37	164,20	-2,39	0,87	1,45	Hypp13803	1	No	No
8	BL95461	NA	NA	NA	NA	58,00	-0,48	NA	1,68	238,25	-2,65	1,05	0,98	OSR2	2	Yes	No
8	BL96168	NA	NA	NA	NA	104,07	NA	NA	1,62	204,61	-1,68	0,81	1,02	ALDH1A1	2	No	No
8	BL96279	NA	NA	NA	NA	88,45	NA	NA	1,69	183,70	-1,60	0,62	1,21	EDIL3	0	No	No
8	BL96280	NA	NA	NA	NA	55,28	-0,56	NA	1,33	115,71	-1,63	0,68	1,69	CNTNAP5	1	No	No
8	BL97481	NA	NA	NA	NA	17,78	NA	NA	1,15	217,43	-2,40	0,65	1,62	SI	6	No	No
9	BL00171	NA	NA	NA	NA	163,28	NA	NA	0,44	64,48	-0,65	NA	0,77	Hypp5471	1	No	No
9	BL00319	NA	NA	NA	NA	227,66	NA	NA	0,35	561,40	-0,43	0,45	0,76	CROT	0	No	No
9	BL01038	NA	NA	NA	NA	76,96	NA	NA	0,87	89,20	-0,92	NA	0,67	ANKH	2	No	No
9	BL01368	NA	NA	NA	NA	255,99	NA	NA	0,45	111,55	-1,07	NA	0,63	SLC17A7	2	No	No
9	BL01921	NA	NA	NA	NA	NA	NA	NA	NA	154,66	-0,58	NA	0,74	FAM151B	0	No	No
9	BL02221	NA	NA	NA	NA	2210,32	NA	NA	0,44	1387,23	-0,88	NA	0,56	GGT1	3	No	No
9	BL02223	NA	NA	NA	NA	61,29	NA	NA	0,77	151,70	-0,75	NA	0,45	PARP12	2	No	No
9	BL02248	NA	NA	NA	NA	104,55	NA	NA	0,65	173,32	-0,78	0,67	0,97	PPIA	7	Yes	No
9	BL02369	NA	NA	NA	NA	105,29	NA	NA	0,53	298,04	-0,68	NA	0,99	Hypp13088	0	No	No
9	BL02568	NA	NA	NA	NA	105,98	NA	NA	0,58	708,40	-0,45	NA	0,79	NMRAL1	1	No	No
9	BL02713	NA	NA	NA	NA	NA	NA	NA	NA	96,67	-0,68	NA	1,05	IL1RAPL1	0	No	No
9	BL03007	NA	NA	NA	NA	166,97	NA	NA	0,98	492,45	-0,71	0,42	0,42	KCTD21	5	No	No
9	BL03396	NA	NA	NA	NA	40,03	NA	NA	0,69	205,05	-0,67	0,59	1,14	GDA	1	No	No
9	BL03570	NA	NA	NA	NA	7971,06	-0,19	0,41	0,74	9243,73	-0,76	0,75	0,66	MEIS1	4	No	No
9	BL03984	NA	NA	NA	NA	141,18	NA	NA	0,89	126,14	-0,76	NA	0,87	Hypp15825	1	No	No
9	BL04504	NA	NA	NA	NA	388,88	NA	NA	0,93	551,07	-0,63	0,55	0,88	SLC47A1	4	No	No
9	BL04730	NA	NA	NA	NA	88,09	NA	NA	1,00	434,35	-0,53	0,59	0,53	SHOX	0	Yes	No
9	BL05110	NA	NA	NA	NA	23,52	NA	NA	0,46	98,55	-0,83	NA	0,95	CREG1	2	No	No
9	BL05228	NA	NA	NA	NA	NA	NA	NA	NA	75,69	-1,64	NA	1,45	NLGN4Y	1	No	No
9	BL05339	NA	NA	NA	NA	1639,04	-0,22	NA	0,72	2050,34	-0,48	0,42	0,65	FADS1	1	Yes	No
9	BL05797	NA	NA	NA	NA	NA	NA	NA	NA	132,48	-0,76	0,51	0,85	SULT2A1	1	No	No
9	BL06383	NA	NA	NA	NA	1374,72	-0,21	NA	0,90	2665,59	-0,45	0,22	0,46	MMP24	0	Yes	No
9	BL06600	NA	NA	NA	NA	57,21	NA	NA	0,89	187,60	-1,01	0,76	1,01	SLCO3A1	3	No	No
9	BL07020	NA	NA	NA	NA	88,14	NA	NA	0,76	707,90	-0,74	NA	0,42	Hypp10247	2	No	No
9	BL07104	NA	NA	NA	NA	48,02	NA	NA	0,86	53,37	-0,88	1,31	1,07	Hypp4830	0	No	No
9	BL07242	NA	NA	NA	NA	98,95	NA	NA	0,74	145,17	-0,99	0,54	0,84	HSD11B1	1	No	No
9	BL07273	NA	NA	NA	NA	129,68	NA	NA	0,74	59,18	-1,42	NA	0,69	SLC22A3	1	No	No
9	BL07485	NA	NA	NA	NA	NA	NA	NA	NA	82,50	-0,84	NA	0,71	Hypp10844	0	No	No
9	BL07563	NA	NA	NA	NA	1880,08	NA	NA	0,41	1223,23	-0,57	0,51	0,71	Hypp1797	1	No	No

		N0				N3				N4							
Cluster	Gene	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS4934	l2FC R1158667	l2FC RA	Auto nomenclature	N.º of RAREs	ATACseq peaks	RARE hubs
9	BL07918	NA	NA	NA	NA	210,33	NA	NA	0,47	438,72	-0,41	0,41	0,66	DHRS4	2	No	No
9	BL08208	NA	NA	NA	NA	10,88	NA	NA	0,46	104,10	-0,63	NA	0,79	GPA33	3	No	No
9	BL08298	NA	NA	NA	NA	311,48	NA	NA	0,66	556,30	-0,59	NA	0,42	SLC27A2	4	No	No
9	BL08486	NA	NA	NA	NA	112,00	NA	NA	0,70	386,09	-0,93	0,52	1,11	C4orf33	2	No	No
9	BL08652	NA	NA	NA	NA	40,39	NA	NA	0,63	159,33	-0,63	NA	0,65	ST8SIA4	1	No	No
9	BL08814	NA	NA	NA	NA	NA	NA	NA	NA	123,62	-0,81	NA	0,58	ACTB	0	No	No
9	BL09025	NA	NA	NA	NA	47,05	NA	NA	0,56	500,18	-0,99	0,33	0,54	ANPEP	0	No	No
9	BL09474	NA	NA	NA	NA	358,49	NA	NA	0,33	202,82	-0,75	NA	0,66	Hypp2014	2	No	No
9	BL09798	NA	NA	NA	NA	161,60	NA	NA	0,96	652,74	-0,60	NA	0,84	GPX3	1	No	No
9	BL10432	NA	NA	NA	NA	NA	NA	NA	NA	52,33	-0,96	0,92	2,02	ENPEP	3	No	No
9	BL10887	NA	NA	NA	NA	15,36	NA	NA	1,03	82,56	-1,05	NA	0,98	PXDN	0	No	No
9	BL11273	NA	NA	NA	NA	NA	NA	NA	NA	56,40	-0,89	NA	0,75	HDC	2	No	No
9	BL11371	NA	NA	NA	NA	168,01	NA	NA	0,51	86,83	-0,88	NA	0,50	GGT1	1	No	No
9	BL11786	NA	NA	NA	NA	568,60	NA	NA	0,40	447,52	-0,75	NA	0,46	ENSG00000262499	4	No	No
9	BL12091	NA	NA	NA	NA	NA	NA	NA	NA	172,65	-1,51	0,65	1,39	HEBP2	3	No	No
9	BL12259	NA	NA	NA	NA	50,66	NA	NA	0,79	98,52	-0,72	0,53	0,86	BHMT	2	No	No
9	BL12283	NA	NA	NA	NA	669,46	-0,29	NA	0,66	1037,33	-0,31	0,30	0,47	WNT7B	0	No	No
9	BL12347	NA	NA	NA	NA	NA	NA	NA	NA	462,14	-0,53	0,40	0,80	HMCN1	1	No	No
9	BL13215	NA	NA	NA	NA	NA	NA	NA	NA	75,00	-1,51	0,84	1,39	MS4A4A	4	No	No
9	BL13430	NA	NA	NA	NA	42,43	NA	NA	0,50	109,88	-0,59	NA	0,92	GSTO1	2	No	No
9	BL13540	NA	NA	NA	NA	315,70	NA	NA	0,80	467,50	-0,54	0,55	0,65	SLC5A2	0	No	No
9	BL13588	NA	NA	NA	NA	52,80	NA	NA	0,73	66,26	-0,77	0,87	1,12	OAS1	3	No	No
9	BL14059	NA	NA	NA	NA	34,72	NA	NA	0,46	66,53	-1,09	NA	0,59	NUDT19	2	No	No
9	BL14292	NA	NA	NA	NA	527,47	NA	0,42	0,40	570,41	-0,34	0,76	0,37	Hypp7199	1	No	No
9	BL14526	NA	NA	NA	NA	NA	NA	NA	NA	152,04	-2,53	NA	1,49	Hypp11602	1	No	No
9	BL14604	NA	NA	NA	NA	NA	NA	NA	NA	98,50	-1,44	0,80	1,57	M6PR	1	No	No
9	BL14606	NA	NA	NA	NA	NA	NA	NA	NA	101,90	-1,27	0,56	1,53	Hypp1138	2	No	No
9	BL14799	NA	NA	NA	NA	48,06	NA	NA	0,88	329,94	-0,59	0,59	0,74	CMTM8	0	No	No
9	BL15078	NA	NA	NA	NA	17,17	NA	NA	1,15	58,09	-1,15	NA	1,01	SMPDL3B	0	No	No
9	BL15153	NA	NA	NA	NA	41,55	NA	NA	0,62	276,43	-0,82	NA	0,78	TSHR	4	No	No
9	BL15252	NA	NA	NA	NA	NA	NA	NA	NA	129,57	-0,56	0,78	0,89	SMPD3	0	No	No
9	BL15408	NA	NA	NA	NA	23,39	NA	NA	0,91	99,11	-0,82	NA	0,72	KLHL9	1	No	No
9	BL15688	NA	NA	NA	NA	18,34	NA	NA	0,45	128,82	-0,89	NA	0,50	PRR5	1	No	No
9	BL15754	NA	NA	NA	NA	369,07	NA	NA	0,66	1020,11	-1,04	0,43	1,11	LCTL	3	No	No
9	BL15778	NA	NA	NA	NA	NA	NA	NA	NA	112,49	-0,99	0,61	1,07	Hypp8768	0	No	No
9	BL16040	NA	NA	NA	NA	249,50	NA	NA	0,68	225,03	-0,71	NA	0,45	Hypp4763	4	No	No
9	BL16116	NA	NA	NA	NA	148,36	NA	NA	0,44	237,90	-0,71	NA	0,45	ALDH2	4	No	No
9	BL16317	NA	NA	NA	NA	16,93	NA	NA	0,87	69,31	-0,85	0,75	0,74	UGT2A1	2	No	No
9	BL16710	NA	NA	NA	NA	NA	NA	NA	NA	74,10	-0,93	0,86	1,11	ENPP7	6	No	No
9	BL16815	NA	NA	NA	NA	NA	NA	NA	NA	52,65	-1,31	NA	0,92	SLC3A1	2	No	No
9	BL17571	NA	NA	NA	NA	148,55	NA	NA	0,51	226,73	-1,08	0,69	0,75	Hypp8255	0	No	No
9	BL17784	NA	NA	NA	NA	NA	NA	NA	NA	126,72	-0,88	NA	0,90	ABCC1	1	No	No
9	BL17982	NA	NA	NA	NA	132,48	NA	NA	0,70	337,38	-0,65	NA	0,51	Hypp1189	1	No	No
9	BL17985	NA	NA	NA	NA	85,69	NA	NA	0,81	425,30	-0,61	NA	0,64	ANXA7	0	No	No
9	BL19525	3278,00	NA	NA	0,17	3339,70	-0,24	0,21	0,43	1386,87	-1,09	NA	0,54	CLYBL	3	No	No
9	BL19854	NA	NA	NA	NA	NA	NA	NA	NA	214,98	-0,61	0,94	1,01	HPGDS	1	No	No
9	BL19907	NA	NA	NA	NA	NA	NA	NA	NA	134,07	-1,02	NA	1,19	Hypp12046	2	No	Yes
9	BL19994	NA	NA	NA	NA	589,96	NA	NA	0,28	582,83	-0,89	0,62	1,38	Hypp1577	2	No	No
9	BL20203	NA	NA	NA	NA	34,45	NA	NA	0,79	53,50	-0,82	NA	0,80	ACE	2	No	No
9	BL20775	147,79	NA	NA	0,30	217,89	NA	NA	0,53	150,79	-0,61	NA	0,79	GAL3ST1	1	No	No
9	BL20840	NA	NA	NA	NA	143,47	NA	NA	0,61	226,23	-0,54	NA	0,34	RTCD1	7	No	Yes
9	BL20909	NA	NA	NA	NA	NA	NA	NA	NA	79,60	-1,46	NA	1,45	Hypp10252	2	No	No
9	BL21069	NA	NA	NA	NA	44,20	NA	NA	0,74	87,67	-0,88	NA	0,84	CTH	0	No	No
9	BL21211	165,57	NA	NA	0,30	179,70	NA	NA	0,57	181,74	-0,58	0,50	0,69	SLC15A4	1	No	No
9	BL21814	NA	NA	NA	NA	356,29	NA	NA	0,67	690,29	-0,56	NA	0,35	Hypp14776	1	No	No
9	BL21859	NA	NA	NA	NA	NA	NA	NA	NA	191,81	-0,81	NA	0,84	POSTN	3	No	No
9	BL22062	NA	NA	NA	NA	99,55	NA	NA	0,69	424,30	-0,55	0,60	0,42	B4GALT2	1	No	No
9	BL22139	NA	NA	NA	NA	NA	NA	NA	NA	55,31	-0,93	NA	1,04	Hypp12264	2	No	No
9	BL22299	NA	NA	NA	NA	119,09	NA	NA	0,86	942,32	-1,39	NA	1,09	TSPAN11	1	No	No
9	BL22304	NA	NA	NA	NA	145,69	NA	NA	0,42	550,48	-0,41	0,56	0,68	TSPAN8	1	Yes	No
9	BL22469	3742,32	NA	NA	0,22	6298,32	-0,29	NA	0,27	6841,07	-0,41	0,61	0,47	Hypp10512	4	No	No
9	BL22598	NA	NA	NA	NA	NA	NA	NA	NA	70,05	-0,78	NA	0,63	GRHPR	3	No	No
9	BL22851	NA	NA	NA	NA	117,74	NA	NA	0,68	413,67	-0,57	NA	0,35	FTH1	5	No	No
9	BL23027	NA	NA	NA	NA	53,28	NA	NA	0,73	98,00	-0,76	0,87	1,22	SLC16A7	2	No	No
9	BL23112	NA	NA	NA	NA	NA	NA	NA	NA	108,08	-0,72	0,75	0,73	AHNAK	3	No	No
9	BL23140	NA	NA	NA	NA	470,17	NA	NA	0,33	451,24	-0,67	NA	0,44	Hypp13571	1	No	No
9	BL23200	NA	NA	NA	NA	349,76	NA	NA	0,57	166,99	-0,52	0,51	0,69	SLC7A8	1	No	No
9	BL23280	NA	NA	NA	NA	120,71	NA	NA	0,70	145,90	-0,45	NA	0,54	EPHX2	3	No	Yes
9	BL23348	NA	NA	NA	NA	NA	NA	NA	NA	141,92	-0,69	NA	0,89	KIAA1161	19	No	No
9	BL23535	NA	NA	NA	NA	160,69	NA	NA	0,69	178,88	-0,47	0,43	0,63	Hypp6305	0	No	No
9	BL23574	NA	NA	NA	NA	NA	NA	NA	NA	187,55	-1,38	0,56	0,97	Hypp13420	1	No	No
9	BL23755	NA	NA	NA	NA	865,54	-0,24	NA	0,67	1023,26	-0,36	0,28	0,49	NID2	0	No	No
9	BL23896	NA	NA	NA	NA	42,06	NA	NA	1,10	86,15	-0,84	NA	0,67	QRSL1	0	No	No
9	BL24138	NA	NA	NA	NA	1083,94	NA	NA	0,47	1038,57	-0,49	0,24	0,64	CH25H	2	No	No
9	BL24271	NA	NA	NA	NA	101,27	NA	NA	1,02	181,02	-0,86	0,74	0,97	TTC38	0	No	No
9	BL24583	15,02	NA	NA	0,28	NA	NA	NA	NA	55,95	-0,67	0,68	0,66	TRIM2	3	No	No
9	BL24609	NA	NA	NA	NA	111,33	NA	NA	0,92	316,69	-0,95	0,67	0,91	Hypp10264	1	No	No
9	BL24654	NA	NA	NA	NA	NA	NA	NA	NA	264,40	-0,78	NA	0,89	Hypp814	0	No	No
9	BL24743	NA	NA	NA	NA	NA	NA	NA	NA	52,30	-0,79	NA	0,74	MFHAS1	4	No	No
9	BL31423	NA	NA	NA	NA	20,43	NA	NA	0,46	99,03	-1,22	NA	0,93	POFUT1	1	Yes	No
9	BL34086	NA	NA	NA	NA	NA	NA	NA	NA	84,60	-0,59	0,55	0,68	Hypnc1693	1	No	No
9	BL39116	NA	NA	NA	NA	157,94	NA	NA	1,24	546,34	-0,69	NA	0,55	KIAA1161	0	No	No
9	BL39793	NA	NA	NA	NA	98,33	NA	NA	0,50	144,52	-0,60	0,66	0,84	Hypnc6658	1	No	No
9	BL42081	NA	NA	NA	NA	368,95	NA	NA	0,34	360,54	-0,48	NA	0,63	Hypnc8653	1	No	No
9	BL43276	NA	NA	NA	NA	33,66	NA	NA	0,55	168,56	-0,70	NA	0,90	Hypnc9684	0	No	No
9	BL44413	NA	NA	NA	NA	29,65	NA	NA	0,47	62,13	-0,70	0,74	1,20	Hypnc10693	0	No	No
9	BL44424	NA	NA	NA	NA	NA	NA	NA	NA	76,78	-0,91	NA	0,63	IGFALS	0	No	No
9	BL45642	NA	NA	NA	NA	85,45	NA	NA	0,87	99,49	-0,90	0,61	0,85	Hypnc11797	1	No	No
9	BL50966	NA	NA	NA	NA	10,37	NA	NA	0,90	50,74	-0,95	NA	1,18	Hypp9107	0	No	No
9	BL51802	NA	NA	NA	NA	NA	NA	NA	NA	296,29	-0,73	0,45	0,46	Hypp12604	0	No	No
9	BL60259	NA	NA	NA	NA	NA	NA	NA	NA	140,07	-1,06	0,59	1,01	Hypnc24674	1	No	No
9	BL62117	NA	NA	NA	NA	467,80	NA	0,37	0,58	606,06	-0,65	0,63	0,51	Hypnc26309	4	No	No
9	BL68479	NA	NA	NA	NA	NA	NA	NA	NA	209,94	-0,68	NA	0,55	Hypnc31862	1	No	No
9	BL70669	NA	NA	NA	NA	32,41	NA	NA	1,00	117,14	-0,91	0,66	1,02	Hypp609	0	No	No
9	BL86400	NA	NA	NA	NA	NA	NA	NA	NA	60,08	-1,76	NA	1,14	Hypnc47547	1	No	No
9	BL91552	14,44	NA	NA	0,22	60,61	NA	NA	0,56	54,09	-0,90	NA	0,64	Hypnc52096	4	Yes	No
9	BL95073	2114,91	NA	NA	0,23	1448,84	-0,30	NA	0,47	326,86	-0,70	0,55	0,73	AKR1A1	1	No	No
9	BL95128	NA	NA	NA	NA	12,64	NA	NA	0,42	170,94	-1,17	NA	0,97	HPGDS	2	No	No
9	BL95404	NA	NA	NA	NA	NA	NA	NA	NA	150,00	-2,27	0,99	1,73	ALPL	1	No	No
9	BL95623	NA	NA	NA	NA	746,17	NA	0,31	0,63	935,27	-0,37	0,52	0,37	PXDN	2	No	No
9	BL95931	NA	NA	NA	NA	232,86	NA	NA	0,90	378,35	-0,80	0,56	0,92	CYP2U1	3	No	No
9	BL96028	54,05	NA	0,41	0,45	76,95	NA	NA	0,75	79,01	-0,74	0,66	0,58	ALDH3A2	3	No	No
9	BL96122	NA	NA	NA	NA	79,21	NA	NA	0,61	102,16	-0,68	NA	0,53	CES2	1	No	No
9	BL96139	NA	NA	NA	NA	NA	NA	NA	NA	55,94	-0,89	0,95	1,82	EPDR1	1	No	No
9	BL96453	NA	NA	NA	NA	356,89	NA	NA	0,73	806,87	-0,85	NA	0,81	Hypp4326	1	No	No
9	BL96454	NA	NA	NA	NA	74,21	NA	NA	0,54	179,35	-1,14	NA	1,09	Hypp4327	0	No	No
9	BL96485	NA	NA	NA	NA	68,26	NA	NA	0,82	252,37	-0,74	0,60	0,50	CYP2U1	3	No	No
9	BL96820	NA	NA	NA	NA	303,58	NA	NA	0,38	627,11	-0,64	0,58	0,88	AEBP1	0	No	No
9	BL96840	NA	NA	NA	NA	NA	NA	NA	NA	64,62	-0,91	NA	1,21	CNTNAP2	1	No	No
9	BL96988	NA	NA	NA	NA	20,29	NA	NA	0,67	172,30	-0,82	1,06	0,94	MFGE8	1	No	No

		N0				N3				N4							
Cluster	Gene	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS4934	l2FC R1158667	l2FC RA	Auto nomenclature	N.º of RAREs	ATACseq peaks	RARE hubs
9	BL97003	NA	NA	NA	NA	152,21	NA	NA	0,87	293,85	-0,74	0,36	0,60	DDAH1	0	No	No
9	BL97458	NA	NA	NA	NA	571,95	NA	0,61	0,66	950,95	-0,65	0,54	0,49	Hypp10701	3	No	No
9	BL97482	NA	NA	NA	NA	NA	NA	NA	NA	137,81	-2,08	0,70	1,55	SI	2	No	No
9	BL97483	NA	NA	NA	NA	NA	NA	NA	NA	88,57	-1,91	NA	1,20	ENSG00000259858	1	No	No
10	BL00390	NA	NA	NA	NA	509,30	-0,43	0,53	1,77	707,15	-0,86	1,16	1,56	Hypp5079 (INTU)	1	Yes	No
10	BL00923	NA	NA	NA	NA	277,20	NA	NA	1,08	168,83	-1,18	1,12	1,21	ZNF214	3	No	No
10	BL01485	NA	NA	NA	NA	69,96	NA	NA	1,29	58,28	-0,70	0,97	1,33	C1QTNF6	3	No	No
10	BL03534	NA	NA	NA	NA	76,99	NA	0,86	1,78	418,10	-1,34	2,04	1,87	CYP3A43	3	No	No
10	BL05626	NA	NA	NA	NA	36,70	NA	NA	1,61	198,46	-1,31	1,01	1,57	Hypp11783	3	No	No
10	BL10272	NA	NA	NA	NA	24,77	NA	NA	1,66	82,65	-0,88	1,56	2,55	Hypp2724	3	No	No
10	BL10283	NA	NA	NA	NA	21,74	NA	NA	1,27	93,22	-1,49	1,30	1,68	Hypp295	2	No	Yes
10	BL10973	NA	NA	NA	NA	58,17	NA	NA	1,75	347,49	-1,39	1,47	1,61	BHMT	3	No	No
10	BL11593	NA	NA	NA	NA	257,60	-0,45	0,37	1,22	455,73	-0,64	0,93	1,30	ASTN2	0	No	No
10	BL12039	NA	NA	NA	NA	21,80	NA	NA	0,69	97,81	-0,81	1,43	1,75	FAM5C	1	No	No
10	BL13116	NA	NA	NA	NA	219,55	-0,51	0,60	1,24	353,27	-0,92	1,83	1,54	Hypp4659	1	No	No
10	BL14770	22,68	NA	NA	0,23	126,04	-0,46	0,83	1,51	284,67	-1,29	1,44	1,24	CYP3A4	1	No	No
10	BL15781	NA	NA	NA	NA	25,56	NA	NA	1,02	61,11	-0,71	1,04	1,74	KBTBD8	2	No	No
10	BL16994	NA	NA	NA	NA	126,67	NA	NA	1,20	259,15	-0,96	0,79	1,08	TTC38	0	No	No
10	BL18699	NA	NA	NA	NA	469,35	-0,61	0,42	1,53	921,72	-0,63	1,07	1,20	BHMT	5	No	No
10	BL18940	NA	NA	NA	NA	105,14	-0,80	0,50	1,96	40,65	-0,79	0,98	1,49	CYP2D6	1	No	No
10	BL21237	NA	NA	NA	NA	61,92	NA	NA	1,51	85,71	-0,94	0,86	1,49	CROCC	3	No	No
10	BL22946	45,84	NA	NA	0,39	285,22	-0,58	0,69	1,65	654,62	-1,08	1,42	1,14	CYP3A5	1	Yes	No
10	BL24660	NA	NA	NA	NA	94,61	NA	NA	1,25	155,20	-0,99	0,96	1,22	CYP2U1	1	No	No
10	BL39398	NA	NA	NA	NA	15,98	NA	NA	0,81	67,44	-0,69	0,91	1,52	UGT2B7	0	No	No
10	BL65403	NA	NA	NA	NA	452,97	-0,53	NA	0,96	331,23	-0,99	1,32	1,65	FUT7	2	Yes	No
10	BL80281	35,06	NA	NA	0,38	145,72	-0,77	0,95	1,56	135,81	-1,04	1,84	1,57	Hypp10313	0	No	No
10	BL95443	NA	NA	NA	NA	559,03	-0,72	NA	1,76	1020,49	-0,80	0,89	1,99	TGM1	4	Yes	No
10	BL95693	NA	NA	NA	NA	42,93	NA	NA	1,02	91,43	-0,63	1,19	1,54	PTGR1	6	No	No
10	BL95757	NA	NA	NA	NA	449,26	NA	NA	1,43	642,97	-0,67	0,73	1,52	CNTN5	5	No	Yes
10	BL96616	NA	NA	NA	NA	49,45	NA	NA	1,94	52,32	-1,01	1,03	1,74	SULT2A1	1	No	No
10	BL97071	NA	NA	NA	NA	16,65	NA	NA	1,32	50,65	-0,98	1,29	1,75	ABCC4	3	Yes	No
11	BL00186	NA	NA	NA	NA	45,44	NA	NA	-0,54	157,42	NA	-1,62	-2,20	GBX2	0	No	No
11	BL00826	NA	NA	NA	NA	NA	NA	NA	NA	240,78	NA	-0,67	-2,76	KREMEN2	3	No	Yes
11	BL07777	NA	NA	NA	NA	53,16	0,79	NA	-1,02	31,12	0,86	-1,55	-1,88	BHLHA9	1	No	No
11	BL12405	NA	NA	NA	NA	69,22	0,93	NA	-1,28	386,54	NA	-1,60	-2,93	PPIB	2	No	No
11	BL14214	NA	NA	NA	NA	121,45	0,65	NA	-1,18	155,38	0,51	-1,73	-2,41	FERD3L	0	No	No
11	BL16282	NA	NA	NA	NA	NA	NA	NA	NA	51,99	NA	-2,44	-2,66	EVX1	2	No	No
11	BL22609	NA	NA	NA	NA	NA	NA	NA	NA	84,82	NA	-1,78	-2,49	WNT9A	2	No	No
11	BL23218	NA	NA	NA	NA	187,86	NA	NA	-1,45	346,49	NA	-1,07	-2,68	FGFA	1	No	No
12	BL00676	NA	NA	NA	NA	NA	NA	NA	NA	209,61	NA	-0,50	-0,96	FLRT1	2	No	No
12	BL00713	NA	NA	NA	NA	72,38	NA	NA	-1,01	683,39	NA	-0,45	-0,85	FOXQ1	0	No	No
12	BL00954	NA	NA	NA	NA	1828,15	NA	NA	-0,41	470,29	NA	-0,81	-0,54	Hypp14222	1	No	No
12	BL01643	NA	NA	NA	NA	NA	NA	NA	NA	62,72	NA	-1,02	-0,77	TMC7	0	Yes	No
12	BL01674	NA	NA	NA	NA	73,34	NA	NA	-1,43	320,86	NA	-1,02	-1,38	NKX2-6	0	No	No
12	BL01977	NA	NA	NA	NA	3725,99	NA	NA	-0,59	1002,78	NA	-0,61	-0,59	CYCS	0	No	No
12	BL01988	NA	NA	NA	NA	228,53	NA	NA	-0,45	1138,37	NA	-0,78	-1,04	TBX1	5	No	No
12	BL02037	NA	NA	NA	NA	74,92	-0,49	NA	NA	85,26	-1,14	-0,91	NA	Hypp6740	3	No	Yes
12	BL02462	NA	NA	NA	NA	24,75	NA	NA	-0,45	191,90	NA	-1,06	-1,50	PADI2	4	No	No
12	BL02582	NA	NA	NA	NA	1039,28	NA	NA	-0,47	287,37	NA	-0,60	-0,52	Hypp5721	0	No	No
12	BL02610	NA	NA	NA	NA	429,32	NA	NA	-0,90	780,71	0,37	-0,59	-0,69	FAM167A	0	No	No
12	BL03164	NA	NA	NA	NA	NA	NA	NA	NA	303,45	-0,60	-0,51	NA	Hypp10905	1	No	No
12	BL03359	NA	NA	NA	NA	31,79	NA	NA	-0,66	61,11	NA	-0,59	-0,93	SCGN	2	No	No
12	BL03761	NA	NA	NA	NA	53,72	NA	NA	-0,61	104,46	NA	-0,63	-1,38	DPEP1	1	No	No
12	BL03998	NA	NA	NA	NA	908,57	0,33	NA	-0,48	364,76	NA	-0,69	-0,81	TYR	1	No	No
12	BL04011	NA	NA	NA	NA	206,82	NA	NA	-0,88	126,36	NA	-0,65	-1,15	RSPO2	4	No	No
12	BL04071	NA	NA	NA	NA	NA	NA	NA	NA	54,39	NA	-0,85	NA	Hypp2620	2	No	No
12	BL04317	NA	NA	NA	NA	857,72	-0,32	-0,29	-0,45	613,15	NA	-1,01	-0,89	CYP1A1	0	No	No
12	BL04369	NA	NA	NA	NA	593,63	NA	-0,27	-1,50	317,28	0,81	-0,63	-1,33	AHNAK2	0	No	No
12	BL04376	NA	NA	NA	NA	1319,15	0,38	-0,24	-0,51	1087,15	NA	-0,71	-0,51	LEFTY2	0	No	No
12	BL04395	NA	NA	NA	NA	NA	NA	NA	NA	119,30	NA	-0,91	-0,99	GCM2	0	Yes	No
12	BL04401	NA	NA	NA	NA	NA	NA	NA	NA	104,41	NA	-0,95	-0,80	BHLHE23	2	No	Yes
12	BL04459	NA	NA	NA	NA	15,59	NA	NA	-0,43	213,57	NA	-1,04	-1,93	PRDM13	1	No	No
12	BL04677	NA	NA	NA	NA	2308,81	0,16	NA	NA	1354,90	NA	-0,53	-0,64	TBX6/16	0	No	No
12	BL04768	NA	NA	NA	NA	2151,03	0,24	NA	-0,63	579,16	NA	-0,64	-0,90	ENSG00000262654	0	No	No
12	BL05178	NA	NA	NA	NA	1163,73	NA	NA	-0,47	5829,69	NA	-0,63	-0,55	Hypp10131	4	No	Yes
12	BL05323	NA	NA	NA	NA	NA	NA	NA	NA	250,84	NA	-1,00	-1,37	CTHRC1	0	No	No
12	BL05533	NA	NA	NA	NA	NA	NA	NA	NA	108,76	NA	-0,58	-1,43	Hypp11350	0	No	No
12	BL05807	NA	NA	NA	NA	77,74	NA	NA	-0,78	156,28	NA	-0,92	-0,89	BARHL1	0	No	No
12	BL05977	NA	NA	NA	NA	NA	NA	NA	NA	761,06	NA	-0,48	-0,83	PAX6	3	No	No
12	BL06206	NA	NA	NA	NA	NA	NA	NA	NA	218,77	NA	-0,46	-0,88	C20orf152	0	No	No
12	BL06405	NA	NA	NA	NA	NA	NA	NA	NA	279,06	NA	-0,71	NA	Hypp2062	0	Yes	No
12	BL06431	NA	NA	NA	NA	3562,90	0,18	NA	-0,49	2451,59	NA	-0,41	-0,74	FGFRL1	1	No	No
12	BL06501	2787,68	-0,30	0,23	0,68	NA	NA	NA	NA	757,82	-0,54	-0,64	-0,65	Hypp15831	1	No	No
12	BL06578	NA	NA	NA	NA	31174,66	NA	NA	-0,38	4091,97	NA	-1,09	-0,82	Hypp13448	0	No	No
12	BL06606	NA	NA	NA	NA	18,20	NA	NA	-0,45	117,37	NA	-0,97	-1,38	PLSCR1	2	No	No
12	BL06759	NA	NA	NA	NA	514,61	NA	NA	-0,34	550,43	NA	-0,59	-0,39	ST6GALNAC4	1	No	No
12	BL07004	1853,29	NA	-0,22	NA	6667,33	NA	-0,27	-0,40	2142,01	NA	-0,58	-0,64	Hypp11579	2	No	No
12	BL07280	NA	NA	NA	NA	NA	NA	NA	NA	123,47	NA	-0,67	-1,26	FOXL2	0	No	No
12	BL07427	NA	NA	NA	NA	NA	NA	NA	NA	235,52	-0,74	-0,68	NA	Hypp2397	2	No	No
12	BL07454	NA	NA	NA	NA	182,42	NA	NA	-0,82	123,82	NA	-0,77	-1,10	DMRTA2	0	No	No
12	BL07471	NA	NA	NA	NA	1621,84	0,19	NA	-0,76	1328,27	NA	-0,46	-0,61	IRX6	0	No	No
12	BL07606	NA	NA	NA	NA	NA	NA	NA	NA	1727,29	-0,60	-0,54	NA	TOB2	0	No	No
12	BL08147	NA	NA	NA	NA	NA	NA	NA	NA	96,45	NA	-0,67	-0,72	CHST15	1	No	No
12	BL08329	NA	NA	NA	NA	404,78	NA	NA	-0,47	396,92	NA	-0,55	-0,59	DIO1	0	No	No
12	BL08389	NA	NA	NA	NA	721,31	NA	NA	-0,23	889,20	NA	-0,50	-0,84	SIX1	0	No	No
12	BL08675	NA	NA	NA	NA	NA	NA	NA	NA	1674,38	-0,86	-0,61	-0,42	Hypp6769	4	No	No
12	BL09554	NA	NA	NA	NA	NA	NA	NA	NA	113,44	NA	-0,57	-0,86	PTH1R	0	No	No
12	BL09761	NA	NA	NA	NA	72,22	NA	NA	-0,51	25,44	NA	-0,87	-1,03	Hypp12773	2	No	No
12	BL10176	NA	NA	NA	NA	9555,23	NA	NA	-0,35	1788,87	NA	-0,47	-0,70	CABP1	0	No	No
12	BL10388	NA	NA	NA	NA	NA	NA	NA	NA	130,73	NA	-0,63	NA	Hypp13114	1	No	No
12	BL10496	NA	NA	NA	NA	79,44	NA	NA	-0,52	100,84	NA	-0,62	-0,81	Hypp9217	1	No	No
12	BL10670	NA	NA	NA	NA	153,01	NA	NA	-0,49	152,70	NA	-0,68	-0,68	Hypp2903	1	No	No
12	BL10950	NA	NA	NA	NA	NA	NA	NA	NA	57,84	NA	-0,60	NA	Hypp6196	0	No	No
12	BL11067	NA	NA	NA	NA	128,10	NA	NA	-0,62	199,19	NA	-0,93	-0,86	FOXG1	0	No	No
12	BL11301	NA	NA	NA	NA	57,82	NA	NA	-0,48	21,43	NA	-0,78	-0,88	NPFFR2	3	No	No
12	BL11304	NA	NA	NA	NA	47,35	NA	NA	-0,56	61,34	NA	-0,96	-1,23	Hypp4548	1	No	No
12	BL11511	NA	NA	NA	NA	NA	NA	NA	NA	360,56	NA	-0,68	-0,87	Hypp6285	1	No	No
12	BL11740	NA	NA	NA	NA	NA	NA	NA	NA	52,59	NA	-0,63	NA	Hypp10057	0	No	No
12	BL11870	NA	NA	NA	NA	47,99	NA	NA	-0,44	667,80	NA	-0,59	-0,98	NMRAL1	1	No	No
12	BL11983	NA	NA	NA	NA	NA	NA	NA	NA	137,33	NA	-0,69	-0,69	C6orf192	0	No	No
12	BL12180	NA	NA	NA	NA	NA	NA	NA	NA	194,85	NA	-1,12	-1,35	TRIM3	0	No	No
12	BL12281	NA	NA	NA	NA	71,90	NA	NA	-0,46	234,82	NA	-0,55	-0,85	Hypp2018	0	No	No
12	BL12305	NA	NA	NA	NA	53,31	NA	NA	-0,46	196,50	NA	-0,73	-0,77	TTN	2	No	No
12	BL12568	NA	NA	NA	NA	NA	NA	NA	NA	315,35	-0,44	-0,62	-0,46	ITPK1	1	No	No
12	BL12801	NA	NA	NA	NA	5098,51	0,30	NA	-0,41	4414,66	NA	-0,68	-0,78	PITX2	0	No	No
12	BL12869	773,03	NA	NA	-0,24	491,26	0,24	NA	-0,65	360,15	NA	-0,49	-0,98	BHLHA15	0	No	No
12	BL12992	NA	NA	NA	NA	42,40	NA	NA	-0,49	279,29	NA	-0,60	-0,84	BARHL1	2	No	No
12	BL13922	NA	NA	NA	NA	NA	NA	NA	NA	328,99	NA	-0,60	-0,84	PAX6	4	No	No
12	BL13992	NA	NA	NA	NA	NA	NA	NA	NA	183,74	NA	-0,74	NA	Hypp1702	1	No	No
12	BL14114	NA	NA	NA	NA	811,86	0,34	NA	-0,51	1124,49	NA	-0,50	-0,74	SAMD7	2	No	No

		N0				N3				N4							
Cluster	Gene	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS4934	l2FC R1158667	l2FC RA	Auto nomenclature	N.º of RAREs	ATACseq peaks	RARE hubs
12	BL14213	NA	NA	NA	NA	1004,29	0,37	NA	-0,42	529,59	NA	-0,47	-0,74	OCA2	1	No	No
12	BL14765	NA	NA	NA	NA	NA	NA	NA	NA	53,10	NA	-1,12	-1,49	SLC45A2	3	No	No
12	BL14792	NA	NA	NA	NA	NA	NA	NA	NA	157,36	NA	-0,74	-0,69	OPN5	1	No	No
12	BL14817	NA	NA	NA	NA	NA	NA	NA	NA	122,48	NA	-0,66	-0,71	NKX1-1	0	No	No
12	BL14839	NA	NA	NA	NA	6881,06	NA	-0,18	NA	7095,06	-0,27	-0,59	-0,23	GLUL	1	No	No
12	BL15774	NA	NA	NA	NA	NA	NA	NA	NA	512,69	NA	-0,82	-1,80	FOXI1	1	Yes	No
12	BL15847	NA	NA	NA	NA	111,05	NA	NA	-0,43	214,12	NA	-0,54	-0,93	SLC1A2	2	No	No
12	BL15943	NA	NA	NA	NA	NA	NA	NA	NA	171,04	NA	-0,61	NA	Hypp5508	0	No	No
12	BL16061	NA	NA	NA	NA	NA	NA	NA	NA	56,14	NA	-0,63	-0,58	PRDM8	1	No	No
12	BL16315	NA	NA	NA	NA	259,21	NA	NA	-0,41	259,60	NA	-0,66	-0,57	HSD17B8	1	No	No
12	BL16360	NA	NA	NA	NA	1379,59	NA	NA	-0,52	1105,25	NA	-0,67	-0,37	Hypp1454	0	No	No
12	BL16522	NA	NA	NA	NA	502,15	0,38	-0,38	-1,57	829,63	NA	-1,14	-1,71	FOXE1 (FOXEa)	0	No	No
12	BL16784	NA	NA	NA	NA	537,16	0,25	NA	NA	249,54	NA	-0,50	-0,60	Hypp3736	2	Yes	Yes
12	BL17050	NA	NA	NA	NA	180,53	NA	NA	-0,46	180,21	NA	-0,63	-0,64	PNLIP	1	No	No
12	BL17351	NA	NA	NA	NA	533,46	NA	-0,47	-0,31	850,58	NA	-0,64	NA	Hypp13545	2	No	Yes
12	BL17436	NA	NA	NA	NA	1982,54	0,21	NA	-0,68	2673,85	NA	-0,73	-0,99	PAX9	1	No	No
12	BL17549	NA	NA	NA	NA	51,81	NA	NA	-0,80	289,12	NA	-0,40	-0,92	MSX1	0	No	No
12	BL18019	NA	NA	NA	NA	NA	NA	NA	NA	121,90	NA	-0,65	-0,80	CA4	0	No	No
12	BL18053	NA	NA	NA	NA	747,53	NA	NA	-0,62	48,11	NA	-0,92	NA	SLC18A2	3	No	No
12	BL18299	NA	NA	NA	NA	142,28	NA	NA	-0,58	220,75	NA	-0,67	-0,89	Hypp12112	1	No	No
12	BL18444	NA	NA	NA	NA	2732,74	NA	NA	-0,58	1517,77	NA	-0,45	-0,88	FJX1	1	No	No
12	BL18832	NA	NA	NA	NA	115,56	NA	NA	-0,38	84,16	NA	-0,64	NA	Hypp13867	2	No	No
12	BL19044	NA	NA	NA	NA	241,34	NA	NA	-0,47	103,40	NA	-0,72	-0,88	ADRB2	0	No	No
12	BL19161	NA	NA	NA	NA	88,35	NA	NA	-0,48	135,90	NA	-0,49	-0,59	ZNF266	3	No	No
12	BL19175	NA	NA	NA	NA	689,46	NA	-0,38	-0,68	2854,46	NA	-0,90	-0,85	Hypp10132	0	No	No
12	BL19312	NA	NA	NA	NA	NA	NA	NA	NA	72,26	NA	-0,65	-1,17	CRHR2	0	No	No
12	BL20013	NA	NA	NA	NA	1064,34	0,34	NA	NA	113,06	NA	-0,66	NA	GLIS3	3	No	No
12	BL20132	NA	NA	NA	NA	NA	NA	NA	NA	599,85	0,39	-0,92	-1,20	PPIB	1	No	No
12	BL20663	NA	NA	NA	NA	NA	NA	NA	NA	431,12	-0,77	-0,50	NA	AGPHD1	0	No	No
12	BL20927	NA	NA	NA	NA	NA	NA	NA	NA	98,68	NA	-0,71	-0,91	FOXL1	2	No	No
12	BL21081	NA	NA	NA	NA	5004,67	0,20	NA	-0,73	1800,06	NA	-0,97	-1,19	SMS	1	No	No
12	BL21466	NA	NA	NA	NA	24,59	NA	NA	-0,48	188,66	NA	-1,15	-1,36	MALL	0	No	No
12	BL21529	NA	NA	NA	NA	12471,24	0,20	NA	-0,31	7611,61	NA	-0,60	-0,51	GBX1	0	No	No
12	BL21667	NA	NA	NA	NA	NA	NA	NA	NA	199,87	NA	-0,56	-0,77	FLT1	2	No	No
12	BL21698	NA	NA	NA	NA	NA	NA	NA	NA	56,63	NA	-0,72	-0,75	Hypp14325	2	No	No
12	BL21886	NA	NA	NA	NA	676,81	0,28	NA	-0,91	988,79	NA	-0,73	-1,31	DMBX1	0	No	No
12	BL21938	NA	NA	NA	NA	372,93	NA	NA	-0,41	156,05	NA	-0,68	-0,45	PTGER4	1	No	No
12	BL22761	NA	NA	NA	NA	13438,23	NA	NA	-0,33	4509,91	NA	-0,49	-0,61	Hypp4833	0	No	No
12	BL22818	NA	NA	NA	NA	804,10	NA	NA	-0,32	480,30	NA	-0,67	-0,76	C20orf152	1	No	No
12	BL22975	NA	NA	NA	NA	172,74	NA	NA	-1,00	198,02	NA	-1,15	-1,82	UNCX	1	No	No
12	BL23478	NA	NA	NA	NA	5510,48	NA	-0,19	-0,40	1569,06	NA	-0,76	-0,39	Hypp1303	0	No	No
12	BL23703	2399,23	NA	NA	-0,22	3787,08	0,20	NA	-0,48	1325,05	NA	-0,71	-0,77	TRIM36	0	No	No
12	BL23919	NA	NA	NA	NA	861,18	0,57	NA	-0,64	3910,29	NA	-0,86	-0,66	Hypp2319	2	Yes	No
12	BL24036	NA	NA	NA	NA	NA	NA	NA	NA	313,56	NA	-0,67	-0,63	Hypp15662	1	No	No
12	BL24073	NA	NA	NA	NA	725,43	NA	NA	-0,33	94,33	NA	-0,76	-0,83	PTGIR	3	No	No
12	BL24618	NA	NA	NA	NA	31,86	NA	NA	-0,54	56,40	NA	-0,67	-0,78	SFTPD	1	No	No
12	BL24661	NA	NA	NA	NA	18134,18	NA	NA	-0,31	2971,05	NA	-0,71	-0,73	Hypp13663	4	No	No
12	BL25760	NA	NA	NA	NA	4073,32	0,22	NA	-0,18	907,31	NA	-0,37	-0,67	Hypp5264	0	No	No
12	BL35269	NA	NA	NA	NA	416,92	NA	NA	-0,32	370,14	NA	-0,72	NA	Hypnc2720	3	No	No
12	BL40162	NA	NA	NA	NA	167,17	NA	NA	-0,92	143,05	NA	-0,45	-0,90	COL5A1	3	No	Yes
12	BL43735	NA	NA	NA	NA	NA	NA	NA	NA	1424,57	-0,75	-0,48	NA	Hypnc10085	4	No	No
12	BL45149	NA	NA	NA	NA	NA	NA	NA	NA	73,90	NA	-0,60	NA	Hypnc11345	0	No	No
12	BL49994	NA	NA	NA	NA	NA	NA	NA	NA	102,52	NA	-0,53	-0,70	POU3F4	0	No	No
12	BL55049	NA	NA	NA	NA	57,19	NA	NA	-0,44	32,96	NA	-0,89	-0,82	Hypp2808	0	No	No
12	BL59545	NA	NA	NA	NA	64,57	NA	NA	-0,78	56,47	NA	-1,04	-1,49	Hypp1760	0	No	No
12	BL60265	NA	NA	NA	NA	590,84	NA	NA	-0,26	749,58	NA	-0,64	NA	Hypnc24680	1	No	No
12	BL63588	NA	NA	NA	NA	NA	NA	NA	NA	94,14	NA	-0,81	-0,73	IL17D	2	No	No
12	BL69926	NA	NA	NA	NA	5824,87	NA	NA	-0,52	1517,83	NA	-0,67	-0,64	Hypp11576	0	No	No
12	BL70322	NA	NA	NA	NA	NA	NA	NA	NA	67,16	NA	-0,71	NA	Hypnc33488	0	No	No
12	BL70670	14518,14	0,15	NA	NA	5060,60	NA	-0,21	-0,54	2268,70	NA	-0,71	-0,49	Hypp609	1	No	No
12	BL73063	NA	NA	NA	NA	111,11	NA	NA	-0,50	70,11	NA	-0,63	-0,58	Hypnc35905	0	No	No
12	BL79214	NA	NA	NA	NA	58,52	NA	NA	-0,49	151,77	NA	-0,59	-0,61	Hypnc41241	0	No	No
12	BL84619	NA	NA	NA	NA	NA	NA	NA	NA	64,10	-0,60	-0,59	NA	Hypnc45972	0	No	No
12	BL90400	NA	NA	NA	NA	554,20	NA	NA	-0,63	478,83	NA	-0,42	-0,58	Hypp4445	3	No	No
12	BL91247	NA	NA	NA	NA	NA	NA	NA	NA	66,47	NA	-0,61	NA	Hypnc51825	0	No	No
12	BL95160	NA	NA	NA	NA	1308,52	NA	NA	-0,21	531,15	-0,57	-0,63	NA	DDAH1	2	No	No
12	BL95233	NA	NA	NA	NA	2780,70	NA	NA	-0,40	1321,26	NA	-0,66	-0,61	DNASE1	1	No	No
12	BL95264	NA	NA	NA	NA	NA	NA	NA	NA	259,43	NA	-0,41	-0,66	GGT1	0	No	No
12	BL95358	NA	NA	NA	NA	1203,07	NA	NA	-0,45	1113,64	NA	-0,79	-1,00	GLIPR2	2	No	No
12	BL95665	NA	NA	NA	NA	NA	NA	NA	NA	275,69	NA	-0,61	-1,01	SDR39U1	1	No	No
12	BL96010	NA	NA	NA	NA	431,05	NA	NA	-0,53	1064,89	NA	-0,62	-0,73	KRT84	0	No	No
12	BL96081	NA	NA	NA	NA	NA	NA	NA	NA	116,29	NA	-0,59	NA	SLCO4A1	0	No	No
13	BL00294	NA	NA	NA	NA	NA	NA	NA	NA	568,77	0,28	NA	-0,62	ENSG00000259777	3	No	Yes
13	BL01192	NA	NA	NA	NA	NA	NA	NA	NA	212,67	NA	NA	-0,83	NMRAL1	2	No	No
13	BL01343	NA	NA	NA	NA	101,10	0,46	NA	-0,64	99,97	NA	NA	-1,14	CHRDL1	2	No	No
13	BL02047	NA	NA	NA	NA	NA	NA	NA	NA	67,27	NA	NA	-0,68	COLEC10	3	No	No
13	BL02411	NA	NA	NA	NA	159,48	NA	NA	-0,36	142,60	NA	NA	-0,74	Hypp591	0	No	No
13	BL02659	NA	NA	NA	NA	NA	NA	NA	NA	1860,02	NA	NA	-0,65	EBF3	3	No	No
13	BL02677	NA	NA	NA	NA	NA	NA	NA	NA	88,51	NA	NA	-0,92	EDIL3	1	No	No
13	BL02840	NA	NA	NA	NA	NA	NA	NA	NA	55,31	NA	NA	-0,89	RASSF9	1	Yes	No
13	BL03226	NA	NA	NA	NA	122,22	NA	NA	-0,38	99,21	NA	NA	-0,72	MATN2	0	No	No
13	BL03815	NA	NA	NA	NA	21613,06	0,20	NA	NA	19783,96	NA	NA	-0,60	CROCC	1	No	No
13	BL03853	NA	NA	NA	NA	NA	NA	NA	NA	64,19	NA	NA	-0,78	TMEM195	0	No	No
13	BL04563	NA	NA	NA	NA	NA	NA	NA	NA	126,02	NA	NA	-0,84	MMP19	4	No	No
13	BL04898	NA	NA	NA	NA	NA	NA	NA	NA	59,53	NA	NA	-0,87	ENSG00000263052	2	No	No
13	BL05053	NA	NA	NA	NA	NA	NA	NA	NA	341,68	NA	NA	-0,70	SLC23A2	0	No	No
13	BL05545	NA	NA	NA	NA	58,56	NA	NA	-0,68	396,97	NA	NA	-0,63	Hypp10198	0	No	No
13	BL06267	NA	NA	NA	NA	NA	NA	NA	NA	56,52	NA	NA	-0,67	BSX	0	No	No
13	BL06285	NA	NA	NA	NA	NA	NA	NA	NA	73,79	NA	NA	-0,59	AOAH	5	No	No
13	BL06373	NA	NA	NA	NA	122,81	NA	NA	-0,64	NA	NA	NA	NA	RERG	2	No	No
13	BL06737	NA	NA	NA	NA	176,35	NA	NA	-0,38	688,88	NA	NA	-1,16	PLIN4	1	No	No
13	BL06903	NA	NA	NA	NA	NA	NA	NA	NA	196,24	0,57	NA	-1,17	Hypp2600	1	No	No
13	BL07393	NA	NA	NA	NA	NA	NA	NA	NA	77,34	NA	NA	-0,87	Hypp12186	1	No	No
13	BL07414	NA	NA	NA	NA	14,85	NA	NA	-0,50	139,11	NA	NA	-0,92	ALX1	2	No	Yes
13	BL07422	NA	NA	NA	NA	283,18	NA	-0,38	-0,88	247,17	NA	NA	-0,67	THBS1	1	No	No
13	BL07663	NA	NA	NA	NA	NA	NA	NA	NA	60,75	NA	NA	-0,86	FCN1	0	No	No
13	BL07839	NA	NA	NA	NA	2318,83	0,69	NA	-0,59	1133,17	0,58	-0,69	-1,14	Hypp12008	2	No	Yes
13	BL07850	NA	NA	NA	NA	NA	NA	NA	NA	65,21	0,89	NA	NA	PKHD1L1	0	No	No
13	BL07852	NA	NA	NA	NA	NA	NA	NA	NA	86,01	NA	NA	-0,70	Hypp6125	0	No	No
13	BL08114	NA	NA	NA	NA	3108,36	NA	NA	-0,66	NA	NA	NA	NA	DNAH3	1	No	No
13	BL08119	NA	NA	NA	NA	109,03	NA	NA	-0,51	235,50	0,42	-0,46	-0,69	CACNA2D2	3	No	No
13	BL08453	NA	NA	NA	NA	NA	NA	NA	NA	60,30	NA	NA	-0,59	KCNK10	2	No	No
13	BL08501	NA	NA	NA	NA	NA	NA	NA	NA	104,42	NA	NA	-0,77	HMCN1	1	No	No
13	BL08945	NA	NA	NA	NA	NA	NA	NA	NA	85,83	NA	NA	-0,72	KCNA2	2	No	No
13	BL09979	NA	NA	NA	NA	NA	NA	NA	NA	58,34	NA	NA	-0,90	FAM43A	0	No	No
13	BL10602	NA	NA	NA	NA	NA	NA	NA	NA	62,54	NA	NA	-0,77	LINGO2	1	No	No
13	BL10689	NA	NA	NA	NA	216,37	NA	NA	-0,67	299,52	NA	NA	-0,49	GPR112	4	No	Yes
13	BL10875	NA	NA	NA	NA	37,73	NA	NA	-0,69	67,56	NA	NA	-1,06	ST8SIA1	2	No	No
13	BL11000	NA	NA	NA	NA	NA	NA	NA	NA	56,79	NA	NA	-0,92	EYS	2	No	No
13	BL11393	NA	NA	NA	NA	164,61	NA	NA	-0,93	412,73	NA	NA	-0,69	SLC7A14	6	No	No

		N0				N3				N4							
Cluster	Gene	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS4934	l2FC R1158667	l2FC RA	Auto nomenclature	N.º of RAREs	ATACseq peaks	RARE hubs
13	BL11849	NA	NA	NA	NA	394,62	NA	NA	-0,36	115,52	NA	NA	-0,84	CACNA1H	5	No	No
13	BL11855	NA	NA	NA	NA	264,55	NA	NA	-0,47	88,05	NA	NA	-0,92	CACNA1G	0	No	No
13	BL12055	NA	NA	NA	NA	47,57	NA	NA	-0,77	269,40	0,57	1,05	-0,97	Hypp13502	1	No	No
13	BL12392	NA	NA	NA	NA	NA	NA	NA	NA	59,47	NA	NA	-0,66	Hypp12759	3	No	No
13	BL12667	NA	NA	NA	NA	NA	NA	NA	NA	174,88	0,62	NA	NA	Hypp11573	3	No	No
13	BL12798	NA	NA	NA	NA	NA	NA	NA	NA	290,19	0,47	NA	-0,66	C3orf63	0	No	No
13	BL12814	NA	NA	NA	NA	NA	NA	NA	NA	84,03	NA	NA	-0,71	Hypp15520	2	No	No
13	BL12897	NA	NA	NA	NA	13,18	NA	NA	-0,44	83,66	1,09	NA	NA	DEGS1	0	No	No
13	BL13229	NA	NA	NA	NA	410,92	NA	NA	-0,27	820,64	0,54	-0,42	-0,62	ITIH3	0	No	No
13	BL13334	NA	NA	NA	NA	38,66	NA	NA	-0,46	83,26	NA	NA	-0,77	VEGFC	2	No	No
13	BL13365	NA	NA	NA	NA	3800,10	0,15	NA	-0,37	4516,73	0,26	-0,52	-0,89	EYA4	9	No	No
13	BL13582	NA	NA	NA	NA	NA	NA	NA	NA	72,61	NA	NA	-0,69	Hypp10046	2	No	No
13	BL13626	NA	NA	NA	NA	NA	NA	NA	NA	65,24	0,60	NA	NA	ACSS2	2	No	No
13	BL13768	NA	NA	NA	NA	796,52	0,42	NA	-0,66	564,73	0,49	NA	-0,69	Hypp13871	1	No	No
13	BL13862	NA	NA	NA	NA	NA	NA	NA	NA	248,66	NA	NA	-0,63	Hypp5233	2	No	No
13	BL13902	NA	NA	NA	NA	NA	NA	NA	NA	87,71	NA	NA	-0,84	CAPN9	1	No	No
13	BL14063	NA	NA	NA	NA	NA	NA	NA	NA	568,02	NA	NA	-0,59	DNAH5	9	No	No
13	BL14590	NA	NA	NA	NA	NA	NA	NA	NA	69,74	NA	NA	-0,82	ONECUT2	5	No	No
13	BL14905	NA	NA	NA	NA	152,32	NA	NA	-0,72	NA	NA	NA	NA	Hypp1839	0	No	No
13	BL15450	NA	NA	NA	NA	NA	NA	NA	NA	174,54	NA	NA	-0,90	KIAA1239	1	No	No
13	BL15600	NA	NA	NA	NA	NA	NA	NA	NA	124,78	0,59	NA	NA	MMP14	0	No	No
13	BL15626	NA	NA	NA	NA	50,42	NA	NA	-0,49	96,86	NA	NA	-0,68	ATOH8	0	No	No
13	BL15654	NA	NA	NA	NA	NA	NA	NA	NA	102,71	NA	NA	-0,74	NRXN2	0	No	No
13	BL15960	NA	NA	NA	NA	NA	NA	NA	NA	73,72	NA	NA	-0,67	ADAM23	3	No	No
13	BL16228	NA	NA	NA	NA	76,86	NA	NA	-0,69	246,46	NA	NA	-0,51	TMEM229A	3	No	No
13	BL16243	NA	NA	NA	NA	67,45	NA	NA	-0,62	79,99	NA	NA	-0,72	Hypp12764	10	No	No
13	BL16265	NA	NA	NA	NA	5947,46	0,60	NA	-0,21	NA	NA	NA	NA	CAPRIN2	0	No	No
13	BL16391	NA	NA	NA	NA	NA	NA	NA	NA	58,32	NA	NA	-0,88	Hypp5858	3	No	No
13	BL16554	NA	NA	NA	NA	2705,75	NA	NA	-0,23	568,52	NA	NA	-0,87	Hypp5263	0	No	No
13	BL17129	NA	NA	NA	NA	NA	NA	NA	NA	96,71	NA	NA	-0,83	GRIK2	4	No	No
13	BL17388	NA	NA	NA	NA	NA	NA	NA	NA	164,64	NA	NA	-0,70	PDE1C	0	No	No
13	BL17625	NA	NA	NA	NA	68,49	NA	NA	-0,56	211,29	NA	NA	-0,75	SLC7A9	0	No	No
13	BL17923	NA	NA	NA	NA	413,34	0,28	NA	-0,67	942,11	NA	NA	-0,52	COL12A1	0	No	No
13	BL18030	NA	NA	NA	NA	NA	NA	NA	NA	66,45	NA	NA	-0,95	LRP2	2	No	No
13	BL18193	NA	NA	NA	NA	NA	NA	NA	NA	71,97	NA	NA	-0,88	LPHN2	1	No	No
13	BL18208	NA	NA	NA	NA	NA	NA	NA	NA	61,52	NA	NA	-0,68	ADCY3	1	No	No
13	BL18410	NA	NA	NA	NA	NA	NA	NA	NA	65,68	NA	NA	-0,64	LHX2	1	No	No
13	BL18751	NA	NA	NA	NA	NA	NA	NA	NA	86,82	NA	NA	-0,65	SYT12	1	No	No
13	BL18760	NA	NA	NA	NA	315,07	NA	NA	-1,01	255,45	NA	NA	-0,70	WIF1	3	No	Yes
13	BL18815	NA	NA	NA	NA	NA	NA	NA	NA	150,42	NA	NA	-0,82	Hypp10073	3	No	No
13	BL18830	NA	NA	NA	NA	278,40	NA	-0,36	-0,92	180,56	0,45	NA	-0,43	TCTEX1D1	0	No	No
13	BL18961	NA	NA	NA	NA	198,47	NA	NA	-0,71	993,96	NA	NA	-0,56	NTRK2	1	No	No
13	BL19296	NA	NA	NA	NA	299,86	NA	NA	-0,31	295,95	0,35	NA	-0,63	RGS22	0	No	No
13	BL19445	NA	NA	NA	NA	NA	NA	NA	NA	71,66	NA	NA	-0,62	Hypp5905	1	No	No
13	BL19611	NA	NA	NA	NA	NA	NA	NA	NA	144,08	NA	NA	-0,94	WDR64	0	No	No
13	BL19671	NA	NA	NA	NA	NA	NA	NA	NA	405,77	0,44	-0,50	-0,77	LMX1B	1	No	No
13	BL19686	NA	NA	NA	NA	NA	NA	NA	NA	141,96	NA	NA	-0,61	STIL	1	No	No
13	BL19804	NA	NA	NA	NA	258,12	0,60	NA	NA	205,42	NA	NA	-1,02	GPR143	1	No	No
13	BL19861	NA	NA	NA	NA	602,79	NA	NA	-0,44	466,38	NA	NA	-0,78	CETN1	0	No	No
13	BL20208	NA	NA	NA	NA	20,82	NA	NA	-0,74	315,38	NA	NA	-0,80	Hypp15421	0	No	No
13	BL20540	NA	NA	NA	NA	1956,39	NA	NA	-0,63	NA	NA	NA	NA	Hypp13154	4	No	No
13	BL21176	NA	NA	NA	NA	29,20	NA	NA	-0,51	53,54	NA	NA	-0,83	DIRAS2	0	No	No
13	BL22043	NA	NA	NA	NA	8718,09	0,37	NA	-0,35	2122,25	0,63	NA	-0,44	CNTN5	3	No	No
13	BL22189	NA	NA	NA	NA	2146,43	NA	NA	-0,27	1041,44	0,28	-0,38	-0,60	SNED1	1	No	No
13	BL22917	NA	NA	NA	NA	NA	NA	NA	NA	220,98	NA	NA	-0,66	HMCN1	1	No	No
13	BL22971	NA	NA	NA	NA	NA	NA	NA	NA	72,18	NA	NA	-0,87	Hypp10993	2	No	No
13	BL24020	NA	NA	NA	NA	97,15	NA	NA	-0,72	118,48	NA	NA	-0,63	GRIN3A	2	No	No
13	BL24048	NA	NA	NA	NA	NA	NA	NA	NA	95,78	NA	NA	-0,68	HHIPL2	0	No	No
13	BL24078	NA	NA	NA	NA	542,60	0,47	NA	-0,64	206,48	0,59	-0,47	-1,05	ADRA1D	0	No	No
13	BL24115	6628,54	NA	NA	-0,23	33291,51	0,23	NA	-0,45	10874,90	0,80	-0,61	-0,86	ALDH1A2	1	No	No
13	BL24736	NA	NA	NA	NA	79,52	NA	NA	-0,42	81,67	NA	NA	-0,67	MAPRE3	0	No	No
13	BL24768	NA	NA	NA	NA	NA	NA	NA	NA	103,53	NA	NA	-0,72	Hypp8185	3	No	Yes
13	BL24841	NA	NA	NA	NA	97,95	NA	NA	-1,03	84,82	0,64	NA	-0,84	RD3	2	No	No
13	BL24890	NA	NA	NA	NA	NA	NA	NA	NA	588,29	NA	NA	-0,70	Hypp14514	1	No	No
13	BL25489	NA	NA	NA	NA	94,99	NA	NA	-0,51	67,16	NA	NA	-0,81	ADRA1D	0	No	No
13	BL29313	NA	NA	NA	NA	NA	NA	NA	NA	105,66	NA	NA	-0,61	Hypp13153	1	No	No
13	BL29509	NA	NA	NA	NA	NA	NA	NA	NA	138,51	NA	NA	-0,80	Hypp5262	0	No	No
13	BL31577	NA	NA	NA	NA	91,57	NA	NA	-0,62	65,06	NA	NA	-0,63	Hypp4301	0	No	No
13	BL31705	NA	NA	NA	NA	NA	NA	NA	NA	697,96	0,65	NA	NA	Hypp3000	0	No	No
13	BL34788	NA	NA	NA	NA	231,64	NA	NA	-0,40	77,14	0,80	-0,78	-0,97	ALDH1A2	0	No	No
13	BL36756	NA	NA	NA	NA	NA	NA	NA	NA	172,15	NA	NA	-0,64	NPTN	1	No	No
13	BL41904	NA	NA	NA	NA	54,15	NA	NA	-0,44	75,70	NA	NA	-0,69	Hypnc8502	0	No	No
13	BL42782	NA	NA	NA	NA	NA	NA	NA	NA	114,95	NA	NA	-0,69	RASD2	0	No	No
13	BL46103	NA	NA	NA	NA	NA	NA	NA	NA	409,65	NA	NA	-0,61	Hypnc12210	0	No	No
13	BL49375	NA	NA	NA	NA	28,60	NA	NA	-0,54	170,00	NA	NA	-0,69	Hypnc15073	0	No	No
13	BL49437	NA	NA	NA	NA	NA	NA	NA	NA	83,25	NA	NA	-0,77	Hypnc15126	0	No	No
13	BL53858	NA	NA	NA	NA	NA	NA	NA	NA	54,82	NA	NA	-0,71	Hypnc19010	2	No	No
13	BL54595	NA	NA	NA	NA	281,86	NA	NA	-0,29	211,62	NA	NA	-0,74	Hypnc19663	0	No	No
13	BL54649	NA	NA	NA	NA	NA	NA	NA	NA	104,97	NA	NA	-0,76	Hypnc19712	1	No	No
13	BL63406	NA	NA	NA	NA	717,86	NA	NA	-0,28	754,64	NA	NA	-0,66	Hypp2695	1	No	No
13	BL66322	NA	NA	NA	NA	202,74	NA	NA	-0,60	NA	NA	NA	NA	TMEM22	8	No	No
13	BL74719	NA	NA	NA	NA	NA	NA	NA	NA	81,30	NA	NA	-0,68	Hypnc37362	1	No	No
13	BL75359	NA	NA	NA	NA	464,87	0,26	NA	NA	272,72	NA	NA	-0,59	Hypnc37898	1	No	No
13	BL79956	NA	NA	NA	NA	NA	NA	NA	NA	91,58	NA	NA	-0,68	Hypnc41908	0	No	No
13	BL87313	NA	NA	NA	NA	58,59	NA	NA	-0,43	377,59	NA	NA	-0,66	GANAB	0	No	No
13	BL88421	NA	NA	NA	NA	82,56	NA	NA	-0,48	69,59	0,67	NA	-0,95	Hypnc49331	1	No	No
13	BL90262	NA	NA	NA	NA	448,10	0,41	NA	-0,64	NA	NA	NA	NA	ZNF782	2	Yes	Yes
13	BL91402	NA	NA	NA	NA	NA	NA	NA	NA	142,27	NA	NA	-0,73	Hypnc51960	0	No	No
13	BL95187	NA	NA	NA	NA	99,35	NA	NA	-0,71	NA	NA	NA	NA	ENSG00000262186	2	No	No
13	BL95190	NA	NA	NA	NA	142,02	NA	NA	-0,60	69,59	NA	NA	-0,56	NCAM1	3	No	No
13	BL95474	NA	NA	NA	NA	NA	NA	NA	NA	237,15	NA	NA	-0,59	SLC1A3	2	No	No
13	BL95759	NA	NA	NA	NA	NA	NA	NA	NA	5248,29	NA	NA	-0,75	CNTN6	7	No	No
13	BL95881	NA	NA	NA	NA	NA	NA	NA	NA	64,20	NA	NA	-0,78	ASMTL	3	No	No
13	BL96181	NA	NA	NA	NA	NA	NA	NA	NA	190,84	NA	NA	-0,67	USH1C	0	No	No
13	BL97370	NA	NA	NA	NA	55,69	0,61	NA	NA	NA	NA	NA	NA	EDIL3	0	No	No
14	BL00452	NA	NA	NA	NA	51,87	NA	NA	0,79	87,35	NA	0,82	0,74	Hypp12273	6	No	Yes
14	BL02233	NA	NA	NA	NA	38,68	NA	NA	1,01	136,01	NA	0,79	1,12	Hypp12931	2	No	No
14	BL02684	NA	NA	NA	NA	989,96	NA	0,43	0,51	1393,50	NA	0,93	NA	PTCHD1	2	No	No
14	BL02758	NA	NA	NA	NA	51,18	NA	NA	0,53	125,91	NA	1,08	0,56	Hypp4826	1	No	No
14	BL03652	41,39	NA	NA	0,29	51,02	NA	NA	1,21	20,05	NA	1,00	1,06	FAM55C	0	No	No
14	BL03708	25,67	NA	NA	0,49	70,61	NA	NA	0,71	111,12	NA	1,56	1,14	RDH12	0	No	No
14	BL03848	NA	NA	NA	NA	686,53	-0,27	NA	0,89	1003,85	-0,55	0,64	1,04	ECHS1	1	No	No
14	BL04010	NA	NA	NA	NA	257,81	-0,36	NA	0,82	967,28	NA	1,01	0,93	Hypp15739	0	No	No
14	BL05107	NA	NA	NA	NA	886,54	-0,34	NA	0,64	3173,87	-0,58	0,64	0,81	SELENBP1	5	No	Yes
14	BL06494	NA	NA	NA	NA	46,45	NA	NA	1,10	344,19	NA	1,09	1,70	CYP24A1	0	No	No
14	BL06645	NA	NA	NA	NA	86,15	NA	NA	0,99	386,11	NA	1,11	0,93	CES2	0	No	No
14	BL06895	52,63	NA	NA	0,18	50,18	NA	NA	0,80	56,94	NA	0,93	0,88	CYP2U1	0	No	No
14	BL07502	NA	NA	NA	NA	1229,47	NA	NA	0,63	3319,19	-0,34	0,82	1,03	PRDX6	1	No	No
14	BL07902	NA	NA	NA	NA	17,07	NA	NA	0,57	84,73	NA	1,09	1,01	PNLIP	1	No	No
14	BL09036	NA	NA	NA	NA	63,66	NA	NA	0,72	150,65	NA	0,99	0,95	NMRAL1	1	No	No

		N0				N3				N4						
Cluster	Gene	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS4934	l2FC R1158667	l2FC RA	Auto nomenclature	N.º of RAREs	ATACseq peaks	RARE hubs
14	BL09971	NA	NA	NA	NA	NA	NA	NA	NA	410,11	NA	1,67	0,66	Hypp7248	0	No	No
14	BL09980	271,37	NA	0,41	0,47	216,90	NA	0,64	0,50	249,16	NA	1,47	0,73	MAOB	0	No	No
14	BL10340	NA	NA	NA	NA	NA	NA	NA	NA	118,47	-0,91	2,17	1,86	FABP2	3	No	No
14	BL11139	NA	NA	NA	NA	51,03	NA	NA	0,68	92,65	NA	0,95	1,24	LPHN3	0	No	No
14	BL11230	NA	NA	NA	NA	286,61	NA	NA	0,57	382,76	NA	0,69	1,29	ABCC4	0	No	No
14	BL13268	542,70	NA	0,29	0,50	3172,53	-0,29	0,62	0,70	4263,64	NA	1,30	0,65	ABCB1	0	No	No
14	BL13899	NA	NA	NA	NA	144,53	NA	0,47	NA	386,58	NA	1,47	0,79	HSD3B7	0	No	No
14	BL14133	NA	NA	NA	NA	50,19	NA	NA	1,00	110,39	NA	1,20	0,98	UGT1A1	2	No	No
14	BL15206	NA	NA	NA	NA	32,10	NA	NA	1,35	104,50	NA	1,13	1,13	HPGDS	2	No	No
14	BL15370	NA	NA	NA	NA	322,25	-0,37	NA	0,79	421,85	-0,36	0,74	0,84	TMEM56	3	No	No
14	BL15878	NA	NA	NA	NA	187,54	NA	NA	0,51	280,99	NA	1,10	1,03	TPMT	1	No	No
14	BL15892	NA	NA	NA	NA	97,21	NA	NA	0,41	205,02	NA	1,39	0,62	MOXD1	3	No	Yes
14	BL16369	11,94	NA	NA	0,25	57,10	NA	0,93	1,36	118,52	NA	0,88	NA	Hypp1744	3	No	No
14	BL16685	NA	NA	NA	NA	959,69	-0,32	0,27	0,64	2029,73	NA	1,44	0,83	NR1H4	3	No	No
14	BL16863	NA	NA	NA	NA	55,36	NA	0,53	0,52	95,34	NA	1,12	1,03	Hypp10509	2	No	No
14	BL17058	NA	NA	NA	NA	12,55	NA	NA	0,92	55,76	NA	2,64	2,03	EPHX1	0	No	No
14	BL17260	NA	NA	NA	NA	28,44	NA	NA	0,87	67,32	NA	2,20	1,91	EHHADH	2	No	No
14	BL17347	NA	NA	NA	NA	578,87	NA	0,89	0,52	2394,85	NA	2,04	NA	CYP2U1	6	Yes	No
14	BL17424	NA	NA	NA	NA	217,30	NA	NA	0,67	335,03	NA	0,78	0,97	CYP46A1	2	No	No
14	BL17665	NA	NA	NA	NA	598,40	NA	0,44	0,66	1809,45	NA	1,11	NA	CYP3A4	2	No	No
14	BL17761	NA	NA	NA	NA	90,96	NA	NA	1,81	171,79	NA	0,85	1,31	IGFALS	0	No	No
14	BL19412	NA	NA	NA	NA	NA	NA	NA	NA	87,99	NA	2,02	1,91	SLCO4A1	1	No	No
14	BL19504	NA	NA	NA	NA	80,46	NA	NA	0,74	98,95	NA	0,73	0,81	DDC	10	No	Yes
14	BL19641	NA	NA	NA	NA	894,17	-0,58	NA	0,46	1420,28	-0,55	0,92	0,78	Hypp10513	4	No	No
14	BL19778	NA	NA	NA	NA	119,61	NA	NA	0,80	195,52	NA	0,72	0,90	CRABP2	2	No	No
14	BL20613	NA	NA	NA	NA	184,43	-0,67	NA	1,02	286,34	NA	1,09	1,12	ISOC2	0	No	No
14	BL20855	NA	NA	NA	NA	106,07	NA	0,44	0,58	289,19	NA	1,43	0,71	MOXD1	1	No	No
14	BL21215	NA	NA	NA	NA	46,54	NA	NA	0,54	315,34	NA	0,66	1,10	Hypp10510	2	No	No
14	BL24259	NA	NA	NA	NA	183,70	NA	NA	1,38	593,08	NA	0,94	0,89	BHLHE40	4	Yes	No
14	BL24666	NA	NA	NA	NA	21,00	NA	NA	0,57	95,65	NA	1,06	1,14	IRF2	3	No	Yes
14	BL31561	NA	NA	NA	NA	16,09	NA	NA	1,09	206,56	NA	2,17	1,02	CREBBP	1	No	No
14	BL39607	NA	NA	NA	NA	181,53	-0,40	NA	0,91	319,88	NA	0,81	1,07	Hypp2249	2	No	No
14	BL41527	NA	NA	NA	NA	201,84	NA	NA	0,35	619,44	NA	1,25	0,59	Hypnc8170	1	No	No
14	BL48592	NA	NA	NA	NA	265,79	NA	NA	0,56	261,61	NA	1,61	0,87	ABCA1	3	No	No
14	BL54118	NA	NA	NA	NA	52,93	NA	NA	0,53	161,25	NA	1,25	0,76	Hypnc19238	2	No	Yes
14	BL59652	72,10	NA	0,17	0,23	130,75	NA	0,66	0,59	171,01	NA	1,19	0,75	Hypnc24141	1	No	No
14	BL60579	NA	NA	NA	NA	192,32	NA	1,16	0,80	783,62	NA	1,82	NA	CYP1B1	2	Yes	No
14	BL64079	NA	NA	NA	NA	37,21	NA	NA	0,64	66,53	NA	1,22	1,31	Hypnc28016	2	No	No
14	BL65209	NA	NA	NA	NA	52,72	NA	NA	0,57	57,89	NA	NA	2,66	Hypp15273	1	No	No
14	BL72668	NA	NA	NA	NA	52,93	NA	NA	0,66	116,30	NA	1,30	1,20	Hypnc35559	2	No	No
14	BL95198	NA	NA	NA	NA	73,26	NA	0,68	NA	142,86	NA	1,75	0,75	KCNMB2	2	No	No
14	BL95209	NA	NA	NA	NA	265,77	NA	1,19	0,59	656,34	NA	1,92	0,70	CYP3A5	4	No	No
14	BL95377	NA	NA	NA	NA	27,99	NA	NA	0,81	53,78	NA	0,95	1,62	FAM55C	0	No	No
14	BL95503	36,13	NA	NA	0,37	87,07	NA	0,58	0,81	119,48	NA	1,36	0,96	UGT2B4	1	Yes	No
14	BL95696	NA	NA	NA	NA	62,86	NA	NA	0,73	111,23	NA	1,00	1,36	PTGR1	5	No	No
14	BL95758	NA	NA	NA	NA	103,89	NA	NA	0,54	151,99	NA	0,55	0,93	CNTN1	1	No	No
14	BL96008	NA	NA	NA	NA	67,58	NA	NA	0,54	259,95	NA	1,11	0,90	AKR1A1	1	No	No
14	BL96466	NA	NA	NA	NA	84,84	NA	NA	0,87	256,95	NA	1,13	0,95	ACHE	3	No	No
14	BL96467	NA	NA	NA	NA	39,51	NA	NA	1,09	183,36	NA	1,44	1,49	CES3	0	No	No
14	BL96471	NA	NA	NA	NA	87,41	NA	0,78	0,74	343,46	NA	1,21	0,64	CES1	3	No	Yes
14	BL96766	NA	NA	NA	NA	357,43	NA	0,74	0,79	649,32	NA	1,11	NA	CORIN	0	No	No
14	BL96767	NA	NA	NA	NA	483,02	NA	0,77	0,79	885,45	NA	1,41	NA	FBN1	1	No	No
14	BL96768	NA	NA	NA	NA	205,02	NA	0,77	0,70	373,76	NA	1,35	0,52	MUC4	1	No	No
14	BL96769	NA	NA	NA	NA	733,74	NA	0,80	0,47	1482,13	0,52	1,59	0,96	MATN2	3	No	No
14	BL97070	NA	NA	NA	NA	72,37	NA	NA	0,49	114,57	NA	0,71	1,04	ABCC4	3	Yes	No
15	BL00604	NA	NA	NA	NA	NA	NA	NA	NA	463,73	NA	0,57	0,71	F11R	1	No	No
15	BL00693	NA	NA	NA	NA	NA	NA	NA	NA	140,15	0,75	1,48	1,41	Hypp13133	0	No	No
15	BL00801	NA	NA	NA	NA	NA	NA	NA	NA	193,33	NA	1,12	1,19	Hypp13487	1	No	No
15	BL01444	NA	NA	NA	NA	NA	NA	NA	NA	890,37	NA	0,37	0,67	IGF2	7	Yes	No
15	BL01507	NA	NA	NA	NA	NA	NA	NA	NA	11384,38	NA	NA	0,75	Hypp488	0	No	No
15	BL01664	NA	NA	NA	NA	NA	NA	NA	NA	193,66	NA	0,71	0,44	CNRIP1	2	No	No
15	BL02166	NA	NA	NA	NA	NA	NA	NA	NA	202,44	NA	NA	0,79	Hypp5994	0	No	No
15	BL02351	NA	NA	NA	NA	NA	NA	NA	NA	289,08	NA	1,26	0,76	ARRDC3	1	Yes	No
15	BL02423	NA	NA	NA	NA	NA	NA	NA	NA	72,64	NA	NA	0,60	TNR	0	No	No
15	BL02558	NA	NA	NA	NA	NA	NA	NA	NA	87,33	NA	NA	0,74	ST8SIA2	1	No	No
15	BL03537	NA	NA	NA	NA	NA	NA	NA	NA	205,78	NA	1,22	1,64	Hypp14906	2	No	No
15	BL03893	NA	NA	NA	NA	5741,28	NA	NA	0,25	4781,62	NA	0,69	0,39	FBP1	2	No	No
15	BL04070	NA	NA	NA	NA	NA	NA	NA	NA	1929,49	0,51	0,62	0,42	ADPGK	0	No	No
15	BL04104	NA	NA	NA	NA	NA	NA	NA	NA	189,70	NA	NA	0,63	JMJD8	2	No	No
15	BL05308	NA	NA	NA	NA	NA	NA	NA	NA	968,65	0,58	0,76	0,58	GSR	0	No	No
15	BL05345	NA	NA	NA	NA	NA	NA	NA	NA	890,35	0,54	0,62	0,54	CTSZ	0	No	No
15	BL05930	NA	NA	NA	NA	NA	NA	NA	NA	366,35	NA	NA	0,85	Hypp12940	0	No	No
15	BL06121	NA	NA	NA	NA	NA	NA	NA	NA	108,95	NA	NA	0,73	Hypp13708	0	No	No
15	BL06552	NA	NA	NA	NA	NA	NA	NA	NA	163,19	NA	0,57	1,30	SERPINB1	0	No	No
15	BL07392	NA	NA	NA	NA	NA	NA	NA	NA	55,87	NA	0,88	0,67	IL17RD	6	No	No
15	BL07565	NA	NA	NA	NA	NA	NA	NA	NA	193,39	NA	0,83	0,50	Hypp1844	2	No	No
15	BL08628	NA	NA	NA	NA	NA	NA	NA	NA	66,03	NA	1,11	1,18	NDUFV2	0	No	No
15	BL08745	NA	NA	NA	NA	NA	NA	NA	NA	891,67	NA	0,67	0,69	BAHD1	2	No	No
15	BL08878	NA	NA	NA	NA	NA	NA	NA	NA	412,98	0,77	0,95	0,73	RP11-1280I22.1	0	No	No
15	BL10004	NA	NA	NA	NA	NA	NA	NA	NA	64,20	NA	NA	0,62	Hypp10646	0	No	No
15	BL10194	NA	NA	NA	NA	NA	NA	NA	NA	87,77	NA	NA	0,73	MCFD2	1	No	No
15	BL10300	NA	NA	NA	NA	NA	NA	NA	NA	55,17	NA	NA	0,88	Hypp14839	4	No	No
15	BL10773	NA	NA	NA	NA	NA	NA	NA	NA	60,41	NA	NA	1,29	OIT3	0	No	No
15	BL11260	NA	NA	NA	NA	9206,92	NA	NA	0,22	9946,53	NA	0,65	0,37	ARRDC2	4	No	No
15	BL11324	NA	NA	NA	NA	NA	NA	NA	NA	305,72	NA	0,83	0,55	HAGH	6	No	No
15	BL12116	NA	NA	NA	NA	NA	NA	NA	NA	220,29	0,62	0,67	0,63	Hypp14051	1	No	No
15	BL12561	NA	NA	NA	NA	NA	NA	NA	NA	82,07	NA	NA	0,84	UGT2B7	0	No	No
15	BL13734	NA	NA	NA	NA	NA	NA	NA	NA	68,18	NA	1,10	0,88	IGF1	9	No	Yes
15	BL14110	NA	NA	NA	NA	NA	NA	NA	NA	1204,15	NA	0,68	0,42	SLC25A28	2	No	No
15	BL14191	NA	NA	NA	NA	NA	NA	NA	NA	58,54	NA	NA	0,65	CYP4F2	0	No	No
15	BL14294	NA	NA	NA	NA	846,97	0,29	0,32	NA	1676,05	NA	0,63	0,42	F5	0	No	No
15	BL16202	NA	NA	NA	NA	NA	NA	NA	NA	51,55	NA	NA	1,10	ENSG00000262654	0	No	No
15	BL16636	NA	NA	NA	NA	NA	NA	NA	NA	617,45	NA	0,70	0,56	GSTO1	0	No	No
15	BL16736	NA	NA	NA	NA	NA	NA	NA	NA	103,65	NA	0,99	0,96	RDH13	1	No	No
15	BL17450	NA	NA	NA	NA	NA	NA	NA	NA	68,36	NA	NA	0,89	TSPAN12	0	Yes	No
15	BL17833	NA	NA	NA	NA	NA	NA	NA	NA	158,50	NA	NA	0,65	BDH1	0	No	No
15	BL19269	NA	NA	NA	NA	NA	NA	NA	NA	85,33	NA	0,58	0,63	MYO1C	0	No	No
15	BL19313	NA	NA	NA	NA	NA	NA	NA	NA	86,01	NA	0,66	0,87	CTHRC1	3	No	No
15	BL19527	NA	NA	NA	NA	NA	NA	NA	NA	98,98	NA	0,56	0,98	PLSCR1	2	No	No
15	BL19717	NA	NA	NA	NA	NA	NA	NA	NA	268,50	NA	NA	0,64	Hypp10322	1	No	No
15	BL20027	NA	NA	NA	NA	NA	NA	NA	NA	481,03	NA	0,59	0,47	TCTEX1D1	2	No	Yes
15	BL23168	NA	NA	NA	NA	NA	NA	NA	NA	58,50	NA	NA	0,66	KLHL30	2	No	No
15	BL23370	NA	NA	NA	NA	NA	NA	NA	NA	1099,93	0,72	1,05	0,66	Hypp10637	1	No	No
15	BL23390	NA	NA	NA	NA	NA	NA	NA	NA	842,31	0,48	0,59	0,38	PPPDE2	4	No	No
15	BL23667	NA	NA	NA	NA	NA	NA	NA	NA	175,59	NA	NA	0,69	SLIT2	5	No	No
15	BL23957	NA	NA	NA	NA	NA	NA	NA	NA	95,79	NA	NA	0,59	Hypp4004	1	No	No
15	BL24458	NA	NA	NA	NA	NA	NA	NA	NA	159,90	NA	NA	0,74	SYT9	0	No	No
15	BL32678	NA	NA	NA	NA	NA	NA	NA	NA	64,75	NA	NA	0,84	Hypnc6091	0	No	No
15	BL37204	NA	NA	NA	NA	192,09	0,44	0,50	NA	221,13	NA	0,67	0,69	Hypnc4393	0	No	No
15	BL39173	NA	NA	NA	NA	NA	NA	NA	NA	57,12	NA	NA	0,99	Hypnc6118	0	No	No

		N0				N3				N4							
Cluster	Gene	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS4934	l2FC R1158667	l2FC RA	Auto nomenclature	N.º of RAREs	ATACseq peaks	RARE hubs
15	BL40421	NA	NA	NA	NA	184,92	NA	0,52	NA	184,69	NA	0,69	0,55	Hypp2529	0	No	No
15	BL40854	NA	NA	NA	NA	NA	NA	NA	NA	173,90	NA	NA	0,59	Hypnc7580	0	No	No
15	BL43394	NA	NA	NA	NA	NA	NA	NA	NA	71,11	NA	0,70	0,66	Hypnc9786	2	Yes	No
15	BL44639	NA	NA	NA	NA	NA	NA	NA	NA	138,58	NA	NA	0,73	Hypnc10892	0	No	No
15	BL46286	NA	NA	NA	NA	NA	NA	NA	NA	109,52	NA	0,67	1,23	Hypnc12365	0	No	No
15	BL47381	NA	NA	NA	NA	NA	NA	NA	NA	94,24	NA	NA	0,75	Hypnc13293	1	No	No
15	BL51198	NA	NA	NA	NA	NA	NA	NA	NA	284,79	NA	NA	0,59	Hypnc16686	0	No	No
15	BL56601	NA	NA	NA	NA	NA	NA	NA	NA	118,44	NA	NA	0,65	Hypnc21445	0	No	No
15	BL56990	NA	NA	NA	NA	NA	NA	NA	NA	57,86	0,87	0,91	0,74	Hypnc21784	0	No	No
15	BL58088	NA	NA	NA	NA	NA	NA	NA	NA	97,39	NA	0,60	0,58	Hypp2120	2	No	No
15	BL61169	NA	NA	NA	NA	NA	NA	NA	NA	99,05	NA	NA	0,59	Hypnc25481	10	No	Yes
15	BL61343	NA	NA	NA	NA	NA	NA	NA	NA	75,56	NA	NA	1,01	Hypp14110	1	No	No
15	BL63537	NA	NA	NA	NA	NA	NA	NA	NA	71,57	NA	0,90	1,02	Hypp13162	1	No	No
15	BL65038	NA	NA	NA	NA	NA	NA	NA	NA	70,49	NA	1,03	1,19	Hypnc28838	2	No	No
15	BL65638	NA	NA	NA	NA	NA	NA	NA	NA	162,95	NA	NA	0,89	LRRC15	0	No	No
15	BL70799	NA	NA	NA	NA	118,36	0,41	NA	NA	112,80	NA	0,64	0,62	Hypnc33905	1	No	No
15	BL72429	NA	NA	NA	NA	NA	NA	NA	NA	172,70	NA	NA	0,62	Hypnc35342	1	No	No
15	BL90874	NA	NA	NA	NA	NA	NA	NA	NA	74,60	NA	NA	0,64	Hypp11805	1	No	No
15	BL91406	NA	NA	NA	NA	NA	NA	NA	NA	212,85	NA	0,45	0,73	Hypnc51964	0	No	No
15	BL95123	NA	NA	NA	NA	NA	NA	NA	NA	192,52	NA	0,57	0,66	HPGDS	2	No	No
15	BL95208	NA	NA	NA	NA	NA	NA	NA	NA	106,51	NA	1,36	1,04	CYP3A7	4	No	No
15	BL95480	NA	NA	NA	NA	NA	NA	NA	NA	52,50	NA	0,81	1,10	EPDR1	0	No	No
15	BL95784	NA	NA	NA	NA	NA	NA	NA	NA	139,56	NA	1,10	0,63	CHIT1	2	No	No
15	BL95795	NA	NA	NA	NA	NA	NA	NA	NA	102,76	NA	1,15	0,92	GDAP1	0	No	No
15	BL96101	NA	NA	NA	NA	NA	NA	NA	NA	125,69	NA	NA	0,82	BCMO1	0	No	No
15	BL96154	NA	NA	NA	NA	NA	NA	NA	NA	182,39	0,93	1,39	1,04	CBR1	2	No	No
15	BL96185	NA	NA	NA	NA	NA	NA	NA	NA	66,19	NA	0,84	0,73	LPHN2	5	No	No
15	BL97223	NA	NA	NA	NA	NA	NA	NA	NA	107,20	NA	NA	0,69	LAMA1	4	No	No
15	BL97259	NA	NA	NA	NA	NA	NA	NA	NA	85,51	NA	1,53	1,21	ENSG00000262243	7	No	No
16	BL00843	NA	NA	NA	NA	NA	NA	NA	NA	53,58	NA	0,76	NA	VWA5A	1	No	No
16	BL00852	NA	NA	NA	NA	NA	NA	NA	NA	103,27	NA	1,56	NA	PTCHD1	0	No	No
16	BL00889	NA	NA	NA	NA	NA	NA	NA	NA	94,44	NA	1,38	NA	LAMA2	0	No	No
16	BL01311	NA	NA	NA	NA	NA	NA	NA	NA	156,64	0,53	0,62	NA	PKHD1L1	0	No	No
16	BL01315	NA	NA	NA	NA	NA	NA	NA	NA	114,01	0,68	0,95	NA	RHOB	0	No	No
16	BL01586	NA	NA	NA	NA	644,20	NA	0,32	NA	689,91	NA	0,69	NA	KY	3	No	No
16	BL01741	NA	NA	NA	NA	NA	NA	NA	NA	250,99	0,77	0,73	NA	Hypp14620	0	No	No
16	BL02073	NA	NA	NA	NA	NA	NA	NA	NA	8264,93	NA	0,68	NA	STOM	1	No	No
16	BL02218	NA	NA	NA	NA	NA	NA	NA	NA	85,44	NA	0,59	NA	Hypp15033	2	No	No
16	BL02265	NA	NA	NA	NA	NA	NA	NA	NA	605,18	NA	0,78	NA	Hypp4440	1	No	No
16	BL02324	NA	NA	NA	NA	NA	NA	NA	NA	64,98	NA	0,81	NA	TBK1	6	No	Yes
16	BL02468	NA	NA	NA	NA	NA	NA	NA	NA	1441,96	NA	0,65	NA	PTGS2	3	No	Yes
16	BL02847	NA	NA	NA	NA	NA	NA	NA	NA	523,22	0,83	1,10	NA	C5orf41	1	No	No
16	BL02869	NA	NA	NA	NA	NA	NA	NA	NA	66,89	1,15	0,80	NA	TRPA1	1	No	No
16	BL02910	NA	NA	NA	NA	NA	NA	NA	NA	386,37	NA	0,99	NA	LZTR1	2	No	No
16	BL03267	NA	NA	NA	NA	NA	NA	NA	NA	197,21	NA	2,18	NA	HPGDS	0	No	No
16	BL03990	NA	NA	NA	NA	NA	NA	NA	NA	93,31	NA	2,25	NA	CYP4F12	3	No	No
16	BL04068	NA	NA	NA	NA	NA	NA	NA	NA	1345,57	NA	0,64	NA	Hypp2622	0	No	No
16	BL04772	NA	NA	NA	NA	669,76	NA	0,26	0,22	503,31	0,64	0,63	NA	LRRK2	0	No	No
16	BL04824	NA	NA	NA	NA	NA	NA	NA	NA	404,06	NA	1,99	NA	Hypp13500	0	No	No
16	BL05149	NA	NA	NA	NA	NA	NA	NA	NA	1167,39	NA	0,76	NA	ATF3	1	No	No
16	BL05241	NA	NA	NA	NA	723,97	NA	NA	0,25	653,39	0,65	0,89	NA	TLL2	1	No	No
16	BL05692	NA	NA	NA	NA	177,96	NA	0,42	NA	197,19	NA	0,67	NA	PKHD1L1	0	No	No
16	BL05720	NA	NA	NA	NA	NA	NA	NA	NA	969,26	0,77	0,87	NA	MAFG	1	No	No
16	BL06073	NA	NA	NA	NA	NA	NA	NA	NA	101,28	NA	0,66	NA	PKHD1L1	0	No	No
16	BL06090	NA	NA	NA	NA	NA	NA	NA	NA	718,04	NA	0,76	NA	ENSG00000262552	1	No	No
16	BL06571	NA	NA	NA	NA	NA	NA	NA	NA	52,00	NA	1,05	NA	MOXD1	3	No	Yes
16	BL07187	NA	NA	NA	NA	NA	NA	NA	NA	2196,41	NA	0,61	NA	TXNRD1	0	No	No
16	BL07377	NA	NA	NA	NA	5232,98	NA	NA	0,19	7526,12	NA	0,71	NA	PCK2	1	No	No
16	BL08115	NA	NA	NA	NA	NA	NA	NA	NA	60,10	1,02	0,95	NA	TRPC5	1	No	No
16	BL08699	NA	NA	NA	NA	NA	NA	NA	NA	1217,77	0,55	0,61	NA	ARG2	2	No	Yes
16	BL08838	NA	NA	NA	NA	NA	NA	NA	NA	142,44	NA	0,59	NA	C2orf84	0	No	No
16	BL08892	NA	NA	NA	NA	NA	NA	NA	NA	1881,30	NA	1,39	NA	Hypp13755	2	No	No
16	BL09241	NA	NA	NA	NA	NA	NA	NA	NA	110,94	NA	1,19	NA	ARRDC3	1	No	No
16	BL10078	NA	NA	NA	NA	NA	NA	NA	NA	138,92	NA	0,81	NA	CYP3A4	2	No	No
16	BL10125	NA	NA	NA	NA	NA	NA	NA	NA	164,66	0,79	0,72	NA	BPNT1	0	No	No
16	BL10242	NA	NA	NA	NA	NA	NA	NA	NA	255,98	NA	0,60	NA	FLNC	2	No	No
16	BL10503	NA	NA	NA	NA	NA	NA	NA	NA	794,73	NA	0,86	NA	Hypp16074	4	No	No
16	BL10938	NA	NA	NA	NA	NA	NA	NA	NA	621,51	NA	0,80	NA	CBS	2	No	No
16	BL11089	NA	NA	NA	NA	NA	NA	NA	NA	321,03	0,69	0,69	NA	ABCB6	1	No	No
16	BL11843	NA	NA	NA	NA	NA	NA	NA	NA	275,23	0,89	0,91	NA	ENSG00000262489	0	No	No
16	BL13020	NA	NA	NA	NA	NA	NA	NA	NA	72,05	NA	0,97	NA	Hypp1350	0	No	No
16	BL13079	NA	NA	NA	NA	NA	NA	NA	NA	925,26	NA	0,61	NA	CD63	1	No	No
16	BL13821	NA	NA	NA	NA	3651,95	NA	0,27	0,28	3012,39	NA	0,62	NA	ALAS1	1	No	No
16	BL14000	NA	NA	NA	NA	NA	NA	NA	NA	169,86	0,59	0,83	NA	RETSAT	2	Yes	No
16	BL14102	NA	NA	NA	NA	NA	NA	NA	NA	60,78	NA	0,64	NA	EFCAB4B	2	No	No
16	BL14198	NA	NA	NA	NA	472,29	NA	NA	0,31	792,90	-0,38	1,10	NA	Hypp10976	2	No	No
16	BL14290	NA	NA	NA	NA	996,06	NA	0,39	NA	12424,00	NA	1,19	NA	HPGDS	0	No	No
16	BL14473	861,16	NA	0,25	NA	908,40	0,30	0,40	NA	1028,92	NA	0,65	NA	Hypp8461	1	No	No
16	BL14775	NA	NA	NA	NA	NA	NA	NA	NA	59,90	NA	1,07	NA	Hypp5344	2	No	No
16	BL15002	NA	NA	NA	NA	NA	NA	NA	NA	389,86	NA	0,66	NA	GLOD4	3	No	No
16	BL15241	NA	NA	NA	NA	NA	NA	NA	NA	1093,95	NA	0,69	NA	SOCS2	0	No	No
16	BL15246	NA	NA	NA	NA	NA	NA	NA	NA	100,07	NA	1,06	NA	ALDH2	0	No	No
16	BL15461	NA	NA	NA	NA	NA	NA	NA	NA	889,14	0,55	0,69	NA	KIAA1239	3	No	No
16	BL15811	NA	NA	NA	NA	NA	NA	NA	NA	360,86	0,66	1,16	NA	CYP3A4	2	No	No
16	BL16259	NA	NA	NA	NA	NA	NA	NA	NA	74,34	NA	0,65	NA	Hypp14621	1	No	No
16	BL16924	NA	NA	NA	NA	NA	NA	NA	NA	494,54	0,76	0,82	NA	HTATIP2	4	No	No
16	BL17037	NA	NA	NA	NA	NA	NA	NA	NA	257,08	NA	1,65	NA	HPGDS	0	No	No
16	BL17799	NA	NA	NA	NA	NA	NA	NA	NA	94,33	NA	1,12	NA	ARSB	0	No	No
16	BL18100	1732,98	NA	0,30	NA	1852,69	NA	0,49	NA	2197,72	NA	0,68	NA	Hypp8215	1	No	No
16	BL18167	NA	NA	NA	NA	NA	NA	NA	NA	1598,41	NA	0,61	NA	Hypp1133	0	No	No
16	BL18339	NA	NA	NA	NA	NA	NA	NA	NA	326,50	0,75	1,53	NA	Hypp849	1	No	No
16	BL18342	NA	NA	NA	NA	NA	NA	NA	NA	273,07	-0,61	0,83	NA	ACCN1	2	No	No
16	BL18544	NA	NA	NA	NA	NA	NA	NA	NA	1690,72	NA	0,62	NA	KLF5	2	No	No
16	BL19110	NA	NA	NA	NA	NA	NA	NA	NA	107,52	NA	1,44	NA	Hypp11526	3	No	No
16	BL19192	NA	NA	NA	NA	NA	NA	NA	NA	450,51	NA	0,61	NA	TTLL8	0	No	No
16	BL19325	NA	NA	NA	NA	NA	NA	NA	NA	184,55	NA	0,59	NA	PKHD1L1	0	No	No
16	BL19382	NA	NA	NA	NA	NA	NA	NA	NA	164,92	NA	1,48	NA	Hypp11524	2	No	No
16	BL19449	NA	NA	NA	NA	NA	NA	NA	NA	504,22	NA	0,73	NA	Hypp4829	1	No	No
16	BL20332	1189,00	NA	0,30	NA	886,32	0,35	0,46	NA	937,31	0,65	0,71	NA	PKHD1L1	2	No	No
16	BL20333	NA	NA	NA	NA	1275,78	NA	0,28	NA	1250,53	0,67	0,61	NA	PKHD1L1	1	No	No
16	BL20698	NA	NA	NA	NA	NA	NA	NA	NA	277,97	0,56	0,67	NA	GDPD1	1	No	No
16	BL20842	NA	NA	NA	NA	NA	NA	NA	NA	106,58	NA	0,64	NA	PLA2G16	1	No	No
16	BL21345	NA	NA	NA	NA	NA	NA	NA	NA	265,40	0,84	0,95	NA	Hypp1261	2	No	No
16	BL21355	NA	NA	NA	NA	NA	NA	NA	NA	777,31	0,79	0,90	NA	PIR	3	No	No
16	BL22079	NA	NA	NA	NA	2041,27	NA	0,20	0,18	2940,96	0,46	0,81	NA	ABCA1	4	No	No
16	BL22225	NA	NA	NA	NA	NA	NA	NA	NA	601,01	NA	0,88	NA	Hypp5857	1	No	No
16	BL22565	NA	NA	NA	NA	875,51	0,26	0,60	-0,42	1304,41	NA	0,71	NA	Hypp13501	1	No	No
16	BL22873	NA	NA	NA	NA	NA	NA	NA	NA	1319,38	0,84	1,02	NA	KEAP1	3	No	Yes
16	BL23044	NA	NA	NA	NA	NA	NA	NA	NA	104,67	0,69	0,66	NA	KCND3	0	No	No
16	BL23088	NA	NA	NA	NA	NA	NA	NA	NA	119,42	NA	0,86	NA	CCS	0	No	No

		N0				N3				N4							
Cluster	Gene	bM	I2FC BMS493	I2FC R115866	I2FC RA	bM	I2FC BMS493	I2FC R115866	I2FC RA	bM	I2FC BMS4934	I2FC R1158667	I2FC RA	Auto nomenclature	N.º of RAREs	ATACseq peaks	RARE hubs
16	BL23116	NA	NA	NA	NA	NA	NA	NA	NA	283,53	0,86	0,54	NA	PARP14	1	No	No
16	BL23430	NA	NA	NA	NA	NA	NA	NA	NA	495,83	NA	0,68	NA	GABRB3	0	No	No
16	BL23989	NA	NA	NA	NA	NA	NA	NA	NA	406,37	0,81	1,18	NA	CYP2J2	0	No	No
16	BL24008	NA	NA	NA	NA	NA	NA	NA	NA	454,35	0,49	0,65	NA	VWA5A	2	No	No
16	BL24796	NA	NA	NA	NA	NA	NA	NA	NA	51,45	NA	1,12	NA	CYP3A5	1	No	No
16	BL24897	NA	NA	NA	NA	NA	NA	NA	NA	745,15	NA	0,81	NA	CBR1	5	No	No
16	BL32106	NA	NA	NA	NA	NA	NA	NA	NA	59,34	NA	0,99	NA	CES2	1	No	No
16	BL32583	NA	NA	NA	NA	NA	NA	NA	NA	336,43	0,64	0,84	NA	Hypp12433	3	No	No
16	BL42250	NA	NA	NA	NA	NA	NA	NA	NA	134,90	NA	0,61	NA	Hypnc8796	0	No	No
16	BL43598	NA	NA	NA	NA	NA	NA	NA	NA	73,45	-0,66	0,62	NA	Hypnc9969	3	No	No
16	BL45120	NA	NA	NA	NA	NA	NA	NA	NA	314,52	NA	1,29	NA	Hypnc11322	1	No	No
16	BL51607	NA	NA	NA	NA	71,65	0,46	NA	NA	75,35	1,15	0,82	NA	Hypp14250	2	No	No
16	BL59483	NA	NA	NA	NA	NA	NA	NA	NA	303,83	0,64	1,41	NA	Hypnc24000	0	No	No
16	BL61707	NA	NA	NA	NA	1364,25	0,26	0,42	NA	145,33	NA	0,62	NA	Hypp1578	0	No	No
16	BL70184	NA	NA	NA	NA	NA	NA	NA	NA	87,43	NA	0,72	NA	Hypnc33372	1	No	No
16	BL83772	NA	NA	NA	NA	NA	NA	NA	NA	68,98	NA	0,61	NA	NR1H3	0	No	No
16	BL84268	NA	NA	NA	NA	NA	NA	NA	NA	97,44	NA	0,75	NA	Hypnc45667	1	No	No
16	BL95109	NA	NA	NA	NA	NA	NA	NA	NA	156,00	NA	0,65	NA	C20orf108	2	No	No
16	BL95242	NA	NA	NA	NA	NA	NA	NA	NA	51,09	NA	0,99	NA	HPGDS	2	No	No
16	BL95245	NA	NA	NA	NA	NA	NA	NA	NA	518,45	NA	0,84	NA	HPGDS	5	No	No
16	BL95361	NA	NA	NA	NA	NA	NA	NA	NA	153,25	0,82	0,71	NA	DENND3	4	No	No
16	BL95445	NA	NA	NA	NA	NA	NA	NA	NA	258,41	0,98	1,11	NA	ABHD4	4	No	No
16	BL95454	NA	NA	NA	NA	NA	NA	NA	NA	110,55	NA	0,61	NA	HSD11B1	3	Yes	No
16	BL95824	NA	NA	NA	NA	NA	NA	NA	NA	5113,52	NA	0,62	NA	ENSG00000262552	0	No	No
16	BL95955	NA	NA	NA	NA	NA	NA	NA	NA	5931,73	0,76	0,64	NA	PROM1	1	No	No
16	BL96118	NA	NA	NA	NA	NA	NA	NA	NA	122,17	0,69	1,09	NA	CES2	0	No	No
16	BL96155	NA	NA	NA	NA	NA	NA	NA	NA	80,36	1,07	1,23	NA	CBR1	2	No	No
16	BL96156	NA	NA	NA	NA	NA	NA	NA	NA	145,23	0,90	1,05	NA	CBR1	2	No	No
16	BL96436	NA	NA	NA	NA	NA	NA	NA	NA	87,98	0,70	0,70	NA	AC008735.1	3	No	No
16	BL96437	NA	NA	NA	NA	NA	NA	NA	NA	327,98	0,62	0,71	NA	ROBO1	3	No	Yes
16	BL96510	NA	NA	NA	NA	NA	NA	NA	NA	59,62	NA	0,94	NA	GSTP1	0	No	No
16	BL96951	NA	NA	NA	NA	NA	NA	NA	NA	50,37	NA	0,66	NA	Hypp13760	0	No	No
16	BL97429	NA	NA	NA	NA	NA	NA	NA	NA	366,04	NA	0,65	NA	DMBT1	0	No	No
16	BL97472	NA	NA	NA	NA	NA	NA	NA	NA	301,35	NA	0,80	NA	EML6	3	No	No

RA treatment - ATACseq peaks

Peak ID	Scaf	start	end	bM	I2FC	Assigned genes	RA N3 vs RA N4 vs ATAC	Motifs NR	Motifs FKH
Peak001	Sc0000000	1652494	1652794	77,45	1,43	BL06743, BL21261, BL57077, BL84925, BL84928	Yes	No	Yes
Peak002	Sc0000000	1655537	1655837	295,09	2,06	BL06743, BL21261, BL57077, BL84925, BL84928	Yes	Yes	No
Peak003	Sc0000000	1675973	1676273	206,68	0,68	BL24569, BL76331, BL76332, BL84923, BL84924, BL84926	No	No	No
Peak004	Sc0000000	1779064	1779364	231,11	1,30	BL20528, BL38784, BL38785, BL38786	No	No	No
Peak005	Sc0000000	1780097	1780397	69,58	0,89	BL20528, BL38785, BL38786	No	No	No
Peak006	Sc0000000	1785237	1785537	408,02	1,22		No	No	No
Peak007	Sc0000000	1786045	1786345	218,06	0,72		No	No	No
Peak008	Sc0000000	1786450	1786750	525,43	0,70		No	No	No
Peak009	Sc0000000	1788197	1788497	670,19	0,59		No	No	No
Peak010	Sc0000000	1827338	1827638	136,41	1,25	BL01409, BL14546, BL91552, BL91553, BL91555	Yes	Yes	Yes
Peak011	Sc0000000	1828151	1828451	700,43	1,42	BL01409, BL14546, BL91552, BL91553, BL91555	Yes	Yes	No
Peak012	Sc0000000	1833063	1833363	326,04	1,36	BL14546, BL91552, BL91555	Yes	No	No
Peak013	Sc0000000	1836697	1836997	650,96	0,60	BL14546, BL91552, BL91555	Yes	Yes	No
Peak014	Sc0000000	1837273	1837573	187,84	1,02	BL14546, BL91552, BL91555	Yes	Yes	Yes
Peak015	Sc0000000	1837805	1838105	225,91	0,79	BL14546, BL91552, BL91555	Yes	No	Yes
Peak016	Sc0000000	1844733	1845033	613,49	0,97	BL14546, BL91552, BL91555	Yes	No	No
Peak017	Sc0000000	1845124	1845424	216,03	1,04	BL14546, BL91552, BL91555	Yes	Yes	No
Peak018	Sc0000000	2373152	2373452	190,87	1,15	BL66618, BL66619, BL96576, BL96577	No	No	No
Peak019	Sc0000000	2373476	2373776	108,81	1,02	BL66618, BL66619, BL96577	No	No	No
Peak020	Sc0000000	2374503	2374803	95,02	1,00	BL66618, BL66619, BL96577	No	No	No
Peak021	Sc0000000	4708085	4708385	112,43	1,05	BL01498, BL03356, BL53391	No	No	No
Peak022	Sc0000000	5191145	5191445	67,34	0,93	BL01414	No	No	No
Peak023	Sc0000000	6832718	6833018	146,00	0,88	BL15145, BL59544	No	No	No
Peak024	Sc0000000	7944671	7944971	133,95	0,77	BL02989	No	No	No
Peak025	Sc0000001	4215241	4215541	140,20	0,78	BL10960, BL32963, BL65642	No	No	No
Peak026	Sc0000001	5681686	5681986	161,50	0,72	BL18976, BL19350, BL62325, BL70267	No	No	No
Peak027	Sc0000001	6739342	6739642	176,04	0,88	BL24920, BL68952	No	No	No
Peak028	Sc0000001	6862260	6862560	162,98	-0,91	BL01676, BL04240, BL68954, BL95023, BL95024	No	No	No
Peak029	Sc0000001	8536906	8537206	187,66	-0,84	BL07206, BL16915, BL16951, BL19687, BL22986, BL54789	No	No	No
Peak030	Sc0000001	9739088	9739388	342,98	-0,81	BL19967, BL47485, BL86257, BL95037	No	No	No
Peak031	Sc0000002	1374366	1374666	167,55	-0,85	BL44394, BL44401, BL83628, BL95040, BL95041	No	No	No
Peak032	Sc0000002	3052884	3053184	246,78	0,89	BL10628, BL16535	Yes	Yes	No
Peak033	Sc0000002	3477878	3478178	73,63	0,84	BL09843, BL32205, BL65097	Yes	Yes	No
Peak034	Sc0000002	4885699	4885999	126,33	0,74	BL02877	No	No	No
Peak035	Sc0000002	4969385	4969685	278,83	0,75	BL00116, BL08127, BL23025	No	No	No
Peak036	Sc0000002	6160324	6160624	79,94	0,90	BL05339, BL09353	Yes	No	No
Peak037	Sc0000002	6269695	6269995	150,73	0,78	BL02579, BL04395	Yes	No	No
Peak038	Sc0000002	6449180	6449480	68,11	0,81	BL02016, BL23908	No	No	No
Peak039	Sc0000002	6984002	6984302	306,08	0,97	BL09347	No	No	No
Peak040	Sc0000002	70175	70475	165,15	0,80	BL05339, BL09353	Yes	No	Yes
Peak041	Sc0000002	70439	70739	191,60	0,82	BL05339, BL09353	Yes	Yes	No
Peak042	Sc0000002	7807660	7807960	134,48	0,76	BL08125	No	No	No
Peak043	Sc0000003	1855536	1855836	200,42	0,85	BL01444, BL37967, BL37979, BL37980, BL53439, BL55637, BL55652	Yes	Yes	No
Peak044	Sc0000003	330002	330302	180,81	1,63	BL08609	No	No	No
Peak045	Sc0000003	331517	331817	108,27	0,97	BL08609	No	No	No
Peak046	Sc0000003	3570795	3571095	154,97	-0,87	BL05781, BL12683, BL22956, BL25075, BL28902	No	No	No
Peak047	Sc0000003	5597600	5597900	136,24	1,19	BL15160, BL30582, BL48937, BL48938, BL96222	No	No	No
Peak048	Sc0000004	1596071	1596371	281,97	-0,77	BL02798, BL16394, BL16784, BL22890, BL60600	Yes	No	No
Peak049	Sc0000004	3103309	3103609	109,87	0,92	BL04636	No	No	No
Peak050	Sc0000004	510090	510390	88,97	0,94	BL00639, BL07309, BL08030, BL86382	No	No	No
Peak051	Sc0000004	510930	511230	426,02	0,68	BL00639, BL07309, BL08030, BL86382	No	No	No
Peak052	Sc0000004	6030510	6030810	87,06	0,85	BL33806, BL72057, BL72058, BL72060, BL72062, BL72063, BL72064, BL72065, BL72066, BL72067	No	No	No
Peak053	Sc0000004	6471843	6472143	127,38	0,80	BL18548, BL19216, BL37474, BL37475, BL59531, BL59534, BL59535, BL76413	No	No	No
Peak054	Sc0000005	1362315	1362615	216,97	-0,75	BL13361, BL17018, BL40197	No	No	No
Peak055	Sc0000005	1905639	1905939	209,46	-0,75	BL02674, BL06336, BL25711, BL51237, BL51238, BL51247, BL69060, BL69061, BL73837, BL73838, BL90869	No	No	No
Peak056	Sc0000005	2412630	2412930	653,97	-0,74	BL17021	No	No	No
Peak057	Sc0000005	5761792	5762092	162,80	0,76	BL02314	No	No	No
Peak058	Sc0000005	5898893	5899193	66,19	0,94	BL07014, BL27820, BL47053	No	No	No
Peak059	Sc0000005	6563627	6563927	153,22	0,79	BL01215, BL47010	No	No	No

Peak ID	Scaf	start	end	bM	I2FC	Assigned genes	RA N3 vs RA N4 vs ATAC	Motifs NR	Motifs FKH
Peak060	Sc0000005	7135023	7135323	169,71	0,72	BL06334, BL63860, BL63864, BL63865, BL95101, BL95103	No	No	No
Peak061	Sc0000006	1397782	1398082	174,68	0,77	BL04856	No	No	No
Peak062	Sc0000006	3160892	3161192	67,05	0,90	BL10091	No	No	No
Peak063	Sc0000006	4290399	4290699	55,96	-0,83	BL60218, BL60219, BL63515, BL96518, BL96519, BL96520	No	No	No
Peak064	Sc0000006	4365672	4365972	123,54	0,75	BL20605, BL60214, BL82187	No	No	No
Peak065	Sc0000006	5627771	5628071	113,86	0,78	BL15451, BL62276, BL62282	No	No	No
Peak066	Sc0000006	6553274	6553574	148,21	0,80	BL06719, BL31450, BL66098	Yes	No	No
Peak067	Sc0000006	6568082	6568382	565,16	0,86	BL06719	Yes	Yes	No
Peak068	Sc0000007	1459121	1459421	133,70	0,79	BL10052, BL32318	No	No	No
Peak069	Sc0000007	2093461	2093761	336,73	0,67	BL08387, BL21325, BL38494, BL38495, BL80551	No	No	No
Peak070	Sc0000007	2268228	2268528	197,76	-0,73	BL02150, BL24607, BL62707, BL62708, BL71842, BL80547	No	No	No
Peak071	Sc0000007	3512698	3512998	378,47	0,65	BL15756	No	No	No
Peak072	Sc0000007	4001243	4001543	62,12	1,37	BL01076, BL75199	No	No	No
Peak073	Sc0000007	5254298	5254598	123,86	0,78	BL56868	No	No	No
Peak074	Sc0000007	7032588	7032888	186,64	0,73	BL18293, BL59310, BL59312, BL59313, BL95153	No	No	No
Peak075	Sc0000008	1209275	1209575	59,28	0,86	BL17965, BL18850	No	No	No
Peak076	Sc0000008	4689990	4690290	143,27	0,87	BL06754, BL10810, BL67079, BL71557, BL71558	Yes	Yes	No
Peak077	Sc0000008	4691815	4692115	91,85	1,15	BL06754, BL10810, BL67079, BL71557, BL71558	Yes	Yes	Yes
Peak078	Sc0000008	4703724	4704024	132,87	1,05	BL06754, BL67079, BL67080, BL71555, BL71556, BL71557, BL71558	Yes	Yes	Yes
Peak079	Sc0000008	4707927	4708227	120,62	1,19	BL06754, BL45046, BL67079, BL67080, BL71555, BL71556, BL71557	Yes	Yes	Yes
Peak080	Sc0000008	6070921	6071221	102,32	0,81	BL09203, BL09559, BL17324, BL18476, BL62563, BL62566, BL74168	No	No	No
Peak081	Sc0000009	2177472	2177772	314,70	1,13	BL06315, BL18793, BL22082, BL22845, BL28065, BL34695, BL36759, BL36760, BL36770, BL43370, BL43371, BL67825	Yes	Yes	No
Peak082	Sc0000009	2177969	2178269	90,96	1,12	BL06315, BL18793, BL22082, BL22845, BL28065, BL34695, BL36759, BL36760, BL36770, BL43370, BL43371, BL67825	Yes	No	No
Peak083	Sc0000009	2179081	2179381	157,20	1,69	BL06315, BL18793, BL22082, BL22845, BL28065, BL34695, BL36759, BL36760, BL36770, BL43370, BL43371, BL67825	Yes	Yes	No
Peak084	Sc0000009	2180612	2180912	139,22	1,12	BL06315, BL18793, BL22082, BL22845, BL28065, BL34695, BL36759, BL36760, BL36770, BL43368, BL43370, BL43371, BL67825	Yes	Yes	No
Peak085	Sc0000009	5713412	5713712	263,80	0,74	BL07042	No	No	No
Peak086	Sc0000010	2489796	2490096	156,04	0,86	BL06597, BL27748, BL86357	No	No	No
Peak087	Sc0000010	2516098	2516398	402,66	0,70	BL03213, BL25314, BL59413, BL59421	No	No	No
Peak088	Sc0000010	4652314	4652614	92,82	0,97	BL01000, BL05647, BL44823, BL47778, BL47780, BL47781, BL84529, BL84530	No	No	No
Peak089	Sc0000010	5043223	5043523	248,74	0,82	BL02854, BL46417, BL46418, BL46419, BL91145, BL91146, BL91147, BL91148	No	No	No
Peak090	Sc0000011	1410707	1411007	78,59	0,86	BL11212	No	No	No
Peak091	Sc0000011	1412256	1412556	150,30	0,92	BL11212	No	No	No
Peak092	Sc0000011	1434919	1435219	126,12	0,78	BL11212	No	No	No
Peak093	Sc0000011	3742456	3742756	68,54	0,84	BL20489	No	No	No
Peak094	Sc0000011	4896006	4896306	272,80	0,86	BL01027, BL20492, BL52848, BL52849	No	No	No
Peak095	Sc0000011	4911217	4911517	63,64	0,90	BL20492	No	No	No
Peak096	Sc0000011	4951229	4951529	237,45	1,06	BL04730, BL52850	Yes	Yes	No
Peak097	Sc0000012	118269	118569	185,21	0,82	BL13745	No	No	No
Peak098	Sc0000012	2875181	2875481	162,99	1,22	BL18032, BL18946, BL22976, BL26205, BL26463, BL28932, BL61064, BL83063	No	No	No
Peak099	Sc0000012	2880477	2880777	151,90	0,81	BL18946, BL61064, BL71171, BL71172, BL83063	No	No	No
Peak100	Sc0000012	3007142	3007442	99,52	-0,83	BL18948, BL25690, BL33831	No	No	No
Peak101	Sc0000013	229768	230068	168,91	-0,76	BL11819, BL47804, BL96825	No	No	No
Peak102	Sc0000013	2534257	2534557	181,18	1,47	BL37627, BL50120, BL59708, BL59709, BL22946	Yes	Yes	Yes
Peak103	Sc0000013	2668663	2668963	157,57	-0,81	BL08901, BL36736, BL53088	No	No	No
Peak104	Sc0000013	4106612	4106912	137,32	1,06	BL08896	Yes	No	No
Peak105	Sc0000013	4110045	4110345	476,14	0,68	BL08896	Yes	Yes	Yes
Peak106	Sc0000013	4110356	4110656	761,16	0,60	BL08896	Yes	No	No
Peak107	Sc0000013	4114915	4115215	255,61	1,08	BL08896, BL19277, BL78634	Yes	Yes	No
Peak108	Sc0000013	4116929	4117229	129,97	1,38	BL08896, BL19277, BL78634	Yes	No	No
Peak109	Sc0000013	4117272	4117572	743,30	1,12	BL08896, BL19277, BL78634	Yes	Yes	No
Peak110	Sc0000013	4117749	4118049	1796,13	1,22	BL08896, BL19277, BL78634	Yes	Yes	No
Peak111	Sc0000013	4118324	4118624	359,04	1,13	BL08896, BL19277, BL78634	Yes	No	No
Peak112	Sc0000013	4143292	4143592	131,71	0,95	BL78634	Yes	No	No
Peak113	Sc0000013	4155169	4155469	120,91	0,82	BL78634	Yes	Yes	No
Peak114	Sc0000014	1880946	1881246	248,83	0,72	BL13385	No	No	No
Peak115	Sc0000015	2176160	2176460	54,99	-0,98	BL00321, BL88994	No	No	No
Peak116	Sc0000015	3930376	3930676	163,57	0,91	BL02605, BL04145, BL69283	No	No	No
Peak117	Sc0000015	721562	721862	107,29	1,14	BL09100, BL18378	No	No	No

Peak ID	Scaf	start	end	bM	l2FC	Assigned genes	RA N3 vs RA N4 vs ATAC	Motifs NR	Motifs FKH
Peak118	Sc0000017	3104726	3105026	121,73	-0,76	BL14679, BL95289, BL95292	No	No	No
Peak119	Sc0000017	812067	812367	100,20	0,82	BL01435, BL05124	No	No	No
Peak120	Sc0000018	2203095	2203395	135,54	-0,83	BL17178, BL24005, BL46044, BL46046, BL46048, BL83379, BL83380, BL97402	No	No	No
Peak121	Sc0000018	3349264	3349564	596,90	0,70	BL08693	No	No	No
Peak122	Sc0000018	3416089	3416389	62,38	0,86	BL20572, BL40994, BL57956	No	No	No
Peak123	Sc0000019	3141360	3141660	199,79	0,71	BL61003, BL86008, BL86011, BL88471	No	No	No
Peak124	Sc0000019	972275	972575	83,06	0,84	BL20426, BL27042, BL53520, BL53522	No	No	No
Peak125	Sc0000021	1599689	1599989	109,51	0,82	BL14795, BL27443	No	No	No
Peak126	Sc0000022	1988117	1988417	179,63	1,02	BL09013	No	No	No
Peak127	Sc0000022	2195507	2195807	148,77	0,77	BL11861, BL20221, BL21074, BL58970, BL78877, BL78878	No	No	No
Peak128	Sc0000022	2319762	2320062	72,99	1,43	BL08133, BL32303, BL78871	No	No	No
Peak129	Sc0000023	1515001	1515301	79,41	0,85	BL02236, BL08714, BL70111, BL74583	No	No	No
Peak130	Sc0000024	214202	214502	395,76	-0,70	BL05551, BL24757, BL83737, BL84555	No	No	No
Peak131	Sc0000025	1016561	1016861	342,18	-0,69	BL22199, BL23428, BL31304, BL31584, BL50231	Yes	No	No
Peak132	Sc0000027	1069301	1069601	499,40	0,86	BL15112	Yes	Yes	No
Peak133	Sc0000027	1076942	1077242	291,70	1,04	BL15112	Yes	Yes	No
Peak134	Sc0000027	1077850	1078150	183,48	1,02	BL15112	Yes	No	No
Peak135	Sc0000027	1081854	1082154	108,88	0,86	BL15112	Yes	No	No
Peak136	Sc0000027	1088925	1089225	889,12	0,83	BL15112	Yes	Yes	No
Peak137	Sc0000027	1089325	1089625	204,41	1,29	BL15112	Yes	No	Yes
Peak138	Sc0000027	1093227	1093527	246,12	1,67	BL15112, BL85547	Yes	No	Yes
Peak139	Sc0000027	1093655	1093955	249,28	1,30	BL15112, BL85547	Yes	Yes	Yes
Peak140	Sc0000027	1095822	1096122	339,18	1,13	BL15112, BL85547	Yes	No	No
Peak141	Sc0000027	1098939	1099239	281,32	1,07	BL14303, BL15112, BL85547	Yes	Yes	No
Peak142	Sc0000027	1686041	1686341	73,01	-0,81	BL09234, BL47279, BL70343, BL81719, BL81720, BL96717, BL96718, BL96719	No	No	No
Peak143	Sc0000027	2609745	2610045	97,11	-1,07	BL02746, BL14308, BL14931, BL20673	No	No	No
Peak144	Sc0000028	2171568	2171868	63,11	0,95	BL19050	No	No	No
Peak145	Sc0000028	2172188	2172488	116,43	1,02	BL19050	No	No	No
Peak146	Sc0000028	227355	227655	369,36	0,64	BL00501, BL07878, BL88341, BL88345	Yes	Yes	Yes
Peak147	Sc0000028	232586	232886	232,92	0,82	BL00501, BL07878, BL31721, BL88341, BL88345	Yes	Yes	No
Peak148	Sc0000029	669528	669828	101,36	1,10	BL15076, BL18235	No	No	No
Peak149	Sc0000030	1847132	1847432	138,30	-0,84	BL03237, BL15378, BL23013, BL38687, BL60820, BL87030, BL87031	No	No	No
Peak150	Sc0000030	1847443	1847743	150,18	-0,90	BL03237, BL15378, BL23013, BL38687, BL60820, BL87030, BL87031	No	No	No
Peak151	Sc0000030	1881116	1881416	65,72	-0,83	BL02351, BL06046, BL18118, BL20199	Yes	Yes	No
Peak152	Sc0000030	195357	195657	115,43	1,23	BL06889	No	No	No
Peak153	Sc0000030	195704	196004	69,63	0,95	BL06889	No	No	No
Peak154	Sc0000031	1557429	1557729	515,41	0,63	BL20788, BL49727, BL75120, BL75121, BL75130, BL86983, BL86984, BL88262	No	No	No
Peak155	Sc0000031	1894128	1894428	258,07	0,79	BL11542, BL91401	No	No	No
Peak156	Sc0000032	804274	804574	271,37	0,90	BL23329	No	No	No
Peak157	Sc0000033	1535381	1535681	94,37	1,09	BL20247	No	No	No
Peak158	Sc0000034	1624740	1625040	75,89	1,03	BL10401	Yes	Yes	Yes
Peak159	Sc0000034	1625145	1625445	74,65	1,19	BL10401	Yes	No	Yes
Peak160	Sc0000034	1648412	1648712	128,93	1,26	BL10401, BL63620	Yes	No	No
Peak161	Sc0000034	1649367	1649667	142,88	0,92	BL10401, BL63620	Yes	No	No
Peak162	Sc0000034	1649655	1649955	104,02	1,58	BL10401, BL63620	Yes	Yes	No
Peak163	Sc0000034	1649866	1650166	135,04	1,77	BL10401, BL63620	Yes	Yes	No
Peak164	Sc0000034	1650275	1650575	229,70	2,44	BL10401, BL63620	Yes	Yes	Yes
Peak165	Sc0000034	1664595	1664895	342,76	2,11	BL10401	Yes	Yes	No
Peak166	Sc0000035	127323	127623	71,68	0,94	BL00095, BL18624, BL29261	Yes	No	No
Peak167	Sc0000035	139690	139990	105,02	1,25	BL00095, BL18624, BL29261	Yes	Yes	No
Peak168	Sc0000036	1950233	1950533	63,06	1,04	BL02842	No	No	No
Peak169	Sc0000036	240715	241015	78,79	1,09	BL13690, BL25663, BL28349, BL51961, BL79859	No	No	No
Peak170	Sc0000037	2268990	2269290	146,51	0,75	BL11415, BL31098, BL35038	No	No	No
Peak171	Sc0000039	638099	638399	106,33	0,89	BL75213, BL97070, BL97071,	Yes	No	Yes
Peak172	Sc0000040	1281096	1281396	216,07	0,89	BL00808, BL25011, BL25014, BL63446	Yes	Yes	No
Peak173	Sc0000040	1281707	1282007	122,20	0,81	BL00808, BL25011, BL25014, BL63446	Yes	No	Yes
Peak174	Sc0000040	566357	566657	159,87	0,72	BL17347, BL60579, BL60581, BL67720, BL95436, BL95437	Yes	No	No
Peak175	Sc0000041	48262	48562	105,71	0,88	BL23135, BL26626, BL95443	Yes	Yes	Yes
Peak176	Sc0000042	119530	119830	300,39	-0,65	BL16702	No	No	No
Peak177	Sc0000042	1413377	1413677	430,55	0,60	BL06533, BL29452, BL90692, BL95454	No	Yes	No
Peak178	Sc0000042	280763	281063	94,47	0,89	BL23525, BL54069, BL55403	No	No	No
Peak179	Sc0000043	1034696	1034996	249,74	0,65	BL04101, BL10607, BL28556	No	No	No
Peak180	Sc0000043	2107931	2108231	90,72	-0,94	BL01357, BL04094, BL06329, BL19871, BL22714, BL34704	No	No	No
Peak181	Sc0000043	784749	785049	60,31	1,00	BL04098, BL45174	No	No	No
Peak182	Sc0000043	785665	785965	775,24	0,56	BL04098, BL45174	No	No	No
Peak183	Sc0000044	374963	375263	178,11	0,94	BL29533, BL29535, BL61353, BL86328, BL95460, BL95461	Yes	Yes	Yes

Peak ID	Scaf	start	end	bM	l2FC	Assigned genes	RA N3 vs RA N4 vs ATAC	Motifs NR	Motifs FKH
Peak184	Sc0000044	376315	376615	147,41	0,83	BL29533, BL29535, BL61353, BL86328, BL95460, BL95461	Yes	No	Yes
Peak185	Sc0000044	377456	377756	272,97	1,51	BL29533, BL29535, BL61353, BL86328, BL95460, BL95461	Yes	Yes	No
Peak186	Sc0000044	378842	379142	742,33	0,61	BL29533, BL29535, BL61353, BL86328, BL95460, BL95461	Yes	No	No
Peak187	Sc0000044	380791	381091	303,36	1,18	BL29533, BL29535, BL95460, BL95461	Yes	Yes	Yes
Peak188	Sc0000044	399144	399444	323,03	0,88	BL61354	No	No	No
Peak189	Sc0000044	399839	400139	225,62	1,18	BL61354	No	No	No
Peak190	Sc0000044	402232	402532	267,96	0,68	BL10465, BL61354	No	No	No
Peak191	Sc0000044	407536	407836	87,18	0,87	BL10465, BL61354	No	No	No
Peak192	Sc0000044	55462	55762	309,50	0,71	BL08554	No	No	No
Peak193	Sc0000048	119323	119623	134,62	-0,88	BL07172, BL14990	No	No	No
Peak194	Sc0000048	703949	704249	71,93	0,83	BL13739	No	No	No
Peak195	Sc0000049	1075866	1076166	127,93	1,08	BL65803, BL75757	Yes	Yes	No
Peak196	Sc0000049	649516	649816	71,46	-0,85	BL77893, BL77894, BL95502	No	No	No
Peak197	Sc0000049	782069	782369	93,25	0,86	BL40355, BL77899, BL95503	Yes	Yes	No
Peak198	Sc0000049	874495	874795	69,63	0,95	BL12697	No	No	No
Peak199	Sc0000049	874768	875068	67,77	1,06	BL12697	No	No	No
Peak200	Sc0000050	265441	265741	268,61	-0,75	BL02286, BL61344	No	No	No
Peak201	Sc0000050	747022	747322	95,08	0,80	BL21170	No	No	No
Peak202	Sc0000051	1196624	1196924	73,50	0,87	BL03600, BL20032, BL50587, BL50589	No	No	No
Peak203	Sc0000051	1568348	1568648	164,41	0,89	BL06930, BL29359, BL38020	No	No	No
Peak204	Sc0000052	465466	465766	370,40	0,64	BL04695, BL04701, BL25394	No	No	No
Peak205	Sc0000053	1306034	1306334	278,97	0,79	BL14000, BL16726, BL43347	No	No	No
Peak206	Sc0000053	1332785	1333085	453,80	0,59	BL16726, BL43348	No	No	No
Peak207	Sc0000053	1344905	1345205	85,16	1,07	BL08243, BL43348	No	No	No
Peak208	Sc0000053	1345281	1345581	266,52	1,73	BL08243, BL43348	No	No	No
Peak209	Sc0000054	549534	549834	332,99	-0,89	BL00734, BL03542, BL51882, BL51883, BL97214, BL97216,	No	No	No
Peak210	Sc0000054	685975	686275	86,78	0,84	BL19332	No	No	No
Peak211	Sc0000055	1169792	1170092	89,00	0,90	BL13483, BL69797, BL73147, BL73148	No	No	No
Peak212	Sc0000055	1266955	1267255	71,46	0,81	BL02248, BL12433, BL35693	Yes	Yes	No
Peak213	Sc0000055	1269000	1269300	122,71	1,01	BL02248, BL12433, BL35693	Yes	No	No
Peak214	Sc0000055	1269267	1269567	72,20	1,05	BL02248, BL12433, BL35693	Yes	Yes	No
Peak215	Sc0000055	1270394	1270694	86,30	1,27	BL02248, BL12433, BL35693	Yes	No	No
Peak216	Sc0000056	406526	406826	107,35	1,06	BL02504, BL22432, BL48160, BL48161, BL80610	Yes	Yes	No
Peak217	Sc0000057	1104207	1104507	166,20	0,79	BL18466, BL73971, BL73972, BL76798	No	No	No
Peak218	Sc0000058	638083	638383	146,33	-0,75	BL16756, BL82094, BL91513	No	No	No
Peak219	Sc0000058	805584	805884	251,84	0,73	BL05204, BL54208, BL54211, BL86601, BL86603	Yes	Yes	Yes
Peak220	Sc0000061	643692	643992	113,51	0,94	BL10379, BL16901, BL20516	No	No	No
Peak221	Sc0000062	1511286	1511586	210,20	1,05	BL02209	No	No	No
Peak222	Sc0000062	1576487	1576787	226,56	0,80	BL02209, BL27511, BL63796	No	No	No
Peak223	Sc0000062	41350	41650	187,64	0,81	BL02215	No	No	No
Peak224	Sc0000062	482943	483243	148,47	-0,78	BL08694, BL08696, BL24889, BL28059, BL31534, BL33159, BL33160, BL33161, BL33162	No	No	No
Peak225	Sc0000063	1014262	1014562	175,19	-0,82	BL06362	No	No	No
Peak226	Sc0000063	1203706	1204006	144,91	0,78	BL11951, BL25226	No	No	No
Peak227	Sc0000063	1215521	1215821	199,10	0,89	BL11951, BL25226	No	No	No
Peak228	Sc0000063	122462	122762	96,56	0,80	BL02648, BL11953	No	No	No
Peak229	Sc0000063	277191	277491	193,70	-0,80	BL20831, BL23357, BL24894, BL32051, BL47543	No	No	No
Peak230	Sc0000063	526899	527199	171,85	0,80	BL06369	No	No	No
Peak231	Sc0000067	39268	39568	54,39	0,89	BL12377, BL16364, BL70433, BL70434, BL70435	No	No	No
Peak232	Sc0000067	434277	434577	73,87	-0,92	BL95593	No	No	No
Peak233	Sc0000067	505538	505838	103,38	0,93	BL07788, BL16365, BL88087, BL88088, BL88089	Yes	Yes	Yes
Peak234	Sc0000071	315279	315579	160,96	0,95	BL14317, BL18893	No	No	No
Peak235	Sc0000072	1008729	1009029	95,92	-0,95	BL07349, BL12618	No	No	No
Peak236	Sc0000072	294008	294308	121,12	1,03	BL25553, BL40934, BL40935, BL40942, BL40943	No	No	No
Peak237	Sc0000074	1050045	1050345	59,41	-0,89	BL04281, BL04287, BL39299, BL40406, BL40407, BL48452, BL48467	No	No	No
Peak238	Sc0000074	560712	561012	631,98	0,84	BL05044, BL20098, BL20564, BL52944, BL52948, BL52949, BL88730	Yes	Yes	No
Peak239	Sc0000074	563084	563384	87,55	0,91	BL05044, BL20098, BL20564, BL52944, BL52948, BL52949, BL88730	Yes	Yes	Yes
Peak240	Sc0000075	1275411	1275711	182,39	1,09	BL00176	No	No	No
Peak241	Sc0000075	591631	591931	239,05	-0,72	BL15106, BL15110, BL37119, BL37120, BL37121, BL37350, BL37351, BL71940, BL82583	No	No	No
Peak242	Sc0000076	1001501	1001801	85,62	0,80	BL22542	No	No	No
Peak243	Sc0000076	360804	361104	74,82	0,95	BL00365, BL16095, BL46073, BL83393	No	No	No
Peak244	Sc0000077	1271197	1271497	260,60	0,78	BL58872, BL84157, BL85189, BL95632, BL95634	Yes	Yes	No
Peak245	Sc0000079	316777	317077	107,70	0,76	BL22552, BL89482	Yes	Yes	No

Peak ID	Scaf	start	end	bM	l2FC	Assigned genes	RA N3 vs RA N4 vs ATAC	Motifs NR	Motifs FKH
Peak246	Sc0000080	648652	648952	286,57	0,72	BL13701, BL42573	No	No	No
Peak247	Sc0000082	232995	233295	92,11	1,20	BL54610, BL54619, BL69546, BL71967	No	No	No
Peak248	Sc0000082	260087	260387	399,56	0,71	BL06291, BL68703	No	No	No
Peak249	Sc0000083	1064442	1064742	257,75	0,64	BL01122, BL07554, BL25921, BL33893, BL69073	No	No	No
Peak250	Sc0000085	1036001	1036301	65,76	-0,85	BL95645	No	No	No
Peak251	Sc0000088	17624	17924	88,01	0,79	BL04008, BL04009, BL04012, BL42489, BL63676	Yes	No	Yes
Peak252	Sc0000088	17908	18208	165,64	0,82	BL04008, BL04009, BL04012, BL42489, BL63676	Yes	No	No
Peak253	Sc0000088	18331	18631	92,75	0,96	BL04008, BL04009, BL04012, BL63676	Yes	No	No
Peak254	Sc0000088	420024	420324	310,22	0,63	BL10537, BL19811, BL57066, BL57068	No	No	No
Peak255	Sc0000089	226408	226708	277,30	-0,70	BL54864, BL90169	No	No	No
Peak256	Sc0000092	324631	324931	118,38	1,03	BL21577	No	No	No
Peak257	Sc0000092	93446	93746	141,28	0,74	BL08620, BL17919, BL31423	Yes	No	No
Peak258	Sc0000093	441117	441417	170,32	1,00	BL10581	Yes	No	Yes
Peak259	Sc0000094	686768	687068	123,23	0,75	BL11936, BL11937, BL59305	No	No	No
Peak260	Sc0000095	838068	838368	208,79	0,70	BL00004	No	No	No
Peak261	Sc0000098	139235	139535	144,48	1,15	BL17936	No	No	No
Peak262	Sc0000098	140421	140721	107,74	1,52	BL17936	No	No	No
Peak263	Sc0000098	141701	142001	289,46	0,88	BL17936	No	No	No
Peak264	Sc0000098	250229	250529	64,93	-1,07	BL07045, BL17931, BL17934	No	No	No
Peak265	Sc0000098	463111	463411	71,55	1,30	BL09521	No	No	No
Peak266	Sc0000103	678843	679143	303,09	1,25	BL04686	No	No	No
Peak267	Sc0000103	680238	680538	247,84	0,68	BL04686, BL77531	No	No	No
Peak268	Sc0000104	289488	289788	95,94	1,04	BL18079	No	No	No
Peak269	Sc0000104	389626	389926	127,25	0,78	BL13241, BL95731, BL95733	No	No	No
Peak270	Sc0000104	390349	390649	131,35	0,88	BL13241, BL95731, BL95733	No	No	No
Peak271	Sc0000104	504656	504956	368,82	0,59	BL13238, BL19627	Yes	Yes	Yes
Peak272	Sc0000104	506973	507273	248,09	1,02	BL13238, BL19627	Yes	No	No
Peak273	Sc0000105	313238	313538	256,23	0,70	BL24312, BL71394, BL71396, BL71397	No	No	No
Peak274	Sc0000109	428315	428615	83,73	0,80	BL15405	No	No	No
Peak275	Sc0000113	148157	148457	153,00	1,06	BL11253, BL22458, BL23538, BL31237, BL31239, BL39602, BL72213, BL72214, BL72215	Yes	Yes	No
Peak276	Sc0000115	389456	389756	294,50	-0,72	BL21242	No	No	No
Peak277	Sc0000116	215863	216163	56,63	-0,88	BL24277, BL27320, BL51640, BL51649, BL66787, BL72684	No	No	No
Peak278	Sc0000116	308567	308867	98,53	1,02	BL17818	No	No	No
Peak279	Sc0000116	638024	638324	54,37	1,12	BL04883, BL08519, BL75543	Yes	Yes	Yes
Peak280	Sc0000119	275807	276107	267,81	1,38	BL15803, BL58136, BL80917	No	No	No
Peak281	Sc0000120	57144	57444	198,79	-0,83	BL20217, BL91071, BL91072	No	No	No
Peak282	Sc0000120	857077	857377	250,17	-0,66	BL13792, BL35531, BL56584, BL58019, BL58020, BL58026, BL59152	No	No	No
Peak283	Sc0000123	238706	239006	377,97	-0,77	BL21710, BL23193	No	No	No
Peak284	Sc0000124	673149	673449	232,23	-0,78	BL01643, BL19248, BL59510, BL67693	Yes	Yes	No
Peak285	Sc0000126	214898	215198	64,82	-0,93	BL59504	No	No	No
Peak286	Sc0000129	853522	853822	158,27	1,83	BL17461, BL38128, BL61361, BL67487, BL67491, BL67494, BL67500, BL67504	No	No	No
Peak287	Sc0000130	393520	393820	174,13	0,76	BL04159	No	No	No
Peak288	Sc0000130	394001	394301	180,41	1,33	BL04159	No	No	No
Peak289	Sc0000134	37104	37404	224,34	0,77	BL21590	No	No	No
Peak290	Sc0000139	274704	275004	258,56	1,04	BL13136, BL41702, BL41703	No	No	No
Peak291	Sc0000139	276654	276954	125,06	0,99	BL13136, BL41703	No	No	No
Peak292	Sc0000139	277139	277439	138,78	1,21	BL13136, BL41703	No	No	No
Peak293	Sc0000139	277726	278026	66,45	0,89	BL13136, BL41703	No	No	No
Peak294	Sc0000139	278055	278355	140,60	1,17	BL01662, BL13136, BL41703	No	No	No
Peak295	Sc0000143	459253	459553	137,61	0,88	BL14704, BL18271, BL43346, BL80803, BL85195, BL85198	No	No	No
Peak296	Sc0000144	656017	656317	249,57	-0,67	BL03558	No	No	No
Peak297	Sc0000146	183538	183838	173,65	-0,75	BL14942, BL16979, BL24136, BL24145, BL72887, BL82168, BL82170	No	No	No
Peak298	Sc0000146	699384	699684	267,09	1,35	BL17675	No	No	No
Peak299	Sc0000152	359171	359471	102,74	0,89	BL12779, BL51871, BL51879, BL82144	No	No	No
Peak300	Sc0000152	378611	378911	207,36	0,95	BL17845	Yes	Yes	No
Peak301	Sc0000156	275757	276057	316,93	-0,75	BL00486, BL29170, BL76284	No	No	No
Peak302	Sc0000157	532335	532635	290,08	0,71	BL12452, BL35954	No	No	No
Peak303	Sc0000158	584667	584967	135,99	1,01	BL23731, BL70166	No	No	No
Peak304	Sc0000159	449131	449431	104,29	0,78	BL05597, BL24161	No	No	No
Peak305	Sc0000159	68289	68589	148,95	0,78	BL13049, BL46888, BL59908	No	No	No
Peak306	Sc0000159	71851	72151	115,90	0,85	BL13049, BL46888, BL59908	No	No	No
Peak307	Sc0000161	641989	642289	392,12	0,62	BL19994, BL49838, BL83641	No	No	No
Peak308	Sc0000161	689316	689616	284,94	0,71	BL43592, BL43597, BL49831, BL83634, BL83635, BL83637, BL83639, BL87088, BL87089	No	No	No
Peak309	Sc0000165	325289	325589	252,56	1,04	BL18599	No	No	No
Peak310	Sc0000168	276899	277199	143,74	-0,91	BL08276, BL13988	No	No	No
Peak311	Sc0000168	414969	415269	79,82	-0,87	BL13989, BL23194, BL88406, BL95856, BL95857	Yes	No	No

Peak ID	Scaf	start	end	bM	l2FC	Assigned genes	RA N3 vs RA N4 vs ATAC	Motifs NR	Motifs FKH
Peak312	Sc0000170	549137	549437	124,12	1,08	BL53141, BL58241, BL58242, BL58243, BL58247, BL58248, BL90450, BL97065	No	No	No
Peak313	Sc0000172	210478	210778	92,04	0,83	BL10405, BL67759, BL67760, BL89665, BL96495	No	No	No
Peak314	Sc0000173	70507	70807	167,11	0,86	BL07417	No	No	No
Peak315	Sc0000176	118014	118314	136,94	-0,82	BL06822, BL10450, BL66167, BL90262, BL90263	Yes	No	Yes
Peak316	Sc0000180	108078	108378	83,11	0,83	BL17450, BL21046, BL35300	Yes	Yes	No
Peak317	Sc0000182	517183	517483	139,02	0,92	BL08826, BL17162, BL72242	No	Yes	No
Peak318	Sc0000183	417523	417823	380,44	0,71	BL00236	No	No	No
Peak319	Sc0000185	440617	440917	148,74	0,93	BL06405, BL12955, BL67648, BL67649, BL85066	No	No	No
Peak320	Sc0000187	110950	111250	130,43	-0,82	BL03259, BL21808	No	No	No
Peak321	Sc0000187	292747	293047	238,87	-0,75	BL06099, BL12526	No	No	No
Peak322	Sc0000188	562421	562721	158,84	0,76	BL02000, BL16119, BL37332, BL37333	No	No	No
Peak323	Sc0000196	216766	217066	136,43	-0,73	BL05363, BL23919, BL27918, BL67475, BL85678, BL85696, BL86422	Yes	Yes	No
Peak324	Sc0000197	296438	296738	212,96	0,67	BL19499	No	No	No
Peak325	Sc0000200	10126	10426	291,37	1,11	BL19441	No	No	No
Peak326	Sc0000210	41001	41301	106,18	1,01	BL24550	No	No	No
Peak327	Sc0000210	41362	41662	158,61	1,12	BL24550	No	No	No
Peak328	Sc0000212	102993	103293	213,03	0,67	BL06716	No	No	No
Peak329	Sc0000212	107862	108162	113,12	0,79	BL06716	No	No	No
Peak330	Sc0000212	89620	89920	114,10	0,80	BL06716	No	No	No
Peak331	Sc0000215	67973	68273	85,65	1,66	BL06879, BL06885, BL36767, BL36768, BL65403, BL65404, BL65408	Yes	No	Yes
Peak332	Sc0000215	68364	68664	185,40	2,07	BL06879, BL06885, BL36767, BL36768, BL65403, BL65404, BL65408,	Yes	Yes	No
Peak333	Sc0000215	68855	69155	144,69	1,24	BL06879, BL06885, BL36767, BL36768, BL65403, BL65404, BL65408	Yes	Yes	Yes
Peak334	Sc0000220	171909	172209	315,59	0,65	BL38779, BL38780, BL76316, BL84979, BL84986, BL84987	No	No	No
Peak335	Sc0000220	172331	172631	447,94	0,57	BL38779, BL38780, BL76316, BL84979, BL84986, BL84987	No	No	No
Peak336	Sc0000223	201037	201337	380,27	0,67	BL00750, BL07229, BL55214, BL55220, BL60429	No	No	No
Peak337	Sc0000223	201893	202193	138,77	1,06	BL00750, BL07229, BL55214, BL55220, BL60429	No	No	No
Peak338	Sc0000224	335751	336051	426,54	0,63	BL10268	No	No	No
Peak339	Sc0000224	341722	342022	81,88	1,22	BL10268	No	No	No
Peak340	Sc0000224	344272	344572	181,25	1,01	BL10268	No	No	No
Peak341	Sc0000229	406730	407030	80,76	1,40	BL14695	No	No	No
Peak342	Sc0000231	354223	354523	152,34	1,74	BL10412	No	No	No
Peak343	Sc0000234	126579	126879	72,19	0,82	BL41346	No	No	No
Peak344	Sc0000234	139370	139670	147,61	0,88		No	No	No
Peak345	Sc0000235	191170	191470	184,45	1,17	BL07657, BL91187, BL95945	No	No	No
Peak346	Sc0000241	243659	243959	76,64	0,80	BL11102, BL30720, BL48659	No	No	No
Peak347	Sc0000253	355430	355730	183,98	0,72	BL16801, BL82306	No	No	No
Peak348	Sc0000255	257277	257577	455,18	0,63	BL06383, BL22145, BL70648, BL70649, BL90597	Yes	Yes	No
Peak349	Sc0000260	69043	69343	60,92	-0,86	BL04676, BL14067	No	No	No
Peak350	Sc0000262	119711	120011	196,16	-0,83	BL11236, BL49257	No	No	No
Peak351	Sc0000266	352614	352914	214,87	0,81	BL24259, BL62722, BL83082, BL83083	Yes	No	No
Peak352	Sc0000270	352010	352310	132,41	1,43	BL04899, BL27574	No	No	No
Peak353	Sc0000271	106161	106461	146,87	0,71	BL02294, BL45138, BL69340	No	No	No
Peak354	Sc0000274	212743	213043	77,52	1,12	BL03672	No	No	No
Peak355	Sc0000285	296035	296335	115,89	0,82	BL03388	No	No	No
Peak356	Sc0000285	296760	297060	160,04	1,16	BL03388	No	No	No
Peak357	Sc0000287	291932	292232	108,33	0,77	BL24379, BL24381, BL59697	No	No	No
Peak358	Sc0000288	7591	7891	238,14	-0,73	BL19072, BL21881, BL69394	No	No	No
Peak359	Sc0000301	168331	168631	341,21	-0,69	BL10392, BL11640, BL14482	No	No	No
Peak360	Sc0000302	254775	255075	144,84	-0,80	BL03374, BL03376, BL12637, BL12642, BL25186, BL38498, BL69245	No	No	No
Peak361	Sc0000302	255028	255328	207,32	-0,67	BL03374, BL03376, BL12637, BL12642, BL25186, BL38498, BL69245	No	No	No
Peak362	Sc0000302	30169	30469	77,75	1,07	BL03370, BL50912, BL50913, BL84443, BL84445, BL84446, BL84447, BL84448	No	No	No
Peak363	Sc0000302	48979	49279	86,69	1,00	BL03370, BL50918, BL84445	No	No	No
Peak364	Sc0000302	54825	55125	96,12	0,79	BL03370, BL50918, BL84445, BL84449	No	No	No
Peak365	Sc0000302	8038	8338	74,24	1,10	BL03368, BL32292, BL42066, BL42067, BL50909, BL50910, BL84442, BL91075, BL91078	No	No	No
Peak366	Sc0000305	112596	112896	90,76	0,77	BL06407, BL25250, BL26430, BL29496, BL37725, BL39819, BL39820, BL39821, BL48696, BL62448, BL86802	No	No	No
Peak367	Sc0000324	131636	131936	238,31	0,93	BL15783, BL22206, BL51702, BL62024, BL96030	Yes	Yes	No
Peak368	Sc0000327	8678	8978	269,01	0,72	BL33373, BL79532	No	No	No
Peak369	Sc0000337	38641	38941	68,53	0,80	BL02840, BL06465	Yes	Yes	Yes
Peak370	Sc0000343	168239	168539	91,92	1,24	BL16767, BL16771, BL30016	Yes	No	No
Peak371	Sc0000343	168721	169021	70,21	0,96	BL16767, BL16771, BL30016	Yes	No	No

Peak ID	Scaf	start	end	bM	I2FC	Assigned genes	RA N3 vs RA N4 vs ATAC	Motifs NR	Motifs FKH
Peak372	Sc0000343	170391	170691	155,29	0,88	BL16767, BL16771, BL30016	Yes	Yes	No
Peak373	Sc0000348	191288	191588	479,69	-0,61	BL09088, BL54580	No	No	No
Peak374	Sc0000358	208260	208560	132,89	-0,82	BL01055, BL08192, BL16840, BL16845, BL84521	No	No	No
Peak375	Sc0000373	224398	224698	218,07	0,67	BL12144	No	No	No
Peak376	Sc0000377	128007	128307	65,01	1,07	BL18636	No	No	No
Peak377	Sc0000377	128282	128582	60,36	0,86	BL18636	No	No	No
Peak378	Sc0000377	128506	128806	66,68	0,95	BL18636	No	No	No
Peak379	Sc0000382	11050	11350	68,60	1,27	BL00390, BL21664, BL59611	Yes	No	No
Peak380	Sc0000382	5940	6240	74,39	0,91	BL00390, BL21664, BL59611	Yes	No	Yes
Peak381	Sc0000382	7425	7725	127,59	0,95	BL00390, BL21664, BL59611	Yes	No	No
Peak382	Sc0000382	7951	8251	161,40	1,35	BL00390, BL21664, BL59611	Yes	Yes	No
Peak383	Sc0000382	9024	9324	80,24	1,12	BL00390, BL21664, BL59611	Yes	Yes	No
Peak384	Sc0000388	99652	99952	55,69	0,93	BL13423	No	No	No
Peak385	Sc0000399	172119	172419	233,90	0,66	BL14293, BL23461, BL78983	No	No	No
Peak386	Sc0000401	72999	73299	55,83	0,96	BL12595	No	No	No
Peak387	Sc0000414	4600	4900	151,55	-0,80	BL01850, BL10774	No	No	No
Peak388	Sc0000417	170689	170989	111,97	-0,77	BL14789, BL80369, BL80370	No	No	No
Peak389	Sc0000427	129772	130072	110,53	0,99	BL13072, BL13078, BL13081, BL84338, BL90931	Yes	Yes	No
Peak390	Sc0000432	117747	118047	156,07	-0,72	BL10205	No	No	No
Peak391	Sc0000433	52410	52710	71,70	0,81	BL00169, BL06814	No	No	No
Peak392	Sc0000450	45489	45789	390,03	0,70	BL02490, BL37452, BL49645, BL62854, BL77022	No	No	No
Peak393	Sc0000450	46587	46887	177,74	0,84	BL02490, BL49645, BL49701, BL49704, BL62854, BL77022	No	No	No
Peak394	Sc0000455	22508	22808	120,88	-0,79	BL02836, BL02837, BL28476	No	No	No
Peak395	Sc0000476	115545	115845	191,61	0,89	BL21357	No	No	No
Peak396	Sc0000480	126293	126593	151,52	0,73	BL02205, BL02208	No	No	No
Peak397	Sc0000480	134930	135230	145,53	0,71	BL02205, BL02208, BL46999	No	No	No
Peak398	Sc0000488	116809	117109	67,09	0,84	BL10847, BL19404	No	No	No
Peak399	Sc0000496	12904	13204	89,01	0,90	BL03018	No	No	No
Peak400	Sc0000517	79876	80176	189,72	0,76	BL22295, BL22303, BL22304	Yes	Yes	Yes
Peak401	Sc0000526	95904	96204	125,39	-0,96	BL07651, BL22520, BL33253, BL46508, BL52315, BL56413, BL56421, BL59772, BL60217, BL73534	No	No	No
Peak402	Sc0000535	70972	71272	103,57	0,82	BL00273, BL39242, BL43384, BL43394, BL66188, BL80860	Yes	Yes	No
Peak403	Sc0000547	72284	72584	60,76	0,82	BL23314, BL23317, BL26639, BL38283, BL38284, BL96124	Yes	No	No
Peak404	Sc0000569	23571	23871	211,03	0,80	BL22259, BL22261, BL62536	No	No	No
Peak405	Sc0000610	51289	51589	102,89	-0,84	BL07613, BL10427, BL10428, BL10429, BL25360	No	No	No
Peak406	Sc0000630	42536	42836	159,25	0,85	BL00332, BL00336, BL79788	No	No	No
Peak407	Sc0000640	15624	15924	317,34	0,70	BL15774, BL62077, BL62079, BL83565, BL83578, BL83584	Yes	Yes	No
Peak408	Sc0000694	13311	13611	134,29	0,76	BL19609, BL26095, BL39430, BL39519, BL54442, BL67749, BL67978, BL67979	Yes	No	Yes
Peak409	Sc0000694	18544	18844	119,74	1,17	BL19609, BL26095, BL39430, BL39519, BL67978, BL67979, BL70565	Yes	Yes	Yes
Peak410	Sc0000694	23622	23922	123,06	0,95	BL19609, BL26095, BL39430, BL67979, BL67988, BL70565	Yes	Yes	Yes
Peak411	Sc0000708	10167	10467	281,36	0,68	BL00088, BL60112, BL79955, BL84080, BL84081	No	No	No
Peak412	Sc0000741	13073	13373	91,08	0,83	BL04040, BL04041, BL04042, BL24849, BL29904	No	No	No
Peak413	Sc0000764	18458	18758	236,84	0,95	BL19425, BL19427, BL43016	Yes	No	No
Peak414	Sc0000764	30530	30830	124,17	2,17	BL19425, BL71549, BL80197	Yes	No	No
Peak415	Sc0000764	31254	31554	132,40	1,65	BL19425, BL71549, BL80197	Yes	No	No
Peak416	Sc0000764	31992	32292	552,79	2,10	BL71549, BL80197	No	No	No
Peak417	Sc0000764	32752	33052	139,60	1,88	BL71549, BL80197	No	No	No
Peak418	Sc0000764	33447	33747	215,50	1,02	BL71549, BL80197	No	No	No
Peak419	Sc0000764	35265	35565	156,30	1,67	BL71549, BL80197	No	No	No
Peak420	Sc0000764	36368	36668	677,47	2,23	BL71549, BL80197, BL87046, BL87086	No	No	No
Peak421	Sc0000764	37450	37750	377,74	2,42	BL71549, BL80197, BL87046, BL87086	No	No	No
Peak422	Sc0000783	30838	31138	262,73	-0,75	BL03013, BL38050	No	No	No
Peak423	Sc0000884	8511	8811	129,00	0,78	BL57675, BL57677, BL57680, BL85963, BL85964, BL85965	No	No	No
Peak424	Sc0001069	36	336	271,49	1,46	BL09970, BL29558, BL29559, BL46916, BL46929	No	No	No
Peak425	Sc0001416	4696	4996	72,22	1,20	BL10029, BL39466	Yes	No	No
Peak426	xfSc0000138	7742	8042	126,79	-0,75	BL51149, BL51150, BL51151	No	No	No
Peak427	xfSc0000451	3476	3776	80,39	-1,11	BL23635	No	No	No
Peak428	xfSc0001959	391	691	90,24	-0,98		No	No	No
Peak429	xfSc0003633	35	335	380,03	-0,63	BL53604	No	No	No
Peak430	xfSc0005846	1054	1354	187,47	-0,90		No	No	No
Peak431	xfSc0006128	462	762	189,98	0,70		No	No	No
Peak432	xfSc0006780	749	1049	131,08	0,92		No	No	No
Peak433	xfSc0007015	1026	1326	82,63	0,84		No	No	No
Peak434	xfSc0019361	67	367	180,15	0,70		No	No	No

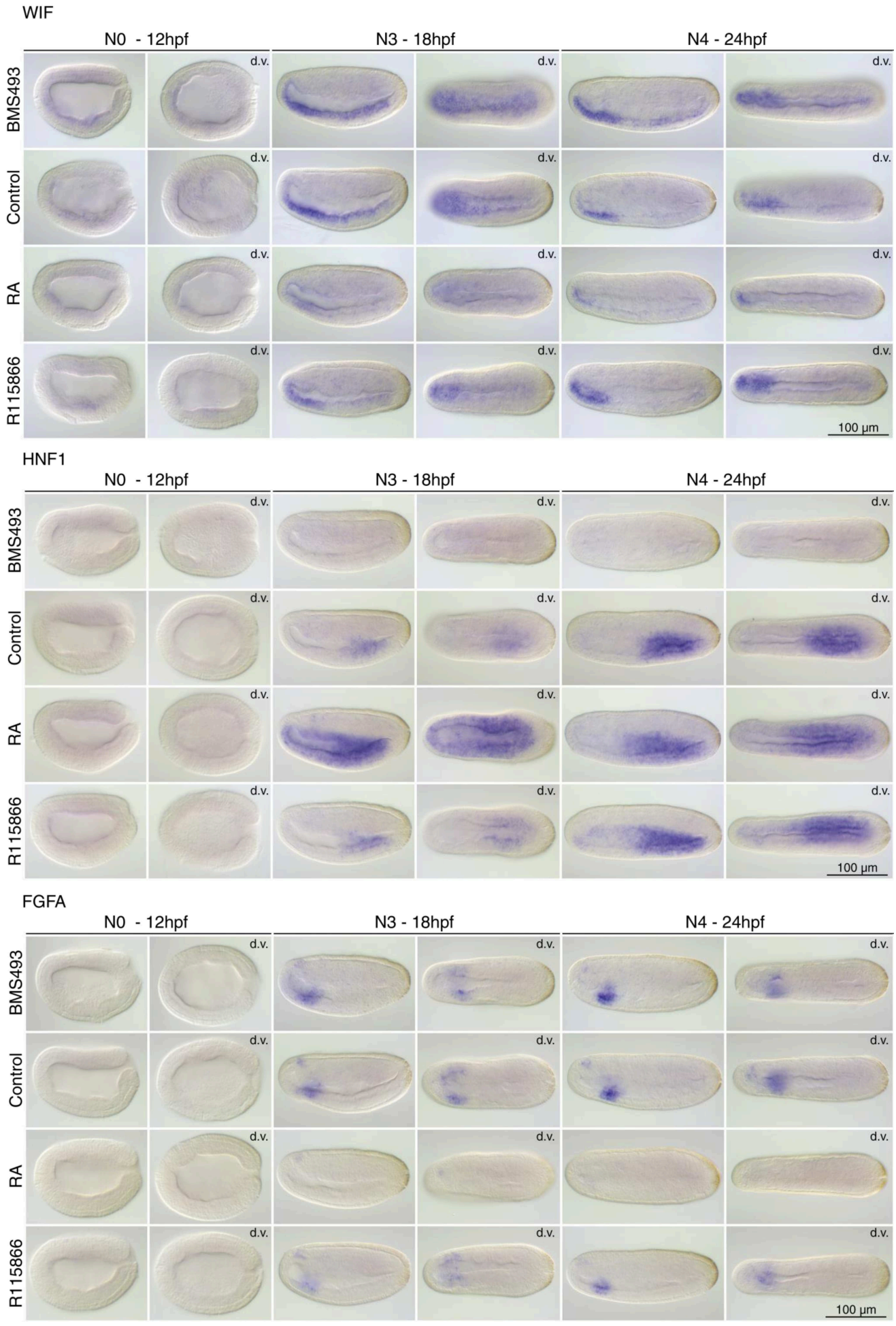
Peak ID	Scaf	start	end	bM	l2FC	Assigned genes	RA N3 vs RA N4 vs ATAC	Motifs NR	Motifs FKH
Peak435	xfSc0024266	91	391	230,31	0,71		No	No	No
Peak436	xpSc0040317	0	299	64,45	-0,86	BL56736, BL56737, BL76333	No	No	No

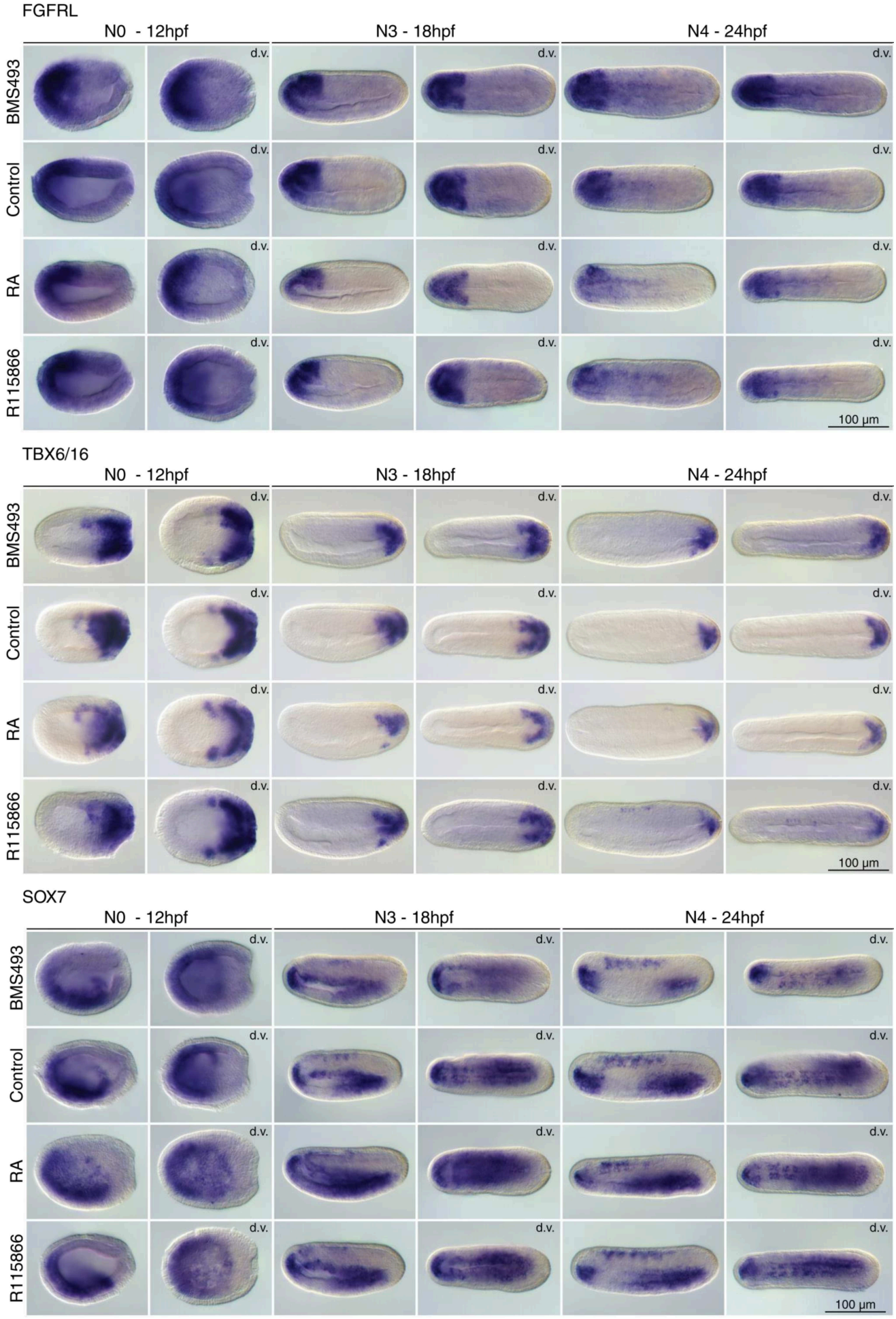
BMS493 treatment - ATACseq peaks

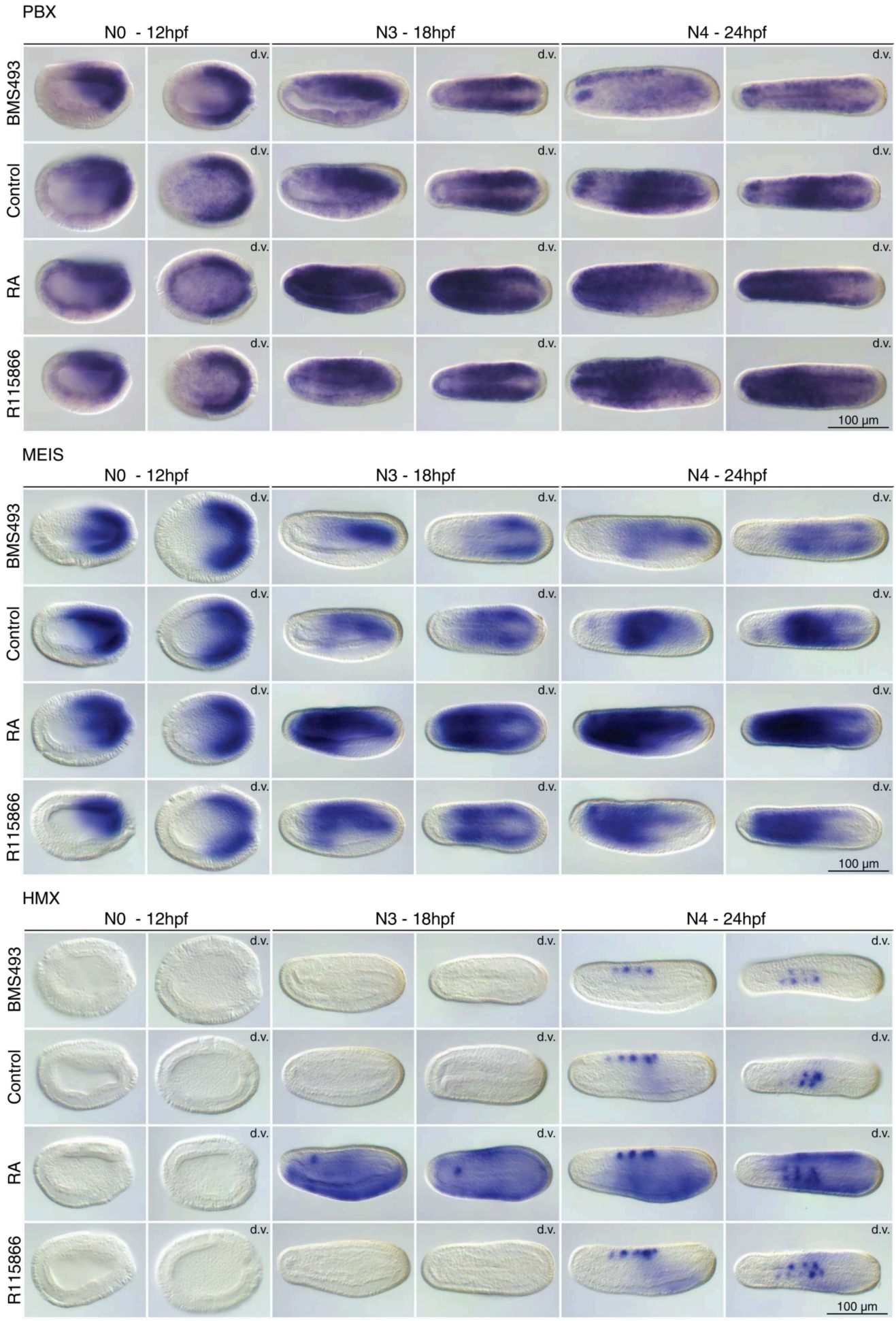
Peak ID	Scaf	start	end	bM	l2FC	Assigned genes
Peak437	Sc0000003	7590635	7590935	259,10	0,85	BL63406
Peak438	Sc0000004	510090	510390	88,97	1,02	BL00639 ,BL07309 ,BL08030 ,BL86382
Peak439	Sc0000015	721562	721862	107,29	1,02	BL09100 ,BL18378
Peak440	Sc0000017	812067	812367	100,20	0,97	BL01435 ,BL05124
Peak441	Sc0000042	280763	281063	94,47	1,00	BL23525 ,BL54069 ,BL55403
Peak442	Sc0000062	41350	41650	187,64	1,06	BL02215 ,BL85066
Peak443	Sc0000158	584667	584967	135,99	1,03	BL23731 ,BL70166
Peak444	Sc0000183	417523	417823	380,44	0,83	BL00236
Peak445	Sc0000185	440617	440917	148,74	1,33	BL06405 ,BL12955 ,BL67648 ,BL67649
Peak446	Sc0000200	10126	10426	291,37	-0,96	BL19441
Peak447	Sc0000327	8678	8978	269,01	0,94	BL33373 ,BL79532
Peak448	Sc0000764	33447	33747	215,50	-0,97	BL71549 ,BL80197

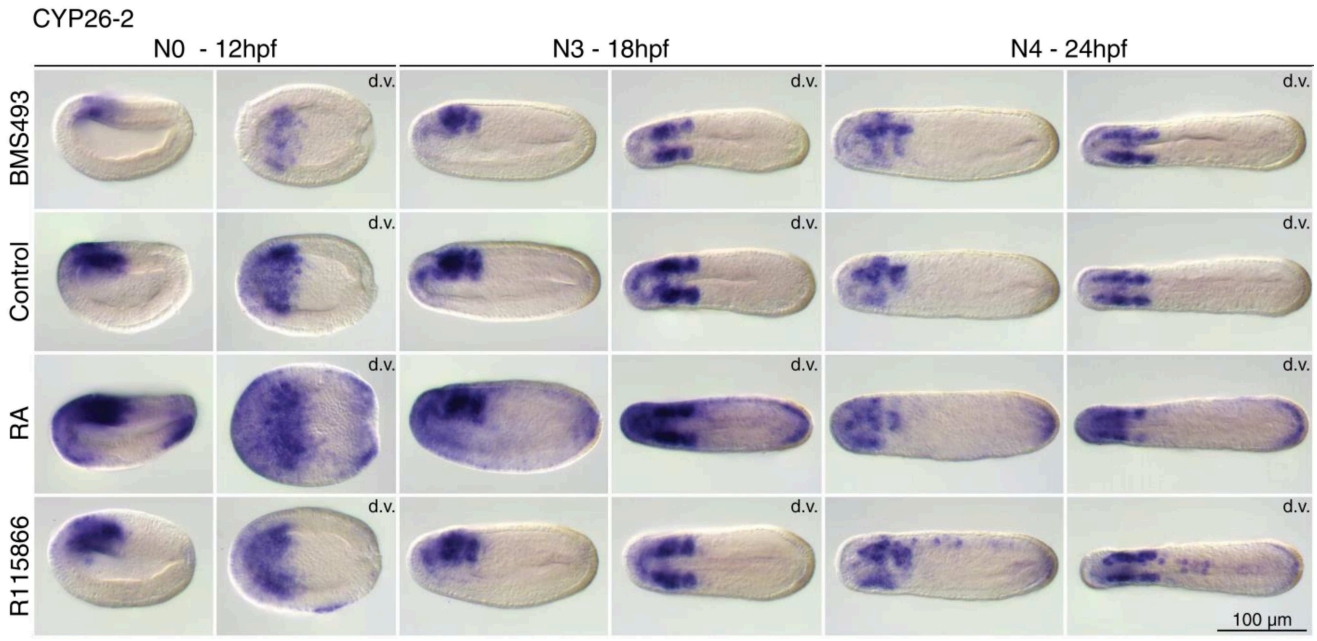
Additional file 3

		N0				N3				N4			
Gene	Auto nomenclature	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS493	l2FC R115866	l2FC RA	bM	l2FC BMS493	l2FC R115866	l2FC RA
BL10935	FOXAa (1)	10396,82	-0,18	NA	0,18	11197,01	NA	0,27	0,33	9599,49	-0,44	NA	0,34
BL01658	FOXAb (2)	NA	NA	NA	NA	10438,60	NA	0,27	0,22	9574,18	NA	NA	0,38
BL22763	FOXB1	NA	NA	NA	NA	1778,10	0,29	NA	NA	NA	NA	NA	NA
BL29729	FOXB1	NA	NA	NA	NA	226,99	0,50	NA	NA	NA	NA	NA	NA
BL11745	FOXC2	NA	NA	NA	NA	4279,25	0,17	NA	NA	NA	NA	NA	NA
BL16522	FOXE1	NA	NA	NA	NA	502,15	0,38	-0,38	-1,57	829,63	NA	-1,14	-1,71
BL11067	FOXG1	NA	NA	NA	NA	128,10	NA	NA	-0,62	199,19	NA	-0,93	-0,86
BL18773	FOXH1	NA	NA	NA	NA	2188,35	0,41	NA	-0,22	1275,89	NA	-0,33	-0,31
BL15774	FOXI1	NA	NA	NA	NA	NA	NA	NA	NA	512,69	NA	-0,82	-1,80
BL19556	FOXJ3	NA	NA	NA	NA	1295,95	NA	NA	0,23	NA	NA	NA	NA
BL12207	FOXK1	NA	NA	NA	NA	2552,19	0,19	NA	NA	NA	NA	NA	NA
BL20927	FOXL1	NA	NA	NA	NA	NA	NA	NA	NA	98,68	NA	-0,71	-0,91
BL07280	FOXL2	NA	NA	NA	NA	NA	NA	NA	NA	123,47	NA	-0,67	-1,26
BL10613	FOXN4	472,35	NA	0,28	NA	NA	NA	NA	NA	NA	NA	NA	NA
BL02241	FOXP1	NA	NA	NA	NA	5003,66	NA	0,18	NA	NA	NA	NA	NA
BL00713	FOXQ1	NA	NA	NA	NA	72,38	NA	NA	-1,01	683,39	NA	-0,45	-0,85
BL11561	CDX4	NA	NA	NA	NA	9419,51	0,25	NA	NA	8435,57	NA	-0,24	-0,35
BL01497	HOXB6	NA	NA	NA	NA	32,32	NA	NA	1,52	32,46	-0,86	1,01	1,22
BL18685	OTX2	NA	NA	NA	NA	5491,42	0,27	NA	-0,27	3386,35	NA	-0,32	NA


WIF
N0 - 12hpf
N3 - 18hpf
N4 - 24hpf
BMS493
Control
RA
R115866
d.v.
100 µm
HNF1
N0 - 12hpf
N3 - 18hpf
N4 - 24hpf
BMS493
Control
RA
R115866
d.v.
100 µm
FGFA
N0 - 12hpf
N3 - 18hpf
N4 - 24hpf
BMS493
Control
RA
R115866
d.v.
100 µm


FGFRL
N0 - 12hpf
N3 - 18hpf
N4 - 24hpf
BMS493
Control
RA
R115866
d.v.
100 µm
TBX6/16
N0 - 12hpf
N3 - 18hpf
N4 - 24hpf
BMS493
Control
RA
R115866
d.v.
100 µm
SOX7
N0 - 12hpf
N3 - 18hpf
N4 - 24hpf
BMS493
Control
RA
R115866
d.v.
100 µm


PBX
N0 - 12hpf
N3 - 18hpf
N4 - 24hpf
BMS493
Control
RA
R115866
d.v.
100 µm
MEIS
N0 - 12hpf
N3 - 18hpf
N4 - 24hpf
BMS493
Control
RA
R115866
d.v.
100 µm
HMX
N0 - 12hpf
N3 - 18hpf
N4 - 24hpf
BMS493
Control
RA
R115866
d.v.
100 µm


CYP26-2
N0 - 12hpf
N3 - 18hpf
N4 - 24hpf
BMS493
Control
RA
R115866
d.v.
100 µm

Appendix 4

Published book chapter

Retinoic acid: metabolism, developmental functions, and evolution.

Carvalho JE, Schubert M.

In: Dakshinamurti K, Dakshinamurti S, eds. Vitamin-binding proteins: functional consequences. CRC Press; 2013. 1–30.

(30 pages)

1 Retinoic Acid
Metabolism, Developmental Functions, and Evolution

João E. Carvalho and Michael Schubert

CONTENTS

1.1 INTRODUCTION

A large amount of data collected during the past one hundred years has established that vitamin A (also known as retinol) and its derivatives, the retinoids (Figure 1.1), play crucial roles during vertebrate development. The retinoids are fat-soluble morphogens that are involved in a variety of processes, such as pattern formation during early development as well as in organogenesis, cell proliferation, cell differentiation, apoptosis, vision, immune responses, and tissue homeostasis (Blomhoff and Blomhoff 2006, Campo-Paysaa et al. 2008, Collins and Mao 1999, Duester 2008, Glover et al. 2006, Hall et al. 2011, Mark et al. 2006, Niederreither and Dollé 2008,

Noy 2010, Theodosiou et al. 2010). Initial insights into retinol-dependent developmental processes were obtained from nutritional excess and deficiency studies carried out in different animal models including pigs and rats (Hale 1933, Wilson and Warkany 1949). In particular, the so-called vitamin A deficiency (VAD) models produced valuable information about the biological roles of retinol in different organs during embryonic development (Wilson et al. 1953). The subsequent emergence of genetic and molecular tools allowed a much more detailed dissection of the developmental functions of retinol and its derivatives including the identification of the molecular machinery underlying retinoid-dependent signaling (Giguere et al. 1987, Mark et al. 2006, Petkovich et al. 1987).

It is now generally accepted that all-*trans* retinoic acid (RA) is the main biologically active form of retinol, although a biological role for other retinoids, such as 9-*cis* RA, is still a matter of debate (Kane 2012, Theodosiou et al. 2010) (Figure 1.1). Moreover, although alternative mechanisms for triggering retinoid-dependent signaling have recently been proposed (Theodosiou et al. 2010), in vertebrates the main cellular mechanism of RA action occurs by binding of RA to heterodimers of two nuclear receptors: the retinoic acid receptor (RAR) and the retinoid X receptor (RXR) (Mark et al. 2006). This classical mode of retinoid action involves fixation of RA by the DNA-bound RXR/RAR heterodimer, the subsequent recruitment of co-activators, and the ultimate activation of transcription of target genes (Figure 1.2) (Chambon 1996).

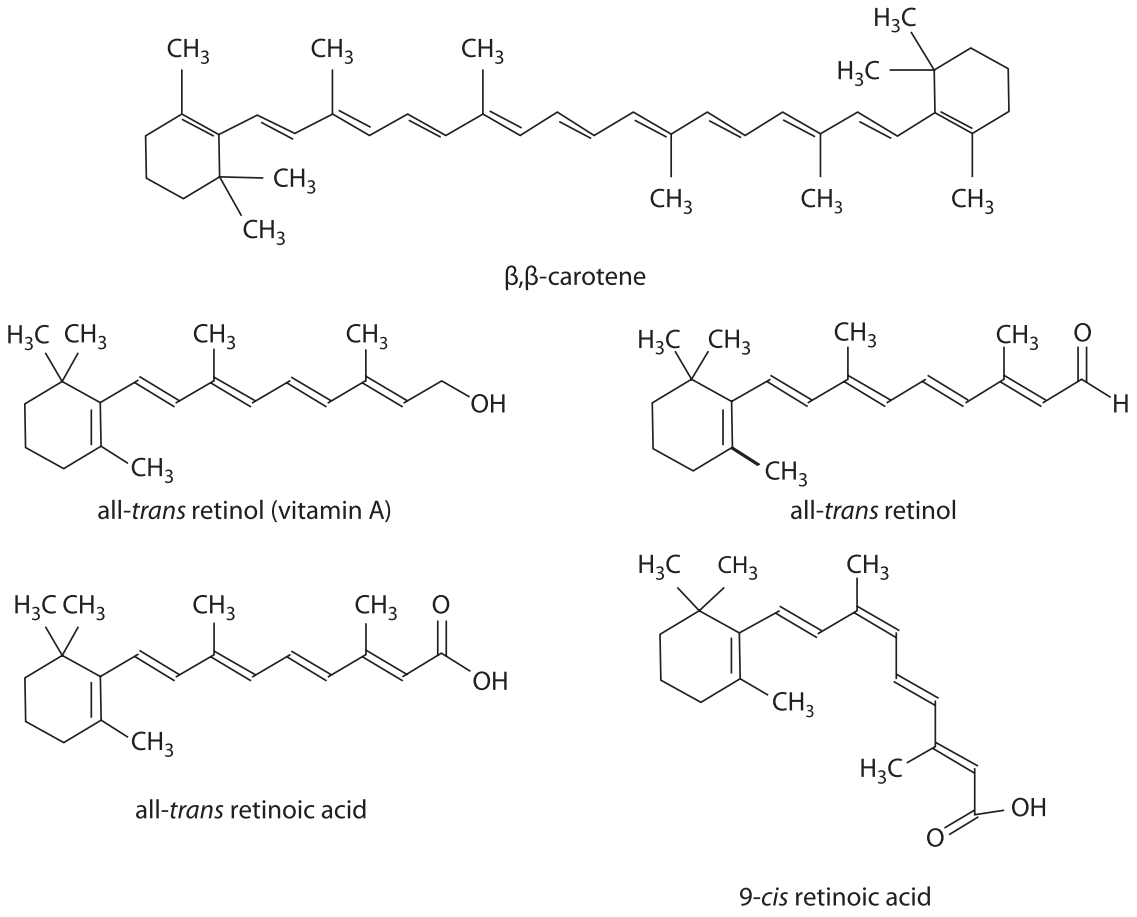


FIGURE 1.1 Chemical structures of different retinoids.

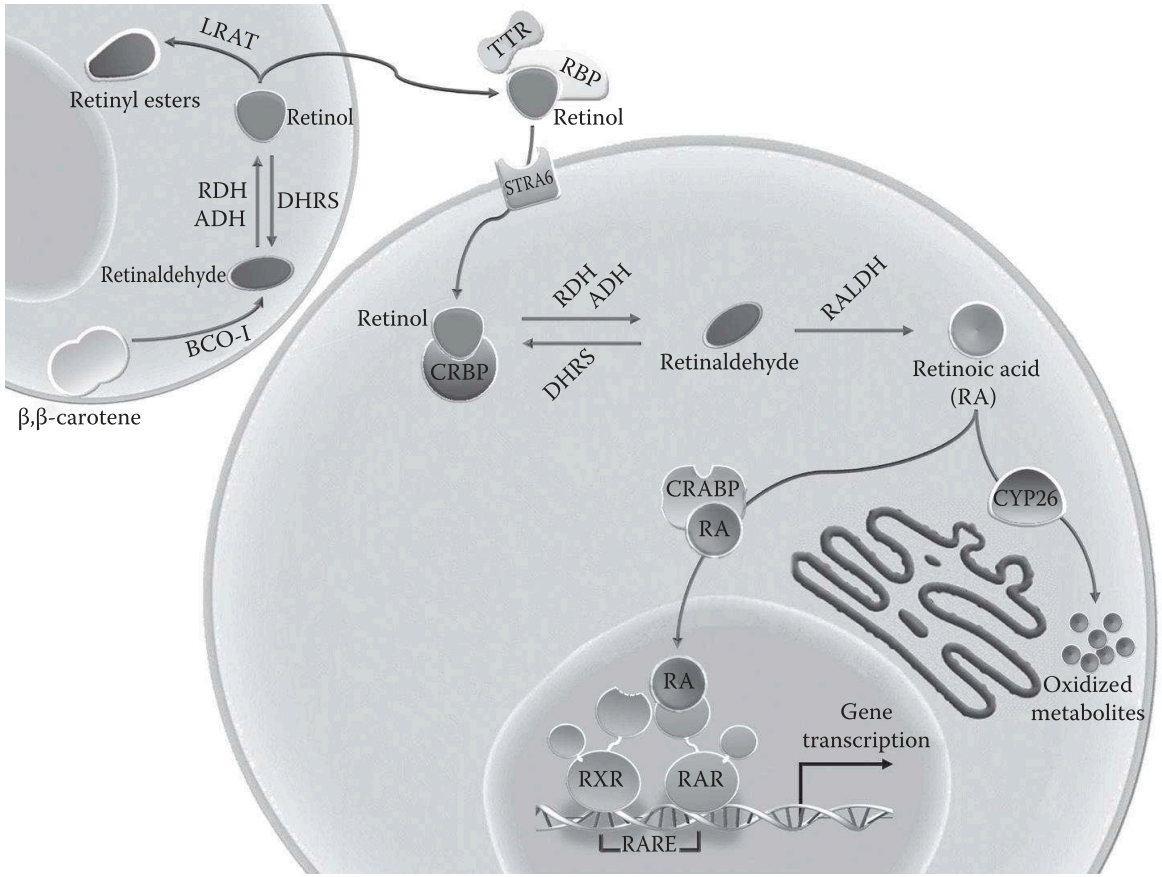


FIGURE 1.2 Retinoid metabolism. Conversion of β,β-carotene to retinol (vitamin A), synthesis of retinoic acid, the major biologically active metabolite, and activation of retinoid-dependent signaling. ADH, alcohol dehydrogenase; BCO-I, β,β-carotene-15,15′-monooxygenase 1; CRABP, cellular retinoic acid-binding protein; CRBP, cellular retinol-binding protein; CYP26, cytochrome P450 family 26; DHRS, dehydrogenase/reductase SDR family member; LRAT, lecithin:retinol acyltransferase; RA, retinoic acid; RALDH, retinaldehyde dehydrogenase; RAR, retinoic acid receptor; RARE, retinoic acid response element; RBP, retinol-binding protein; RDH, retinol dehydrogenase; RXR, retinoid X receptor; STRA6, stimulated by retinoic acid gene 6; TTR, transthyretin.

In this review, we first describe the canonical RA pathway in vertebrates, whose components include proteins responsible for RA production, degradation, and transport. Later, we discuss recently described alternative mechanisms of RA function, followed by a summary of the main developmental functions of RA signaling in vertebrates. We conclude by comparing the available data on RA in vertebrates with the latest information on RA signaling obtained from non-vertebrate animal models.

1.2 RETINOID METABOLISM

Due to their inability to synthesize retinol *de novo*, animals need to obtain retinoids through the diet, mainly in the form of carotenoids, such as β,β-carotene (Figure 1.1). Carotenoids are isoprenoids that are widespread in nature and are typically associated with the yellow, orange, red, or purple colors found in vegetables, fruits, flowers, birds, and crustaceans (Fraser and Bramley 2004). Alternatively, retinol can also be taken up directly in the form of pro-vitamin A carotenoids (Blomhoff and Blomhoff 2006).

The conversion of ingested carotenoids, such as β,β-carotene, into retinaldehyde is the first step of the classical retinoid metabolism (Figure 1.2). This process is driven by the enzyme β,β-carotene-15,15′-monooxygenase 1 (BCO-I) (von Lintig et al. 2005). This enzyme is crucial for proper embryonic development, at least in diets with β,β-carotene as the only source of retinol. For example, BCO-I knockout mice accumulate high levels of β,β-carotene (Hessel et al. 2007), a condition that is similar in zebrafish embryos (Lampert et al. 2003). Additional to BCO-I, β,β-carotene-9,10-dioxygenase 2 (BCO-II) has been described as a second carotenoid oxygenase in vertebrates (Kiefer et al. 2001). BCO-II mainly functions in regulating cell survival, proliferation, and preventing oxidative stress in accordance with its mitochondrial location (Lobo et al. 2012). Additionally, BCO-II might also function as the key enzyme for a non-canonical pathway for RA production (Simões-Costa et al. 2008), but this involvement of BCO-II in retinol metabolism remains to be confirmed *in vivo* (Spiegler et al. 2012).

Retinaldehyde is subsequently hydrolyzed to retinol, which can be used for two different purposes: esterification and tissue storage or oxidative metabolism leading to synthesis of RA, the main biologically active retinoid. The process of esterification is conducted by the enzyme lecithin:retinol acyltransferase (LRAT), which produces retinyl esters (Figure 1.2). Under retinoid oversupply, LRAT has been proposed to play a crucial role in maintaining retinol homeostasis (Ross 2003). For RA synthesis, retinol is transported by retinol-binding protein (RBP) to target tissues (D'Ambrosio et al. 2011). RBP is a specific transport protein circulating in the blood and is associated one to one with another serum protein, transthyretin (TTR) (Figure 1.2) (Richardson 2009, Yamauchi and Ishihara 2009).

In mice, RBP, which cannot cross the placenta (Quadro et al. 2004), has been detected at the junction of the uterine wall and the placenta at E13.5 (embryonic day 13.5) and during subsequent development in the visceral endoderm of the yolk sac (Spiegler et al. 2012). Similarly, RBP has been detected in developing chick and zebrafish embryos (Li et al. 2007, Quadro et al. 2005). Given that RBP is the only known specific carrier protein of retinol, it is rather surprising that mice lacking

RBP ($RBP^{-/-}$) are viable, showing no major malformations when exposed to a retinol-sufficient diet (Barron et al. 1998). Moreover, using an elegant genetic approach, it was shown that, in a BCO-$I^{-/-}$/$RBP^{-/-}$ background, BCO-$I^{+/-}$/$RBP^{-/-}$ offspring can develop normally and maintain retinoid homeostasis, when BCO-$I^{-/-}$/$RBP^{-/-}$ pregnant mice are maintained under a VAD diet with β,β-carotene being the only source of retinoids administrated via intra peritoneal injection (Kim et al. 2011). These results also suggest that β,β-carotene alone can be an adequate source of maternal retinoids (Kim et al. 2011).

The cellular uptake of retinol is carried out by an RBP membrane receptor, called stimulated by retinoic acid gene 6 (STRA6) (Kawaguchi et al. 2007) (Figure 1.2). STRA6 works as a bidirectional retinol transporter, and its activity is controlled by intracellular retinoid levels (D'Ambrosio et al. 2011). In addition, expression of STRA6 is also regulated by RA (D'Ambrosio et al. 2011). Interestingly, STRA6 does not seem to function like other known membrane receptors, transporters, or channels, and there are, at present, no data available about the molecular mechanisms of STRA6-dependent substrate uptake (Sun 2012).

Once inside the cell, RA is synthesized from retinol in two oxidation steps. The first step consists of the reversible oxidation of retinol into retinaldehyde, which is carried out by enzymes of two different classes: the cytosolic alcohol dehydrogenases (ADHs), belonging to the medium-chain dehydrogenase/reductase family, and the microsomal retinol dehydrogenases (RDHs), which are members of the short-chain dehydrogenase/reductase family (Parés et al. 2008) (Figure 1.2). Several members of the ADH family, including ADH1, ADH3, and ADH4, have been shown to catalyze the oxidation of retinol into retinaldehyde *in vitro* (Duester 2008, Parés et al. 2008). Mice carrying null mutations for ADH1, ADH3, and ADH4 have revealed partially redundant functions of these three ADHs (Parés et al. 2008). Moreover, while both ADH1 and ADH4 seem to be required specifically in conditions of retinoid deficiency or excess, ADH3 can probably be regarded as a ubiquitous ADH (Theodosiou et al. 2010). Like for the ADHs, multiple RDHs, such as RDH10, have been shown to be capable of synthesizing retinaldehyde from retinol *in vitro* (Parés et al. 2008). However, recent experimental evidence suggests that the most important player in retinaldehyde synthesis during embryonic development is RDH10 (Farjo et al. 2011). Loss of RDH10 functions in mice is lethal between E10.5 and E14.5 with embryos showing severe defects in embryonic patterning and morphogenesis (Sandell et al. 2007). It has recently been suggested that the regulation of endogenous retinaldehyde levels represents an additional mechanism for controlling RA signaling activity in developing embryos: within a given cell, RA production is thus dependent on the balance between retinaldehyde synthesis from retinol and retinaldehyde reduction to retinol. While in developing embryos the former reaction, as mentioned previously, is catalyzed by ADH or RDH10, the latter reaction can be carried out, for example, by DHRS3a (dehydrogenase/reductase SDR family member 3a), which, like RDH10, is a member of the short-chain dehydrogenase/reductase family (Feng et al. 2010).

The second oxidation step is the conversion of retinaldehyde into RA, an irreversible reaction carried out by retinaldehyde dehydrogenases (RALDHs) (Figure 1.2). Three RALDHs of the ALDH1 class (RALDH1, 2, and 3 also called ALDH1A1, ALDH1A2, and ALDH1A3, respectively) and one RALDH of the ALDH8 class

(RALDH4 also called ALDH8A1) have been implicated in endogenous RA production with each enzyme presenting a distinct and tissue-specific distribution (Theodosiou et al. 2010). RALDH2 is the earliest RALDH enzyme to be expressed during development, and its expression is crucial for early embryogenesis, since mice lacking RALDH2 exhibit severe abnormalities and die before mid-gestation (Niederreither et al. 1999). These phenotypes can be almost completely rescued by maternal administration of RA, suggesting that RALDH2 is indeed required for endogenous RA synthesis during development (Mic et al. 2002, Niederreither et al. 1999). RALDH2 homologs in the frog *Xenopus laevis* and in zebrafish have been shown to have similar roles during early embryonic development (Begemann et al. 2001, Chen et al. 2001). RALDH1, RALDH3, and RALDH4 are expressed at later developmental stages, with RALDH1 and RALDH3 contributing to the patterning of the respiratory and visual systems (Duester et al. 2003, Rhinn and Dollé 2012) and RALDH4 being detectable in the fetal liver (Lin et al. 2003).

Some reports have suggested that RA synthesis during embryonic development may also occur independently of RALDHs, possibly through the action of members of the cytochrome P450 family of monooxygenases (Collins and Mao 1999). Functional studies point to CYP1B1 as a potential candidate (Chambers et al. 2007), because this enzyme can efficiently oxidize retinol to retinaldehyde and subsequently to RA and exhibits an expression pattern consistent with RA synthesis patterns in the developing embryo (Chambers et al. 2007).

Catabolism of endogenous RA is driven by enzymes of the cytochrome P450 subfamily 26 (CYP26A1, CYP26B1, and CYP26C1), which are responsible for the oxidation of RA into metabolites (Figure 1.2), such as 4-hydroxy-RA, 4-oxo-RA, or 18-hydroxy-RA (Bempong et al. 1995). All three CYP26 enzymes function in embryonic development and exhibit dynamic expression patterns both in time and space (Reijntjes et al. 2004, Ross and Zolfaghari 2011). CYP26A1 expression starts at the gastrula stage in mouse endoderm, in chick it is initially expressed in ectoderm and mesoderm close to Hensen's node, whereas in frogs and zebrafish the gene is found in the anterior ectoderm during early gastrulation (White and Schilling 2008).

Mice devoid of CYP26A1 die before the first postnatal day presenting a posteriorized hindbrain and vertebral column as well as severe caudal truncations (Abu-Abed et al. 2001). In contrast, CYP26B1 seems to be fundamentally involved in limb patterning and outgrowth (Yashiro et al. 2004) and is required for maintenance of germ cells during testes development (MacLean et al. 2007). Like CYP26A1, CYP26C1 is expressed in the hindbrain, but mice deficient for CYP26C1 are viable and do not show severe anatomical abnormalities, suggesting some degree of functional redundancy between CYP26A1 and CYP26C1 (Uehara et al. 2007). Double mutants for CYP26A1 and CYP26C1 have a much stronger phenotype when compared to both CYP26A1 or CYP26C1 single mutants. Simultaneous loss of RALDH2 partially rescues these phenotypes, confirming that the malformations observed after loss of both CYP26A1 and CYP26C1 are due to elevated levels of RA (Uehara et al. 2007).

Retinoids present within a given cell are bound by intracellular-binding proteins, such as cellular retinol-binding proteins (CRBPs) or cellular retinoic acid-binding proteins (CRABPs) (Figure 1.2), all of which exhibit extremely high affinities for their substrates (Noy 2000). CRBP-I, for example, protects retinol from oxidation

and isomerization by limiting the interaction with enzymes capable of recognizing retinol (Napoli 2000). During development, CRBP-I is expressed in several tissues including motor neurons, spinal cord, lung, liver, and placenta (Ghyselinck et al. 1999). Mice lacking CRBP-I do not exhibit any morphological abnormalities, suggesting that RA metabolism is not dependent on CRBP-I function (Ghyselinck et al. 1999). Nonetheless, in the absence of CRBP-I, retinol stocks are severely depleted (Matt et al. 2005). Compared to CRBP-I, CRBP-II has a more restricted expression pattern, being mainly localized in fetal intestine and liver, but also in the yolk sac from E10.5 to E15.5 (Noy 2000). Its expression suggests a role as binding protein for recently absorbed retinol (Suruga et al. 1997). Interestingly, developmental expression of CRBP-I overlaps significantly with that of CRABP-II, while the patterns of CRBP-I and CRABP-I are mutually exclusive (Napoli 1999).

Several functions have been proposed for the two CRABP proteins, but their antagonistic effect on RA transport seems to be the most important: while CRABP-I seems to mediate CYP26-dependent RA degradation (Napoli 1999), CRABP-II is probably responsible for the transport of RA to its receptors in the nucleus, a fact supported by the increase of RA signaling when CRABP-II is overexpressed in cell lines and *Xenopus* embryos (Dong et al. 1999, Noy 2000). Although CRABP-I$^{-/-}$ and CRABP-II$^{-/-}$ single mutant as well as CRABP-I$^{-/-}$/CRABP-II$^{-/-}$ double mutant mice failed to reveal the biological functions of CRABPs (Napoli 2000), a recent analysis in zebrafish has highlighted the importance of CRABPs in stabilizing the RA morphogen gradient in the developing hindbrain by delivering RA both to its receptors and to CYP26 for degradation (Cai et al. 2012).

1.3 CLASSICAL RETINOID SIGNAL TRANSDUCTION: THE RXR/RAR HETERODIMER

Until very recently, the main dogma of retinoid research stated that RA functions are mediated by heterodimers of two nuclear receptors: RAR and RXR (Chambon 1996, Gronemeyer et al. 2004). There are usually three RAR genes encoded in mammalian genomes (RARα, RARβ, and RARγ) (Linney et al. 2011), while the zebrafish, for example, has lost RARβ but duplicated both RARα (yielding RARαa and RARαb) and RARγ (yielding RARγa and RARγb) (Waxman and Yelon 2007). Congruent with the overall RAR toolkit, mammalian genomes generally also possess three RXR genes (RXRα, RXRβ, and RXRγ). In contrast, the zebrafish genome, for instance, encodes six RXR genes (RARαa, RARαb, RARβa, RARβb, RARγa, and RARγb) (Waxman and Yelon 2007). The RXR/RAR heterodimer acts as ligand-dependent transcription factor by binding to RA response elements (RAREs) located within the regulatory DNA of RA target genes. *In vitro* studies have shown that RAR is able to bind and to be activated by both all-*trans* RA and 9-*cis* RA, whereas RXR binds and is activated only by 9-*cis* RA. However, it remains to be determined whether 9-*cis* RA is present in vertebrate embryos and if its interaction with RXR effectively triggers a biological process (Mic et al. 2003).

RXR/RAR typically associates with two DNA sequence stretches composed of direct repeats (DRs) with the conserved nucleotide sequence (A/G)G(G/T)(G/T)(G/C)A. The spacing between the two DRs is typically one, two, or five nucleotides

long (Balmer and Blomhoff 2005, Chambon 1996, Ross et al. 2000). According to the canonical model, RXR/RAR heterodimers bind to DNA even in the absence of ligand, thereby recruiting a co-repressor complex that mediates chromatin compaction and target gene repression (Vilhais-Neto and Pourquié 2008). In the presence of ligand, due to a conformational change of the receptor induced by the ligand, there is dissociation of the co-repressor complex and recruitment of a co-activator complex. This binding of co-activators mediates chromatin decondensation and assembly of the transcriptional pre-initiation complex ultimately leading to target gene transcription (Rhinn and Dollé 2012).

Genetic studies have established that the three mammalian RAR paralogs have both specific and partially redundant functions during embryogenesis (Matt et al. 2003). Interestingly, RARα$^{-/-}$, RARβ$^{-/-}$, and RARγ$^{-/-}$ single mutant mice are able to reach adulthood, presenting only minor morphological phenotypes indicative of VAD syndromes (Mark et al. 2009). In contrast, RARα$^{-/-}$/RARβ$^{-/-}$, RARα$^{-/-}$/RARγ$^{-/-}$, and RARβ$^{-/-}$/RARγ$^{-/-}$ double mutant mice die during embryonic development exhibiting severe malformations typical of VAD (Mark et al. 2009). RXRα$^{-/-}$ mice present several severe abnormalities and die during development, with most of the abnormalities equally found in the different RAR$^{-/-}$ compound mutants (Mark et al. 2009). In contrast, RXRβ$^{-/-}$/RXRγ$^{-/-}$ double mutant mice do not show any evident morphogenetic defects, indicating that RXRα is the most important RXR implicated in morphogenetic patterning during development (Krezel et al. 1996). RAR$^{-/-}$ and RXRα$^{-/-}$ mutations seem to act in synergy, since null mutations for individual RAR paralogs combined with null mutations of RXRα reveal potentially redundant functions of the different RAR paralogs. These different compound mutants thus reveal the fundamental functions of the RXRα/RARα, RXRα/RARβ, and RXRα/RARγ heterodimers in transducing the RA signal during embryogenesis (Samarut and Rochette-Egly 2012).

1.4 NON-CLASSICAL RETINOID SIGNAL TRANSDUCTION

In addition to the classical mode of retinoid action, mediated by RXR/RAR heterodimers, several alternative mechanisms for triggering retinoid-dependent signaling cascades in vertebrates have recently been proposed. For example, RA has been shown to specifically activate the orphan receptor peroxisome proliferator-activated receptor β/δ (PPARβ/δ), but neither one of its paralogs, PPARα and PPARγ (Shaw et al. 2003). Interestingly, CRABP does not play a role in the delivery of RA to this receptor, as this function is carried out by FABP5 (Schug et al. 2007, Tan et al. 2002). Despite the evident existence of this alternative pathway, the affinity of RA for association with FABP5/PPARβ/δ is much lower than its affinity for CRABP/RAR (Schug et al. 2007). Importantly, the ratio of the proteins involved in intracellular RA delivery (FABP5 and CRABP) has been shown to be a crucial component for triggering alternative versus classical RA responses (Schug et al. 2007). Given that the activation of the FABP5/PPARβ/δ pathway leads to a very different biological output both in terms of cell survival (Schug et al. 2007) and differentiation (Berry and Noy 2009), this alternative pathway thus significantly increases the complexity of the RA signaling response.

A number of other nuclear receptors have also been shown to bind and be activated by retinoids. For example, testicular receptor 4 (TR4) has recently been suggested to be a retinoid receptor, because the binding of both retinol and RA can induce conformational changes of TR4 leading to the activation of this receptor (Zhou et al. 2011). TR4 can either function as homodimer or heterodimerize with testicular receptor 2 (TR2) to function in spermatogenesis, lipid and lipoprotein regulation, and central nervous system development (Zhou et al. 2011). However, additional experiments are required to reveal the physiological and developmental implications of the activation of TR4 by retinoids, such as RA (Zhou et al. 2011).

Like TR4, COUP-TFII (chicken ovalbumin upstream promoter-transcription factor II) is capable of recruiting co-activators and thus of activating reporter construct expression in the presence of retinoids (Kruse et al. 2008). Both all-*trans* RA and 9-*cis* RA can associate with, and activate, COUP-TFII. However, the biological relevance of this interaction is questionable because the retinoid concentration needed to trigger a response might be significantly higher than endogenous retinoid levels (Kruse et al. 2008). Future work will thus have to assess the potential role of retinoids in mediating COUP-TFII-specific functions, for example, in neuronal development, cell fate determination, and circadian rhythm (Kruse et al. 2008).

RA is also able to bind to RA receptor-related orphan receptor β (RORβ) *in vitro*. This receptor functions in the regulation of the circadian rhythm, and its activity has been shown to be regulated by RA *in vivo* (Stehlin-Gaon et al. 2003). Given that RAR has also been shown to inhibit a fundamental protein of the circadian clock, CLOCK/BMAL1, at least *in vitro* (McNamara et al. 2001, Shirai et al. 2006), the modification of RORβ functions by RA might represent an alternative pathway for integrating the circadian clock. Further studies are evidently required to understand how RORβ acts in association with RA.

Members of completely different protein families have also been suggested to convey retinoid signals. For instance, the association of retinol-bound RBP with STRA6 has been shown to trigger a signaling cascade mediated by the Janus kinase JAK2 and its associated transcription factors, the signal transducers and activators of transcription (STATs) (Berry and Noy 2012). Activation of the JAK/STAT cascade by retinol-bound RBP also leads to an increase of expression of STAT target genes, such as inhibitors of insulin signaling and regulators of lipid homeostasis (Berry and Noy 2012).

Moreover, RA has been shown to modulate the activity of protein kinase C α (PKCα). Thus, all-*trans* RA and acidic phospholipids compete for binding to PKCα, and this competitive binding is responsible for controlling PKCα activity (Ochoa et al. 2003). RA can also activate cAMP response element-binding protein (CREB) in cell lines, where RXR/RAR heterodimers have been silenced (Aggarwal et al. 2006). RA rapidly activates CREB, which leads to a concomitant increase of DNA binding by activated CREB (Aggarwal et al. 2006). RA is also responsible for an increase in the activity of both extracellular-signal-regulated kinases 1/2 (ERK1/2) and ribosomal s6 kinase (RSK), an important kinase acting upstream of CREB (Aggarwal et al. 2006). It has been suggested that PKC is involved in this RA-dependent activation (Aggarwal et al. 2006). Another study has shown that RA activates ERK1/2 through induction of mitogen-activated protein kinase kinase 1/2 (MEK1/2), and that this specific activation leads to caspase-3-dependent apoptosis (Zanotto-Filho et al. 2008).

The PI3K/Akt (phosphatidylinositol 3-kinase/v-akt murine thymoma viral oncogene homolog) signaling pathway has also been shown to be activated by RA in a transient process essential for neural differentiation (López-Carballo et al. 2002). Interestingly, an atypical orphan nuclear receptor, DAX1 (dosage-sensitive sex reversal-adrenal hypoplasia critical region on the chromosome X protein 1), has been implicated in this process (Nagl et al. 2009): acting through PI3K/Akt signaling, RA increases DAX1 expression, which leads to the modification of nitric oxide signaling levels, hence modulating neural differentiation (Nagl et al. 2009).

In addition to the aforementioned classical mechanisms of RA signaling, which involve binding of RXR/RAR heterodimers to DNA, RXR/RAR-dependent RA activity may also occur independently of DNA binding, a feature referred to as non-genomic signaling. One example of a non-genomic signaling response to RA is the RAR-dependent activation of p38MAPK (p38 mitogen-activated protein kinase) and of the downstream mitogen and stress-activated kinase 1 (MSK1) (Duong and Rochette-Egly 2011, Piskunov and Rochette-Egly 2011). This non-genomic signaling mechanism that ultimately leads to the activation of cdk7 (cyclin-dependent kinase 7) is mediated by a fraction of cellular RAR anchored in lipid rafts at the cell membrane, where RAR forms a complex with G protein alpha q (Gαq) (Piskunov and Rochette-Egly 2011). A fraction of RAR proteins thus functions in an RA-dependent, non-genomic manner to integrate processes occurring at the cell membrane. These roles played by RA and RAR might serve as fine-tune mechanisms for the rapid integration of extracellular cues. Future studies will certainly be needed to address the biological roles of these non-genomic signaling cascades during development.

1.5 RETINOID FUNCTIONS DURING DEVELOPMENT

During vertebrate development, RA regulates a significant number of pleiotropic effects, most of which, however, are not exclusively dependent on RA (Gutierrez-Mazariegos et al. 2011). These processes also rely on signals from other intercellular signaling pathways, such as the fibroblast growth factor (FGF) (Yasuda et al. 1992), sonic hedgehog (SHH) (Ribes et al. 2009), WNT, NODAL (Engberg et al. 2010), and bone morphogenetic protein (BMP) (Sheng et al. 2010) cascades. This combined action of RA and other signaling pathways can have very different architectures. Interactions can be synergistic or antagonistic (Gutierrez-Mazariegos et al. 2011), they may act directly or indirectly on a given target gene (Gutierrez-Mazariegos et al. 2011), they may also regulate upstream or downstream effectors of a particular cascade (Gutierrez-Mazariegos et al. 2011), or be involved in one or several regulatory loops (Gutierrez-Mazariegos et al. 2011). In the following section, we will discuss various tissue-specific roles of RA signaling during development (Figure 1.3) and highlight examples of some of the interactions of the RA pathway with other signaling cascades.

1.5.1 Central Nervous System Patterning Driven by RA

There is an extensive amount of experimental evidence describing roles of RA signaling in the developing central nervous system (CNS). For example, it has been

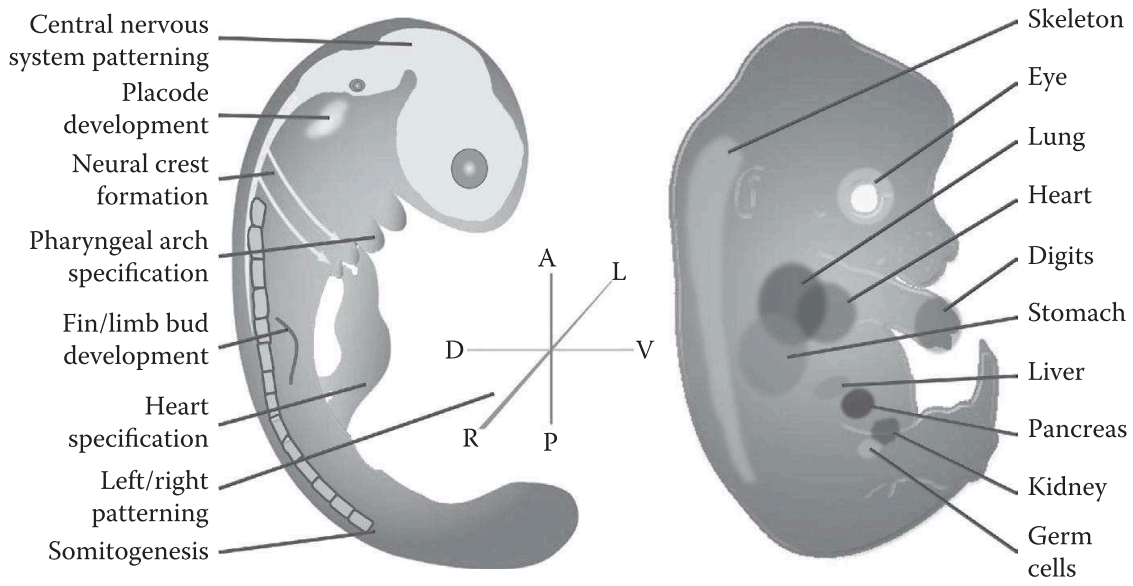


FIGURE 1.3 Summary of retinoic acid (RA) signaling functions during development. The schematic drawings of two vertebrate embryos chiefly correspond to early somitogenesis (left) and organogenesis (right) stages. RA signaling is involved in a multitude of developmental processes (Duester 2008, Kam et al. 2012, Rhinn and Dollé 2012) and the illustration highlights the roles of this signaling cascade detailed in the text.

known for a long time that treatment of vertebrate embryos with exogenous RA results in the loss of forebrain/midbrain territories with a significant expansion of the hindbrain (Avantaggiato et al. 1996, Pourquié 2003, White et al. 2000, Zhang et al. 1996). Thus, the precise regulation of RA function during development is fundamental for the proper establishment of anteroposterior (AP) polarity in the vertebrate CNS (Avantaggiato et al. 1996, Durston et al. 1989, White et al. 2000).

The activity of RA in the hindbrain is regulated by dynamic expression of RAR and RXR (i.e., of the RA receptors), of RALDH2 (i.e., the RA source), and of the CYP26s (i.e., the RA sinks) (Hernandez et al. 2007, Sirbu et al. 2005, White and Schilling 2008). In particular, the expression of RAR paralogs and the activity of different CYP26 enzymes regulate RA signaling in the hindbrain leading to rhombomere-specific activation of *hox* genes (Rijli et al. 1998, White and Schilling 2008).

The AP regionalization of the hindbrain is thus dependent on RA signaling, and this function of RA is directly mediated by *hox* genes (Glover et al. 2006, Rijli et al. 1998): within the hindbrain, the direct regulation by RA of the overlapping expression of different *hox* genes, the so-called *hox* code, ensures the proper patterning of this CNS region (Marshall et al. 1996, Rijli et al. 1998). Interestingly, in addition to its role in transcriptional activation of *hox* genes, RAR and RXR may also promote *hox* gene expression through interactions with chromatin-remodeling enzymes (del Corral and Storey 2004).

The anterior limit of *hox* expression in the CNS is at least in part defined by diffusible signals emanating from the midbrain-hindbrain boundary region, such as FGF8, working antagonistically to the RA signal (Irving and Mason 2000). RA/FGF antagonism seems to be a recurrent scheme for the regulation of *hox* genes, given

that such a crosstalk has also been described for the regulation of *hox* genes along the AP axis of the vertebrate CNS (White and Schilling 2008). RA thus preferentially activates anteriorly expressed *hox* genes, while FGF induces progressively more posteriorly expressed *hox* genes (Irving and Mason 2000, Liu et al. 2001).

In addition to AP regionalization of the CNS, RA signaling also controls neuronal differentiation and specification and is also involved in the patterning of the neural tube along the dorsoventral (DV) axis (Maden 2002). In the spinal cord, RA interacts with SHH to establish ventral determinants that later will give rise to motor neurons. This specification consists of the induction of genes, such as *pax6*, *nkx6.1,* or *olig2*, that are responsible for mediating DV patterning of the spinal cord (Novitch et al. 2003). When RA signaling is not present, the ventral neurons are not induced to differentiate from the spinal cord neuroectodermal progenitor cells (Molotkova et al. 2005, Novitch et al. 2003). Finally, neurite outgrowth during development is also regulated by RA, since lack of RA signaling results in abnormal axonal projections (Corcoran et al. 2000, White et al. 2000).

1.5.2 RA Signaling in Neural Crest Cells (NCCs) and Placode Development

NCCs are a transient population of cells in vertebrates that originate at the dorsal neural tube from where they delaminate and subsequently migrate through the embryo (Krispin et al. 2010, Martínez-Morales et al. 2011). NCCs ultimately differentiate into a wide variety of cell types, such as sensory neurons, Schwann cells, myoblasts, chondrocytes, osteocytes, or melanocytes, among others (Hall 2009). Migrating NCCs are already specified to a particular cell lineage according to their place of origin along the AP axis of the neural tube and hence actively participate in the regional patterning of the embryo (Minoux and Rijli 2010, Trainor and Krumlauf 2001). It has been shown that exogenous RA triggers abnormal migration of NCCs from rhombomere 4 into the first pharyngeal arch, as opposed to their normal migration to the second pharyngeal arch (Minoux and Rijli 2010, Plant et al. 2000). RA signaling has also been implicated in controlling differentiation and survival of NCCs at different developmental stages (Begemann et al. 2001, Martínez-Morales et al. 2011). One of the most important processes NCCs need to undergo to become migratory is an epithelial-mesenchymal transition, a process that is regulated by interaction of RA and FGF signaling (Martínez-Morales et al. 2011). Indeed, alterations of RA or FGF signaling activity disrupt the timing of epithelial-mesenchymal transitions of NCCs along the AP axis of the CNS, which ultimately leads to changes in the cell fates of the NCCs (Krispin et al. 2010).

Placodes are discrete, specialized ectodermal areas located at stereotyped positions in the embryonic head (Schlosser 2010). They give rise to a wide range of cell types, including sensory neurons that contribute to cranial ganglia (Schlosser 2010). RA signaling has been implicated in the development of the otic (Dupé et al. 1999), lens (Matt et al. 2005), olfactory (Song et al. 2004), and lateral line placodes (Gibbs and Northcutt 2004), but the precise functions of RA in the different placodes remain elusive. Eye morphogenesis and retina differentiation are also controlled by

RA, although the RA-dependent signal seems to act primarily on the NCC-derived periocular mesenchyme (Matt et al. 2005, McCaffery et al. 1999).

1.5.3 RA Signaling and Left/Right (LR) Axial Patterning

In addition to patterning along the AP and DV body axes, RA signaling is also implicated in the early determination of LR asymmetry during vertebrate development. Indeed, RA seems to be involved in the side-specific activation of genes involved in LR specification, such as *lefty*, *pitx*, and *nodal* (Wasiak and Lohnes 1999). It has initially been suggested that the RA-dependent establishment of LR asymmetry is controlled by vesicles loaded with RA and SHH that are delivered to the future left side of the embryo by rotating cilia in an FGF-dependent process (Tanaka et al. 2005). This hypothesis was subsequently challenged by the finding that an antagonism between FGF8 and RA in the ectoderm is required for generating symmetry. Failure of this antagonism to be established generates excessive FGF8 signaling in adjacent mesoderm that leads to reduction of somite size and LR asymmetry (Sirbu and Duester 2006). In addition, it has been shown that RA action during somitogenesis is associated with the maintenance of symmetric somite formation along the LR axis (Vermot and Pourquié 2005). RA thus works as a buffer to stabilize and balance the LR signaling cues in the presomitic mesoderm (PSM) (Vermot and Pourquié 2005).

1.5.4 RA Signaling in Somitogenesis

Somites are metameric structures of mesodermal origin that give rise to vertebrae and their associated muscles and tendons, the skeletal muscles of the body wall and limbs, and to the dermis of the back (Brent and Tabin 2002, Dubrulle and Pourquié 2004). The biological process of forming new somites appears to be highly conserved during vertebrate embryogenesis and is controlled by rhythmic gene transcription within the PSM, involving mainly NOTCH and WNT signaling components that establish the so-called segmentation clock (Pourquié 2003). In addition to its role in the maintenance of LR asymmetry, RA produced by RALDH2 in anterior somites is involved in repressing FGF8 signals emanating from the posterior ectoderm (Duester 2007). In this way, reciprocal inhibition of FGF8 signals by RA and of RA signals by FGF8 is crucial for positioning segment boundaries in the PSM (Moreno and Kintner 2004). This interaction is responsible for the translation of pulsations of the segmentation clock into the periodic establishment of segment boundaries, with RA canceling the effects of FGF8 and hence triggering the expression of segmentation genes responsible for somite formation (Duester 2007, Moreno and Kintner 2004). Furthermore, through the control of *hox* gene expression, the RA/FGF antagonism also has a very important function in defining AP positional information in structures derived from the PSM (del Corral and Storey 2004).

1.5.5 RA in Heart Field Specification

RA signaling also plays important roles in patterning the developing heart, in particular during cardiac field specification, during AP regionalization of the heart tube,

and in the establishment of LR asymmetry during heart looping (Chazaud et al. 1999, Niederreither et al. 1999, Rosenthal and Xavier-Neto 2000, Xavier-Neto et al. 2001). Interestingly, at least the role of RA in AP patterning of the heart is well conserved within vertebrates, given that exogenous RA leads to similar heart regionalization defects in mice (Lin et al. 2010), chicken (Hochgreb et al. 2003), and zebrafish (Stainier and Fishman 1992). In vertebrates, RALDH2 is expressed early in the heart mesoderm in an AP wave, hence defining cardiac precursors and delimitating the heart field along the AP axis (Hochgreb et al. 2003). RA signaling exerts its effect on AP regionalization of the heart, at least in part, by downregulating *isl1* expression, probably by reducing FGF signaling levels (Sirbu et al. 2008).

1.5.6 RA Signaling and Kidney Development

The kidney is another target tissue of RA signaling during development. The kidney field is derived from intermediate mesoderm, and recent studies have shown that RA signaling is required for early specification of this kidney field (Cartry et al. 2006, Serluca and Fishman 2001, Wingert et al. 2007). In the frog *Xenopus laevis*, for example, early determinants of a pronephric fate, such as *pax8* and *lhx1*, are controlled by RA signaling (Cartry et al. 2006). In zebrafish, RA is required for the positioning and segmentation of the pronephric kidney (Wingert et al. 2007). Moreover, the frog *pteg* and zebrafish *wt1* genes, both of which play important roles during kidney development in the respective animal, are direct targets of RA signaling (Bollig et al. 2009, Lee et al. 2010). In mice, it has further been shown that RA signaling, by controlling *ret* expression in ureteric bud cells, is crucial for ureteric bud formation and for branching morphogenesis within the developing collecting duct system (Rosselot et al. 2010). Interestingly, RA signaling activity in ureteric bud cells depends mainly on RA generated by RALDH2 in adjacent stromal mesenchyme (Rosselot et al. 2010).

1.5.7 Relevance of RA Signaling for Body Appendage Development

Based on pharmacological experiments that had shown, for example, that exogenous RA induces mirror-image duplications in chick wing buds (Tickle et al. 1982), it had initially been suggested that RA signaling plays crucial roles in limb bud patterning of tetrapod vertebrates (Campo-Paysaa et al. 2008). However, more recent work suggests that RA signaling might be less important for tetrapod limb development than initially expected. Albeit required during forelimb induction, RA might not be necessary at all for hind limb budding and patterning (Zhao et al. 2009). Moreover, the involvement of RA in forelimb formation seems to be indirect, by creating a permissive environment for FGF signaling to develop a limb bud (Zhao et al. 2009). Conversely, in zebrafish, RA is indispensable for pectoral fin induction (Gibert et al. 2006, Grandel and Brand 2011), which is suggestive of a divergence in RA signaling functions between fish fin buds and tetrapod limb buds (Campo-Paysaa et al. 2008). Later in development, RA signaling likely controls tissue remodeling and apoptosis at the digit-interdigit junction, with RALDH2 being

expressed in interdigital mesenchyme and CYP26B1 in digits, hence restricting RA action to the interdigital zones (Zhao et al. 2010).

1.5.8 Endoderm Specification and RA

RA signaling is also essential for patterning the developing endoderm. RA is required for the determination of endodermal fields along the AP axis, each of which will give rise to a particular organ (Bayha et al. 2009). In particular, exposure to exogenous RA prevents the expression of genes normally present in the anterior-most endoderm and, at the same time, activates and shifts anteriorly the expression of genes normally expressed at more posterior endoderm levels (Bayha et al. 2009).

In the anterior endoderm, RA signaling is required for the regional patterning of the pharyngeal arches. The pharyngeal arches are structures that develop along the embryonic foregut endoderm in a very complex process that involves the interaction between cells from all germ layers (ectoderm, mesoderm, endoderm, and the neuroectoderm-derived NCCs). RA is produced locally in the pharyngeal arches by RALDH2 in the mesoderm adjacent to the endoderm (Bayha et al. 2009, Niederreither et al. 2003). When RA signaling is disrupted, the first and second pharyngeal arches develop normally, while severe defects are observed in structures derived from more posterior pharyngeal arches (Niederreither et al. 2003). Interestingly, RA seems dispensable for the specification of pharyngeal endoderm, but required for morphogenesis and segmentation of the posterior pharyngeal arches (Kopinke et al. 2006).

Some organs derived from endoderm in the posterior foregut also require RA for proper development (Wang et al. 2006). These organs include the lungs, stomach, liver, and pancreas (Wang et al. 2006). In the lung, for example, loss of RA signaling causes a disruption of lung bud outgrowth (Wang et al. 2006). This defect is due to a reduction of FGF10 levels, which, together with BMP4, is required for mediating proper branching morphogenesis (Wang et al. 2006). Activation of FGF signaling by RA is also required for early stomach and liver development (Wang et al. 2006), while, during pancreas development, mesodermal RALDH2 is required for specification of the dorsal, but not the ventral, pancreatic bud (Molotkov et al. 2005, Stafford et al. 2004).

1.6 EVOLUTION OF RA SIGNALING

Retinoid signaling was long thought to be vertebrate specific, but recent studies have revealed major roles for RA signaling that are conserved in all chordates (Figure 1.4), which, in addition to the vertebrates, include two invertebrate groups: the cephalochordates and the urochordates (also called tunicates) (Campo-Paysaa et al. 2008, Cañestro et al. 2006, Marletaz et al. 2006). While the tunicates constitute the sister group of vertebrates, the cephalochordates are located at the base of the chordate phylum (Schubert et al. 2006). Nonetheless, of all invertebrates, the cephalochordates, such as the lancelet or amphioxus (for example *Branchiostoma floridae,* the Florida amphioxus), are characterized by the most vertebrate-like retinoid signaling system, both in terms of molecular composition and biological functions (Campo-Paysaa et al. 2008, Koop and Holland 2008, Schubert et al. 2006, Theodosiou et al. 2010).

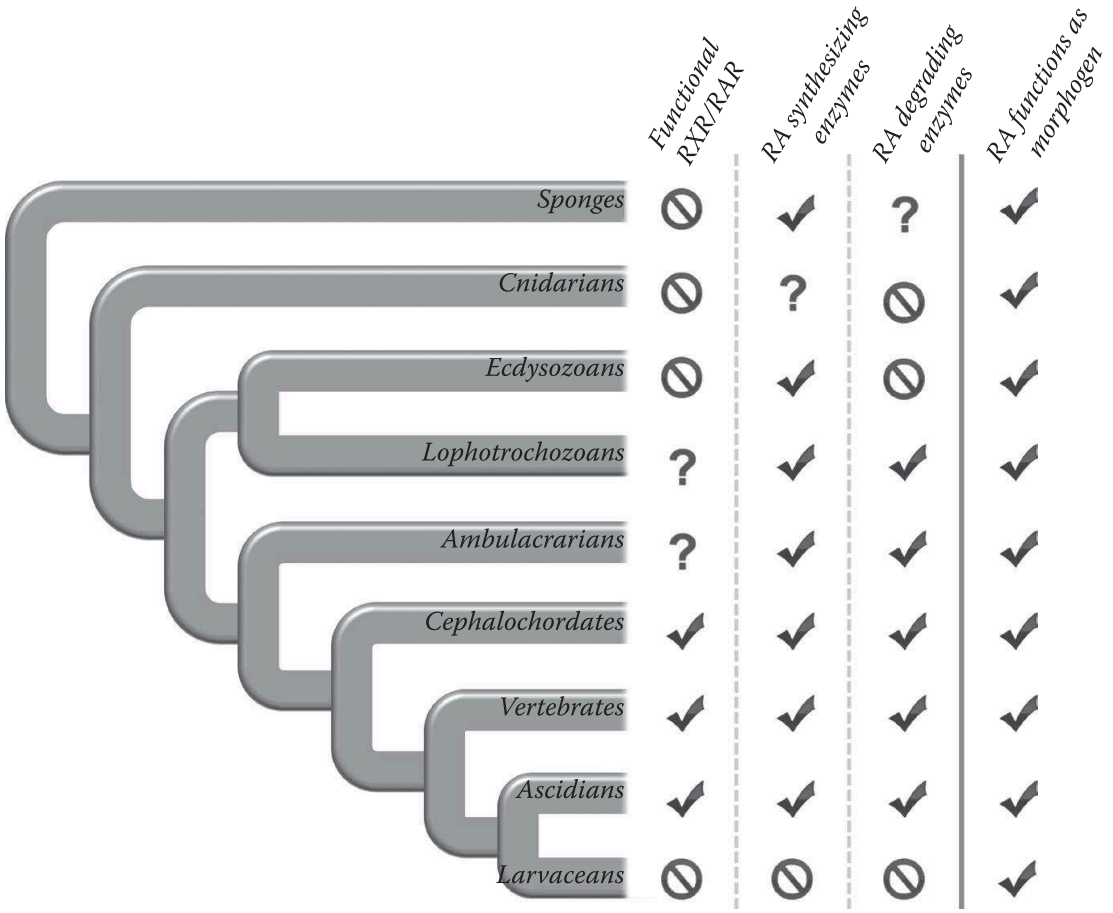


FIGURE 1.4 Evolution of RA signaling components and functions. A simplified phylogenetic tree of metazoan animals is shown and the presence (✓) or absence (⊘) of functional RXR/RAR heterodimers, of RA synthesizing (RALDH), and of RA degrading (CYP26) enzymes is highlighted. Uncertainty over presence or absence is illustrated by (?). The phylogenetic tree also indicates the distribution of intercellular signaling functions of RA (as morphogen) in the different lineages. RA, retinoic acid; RAR, retinoic acid receptor; RXR, retinoid X receptor.

1.6.1 Chordate RA Signaling

When considering the key players of RA metabolism, some conserved and divergent features are observable when comparing vertebrates with invertebrate chordates (Figure 1.4). For example, while vertebrate genomes encode multiple RAR and RXR paralogs, amphioxus has only one RAR and one RXR, forming a single heterodimer (Escriva et al. 2002). Like its vertebrate counterparts, this amphioxus RXR/RAR heterodimer binds to and is activated by RA (Escriva et al. 2002). Using ligands specific for each of the three mammalian RAR paralogs (RARα, RARβ, and RARγ), it was shown that the ligand-binding capacities of amphioxus RAR resemble those of mammalian RARβ, suggesting that the ancestral RAR of vertebrates was of the RARβ type (Escriva et al. 2006). Comparison of developmental gene expression patterns between amphioxus RAR and the three vertebrate RAR paralogs further supported this notion and suggested that, while RARβ kept the ancestral features in terms of ligand specificity and development expression, after duplication, RARα and RARγ might have acquired novel functions in the course of vertebrate diversification (Escriva et al. 2006).

Although the amphioxus genome encodes orthologs of the enzymes required in vertebrates for RA synthesis (RALDH) and degradation (CYP26), it remains to be determined whether these cephalochordate RALDH and CYP26 are indeed capable of, respectively, RA synthesis and degradation (Albalat et al. 2011, Sobreira et al. 2011). Regarding the endogenous production of retinaldehyde from retinol, a bona fide ADH has previously been characterized biochemically in amphioxus (Cañestro et al. 2000, Cañestro et al. 2003, Dalfó et al. 2007), but the functional capacities of the RDH orthologs that have been identified in cephalochordate genomes still need to be assessed (Albalat et al. 2011, Holland et al. 2008). Interestingly, although a candidate for an intracellular retinol-binding protein (CRBP) has previously been described in amphioxus, there is no convincing evidence for the existence of an amphioxus intracellular RA-binding protein (CRABP) (Albalat et al. 2011, Holland et al. 2008, Jackman et al. 2004), and the concerted absence in amphioxus of the main components for retinol storage, transport, and cellular uptake further suggest that these systems might be functional innovations of vertebrates (Albalat et al. 2011).

Several studies have assessed the roles of RA signaling during amphioxus development (Campo-Paysaa et al. 2008, Escriva et al. 2002, Holland and Holland 1996, Koop and Holland 2008). For example, in the amphioxus CNS, it was shown that RA regulates both AP regional patterning and neuronal specification in a *hox*-dependent manner (Schubert et al. 2006). Likewise, the amphioxus ectoderm is also patterned by RA: disruption of endogenous RA signaling levels lead to alteration of an ectodermal *hox* code, which in turn results in AP regionalization defects and changes in the combinatorial code defining specific populations of ectodermal sensory neurons (Schubert et al. 2004). Recently, it has also been shown that, at later developmental stages, RA signaling is involved in the formation of the amphioxus tail fin, which arises from the posterior ectoderm (Koop et al. 2011).

In the endoderm, RA signaling via *hox1* has been shown to be required for the definition of the posterior limit of the amphioxus pharynx by limiting expression of pro-pharyngeal genes to the anteriormost (thus pharyngeal) endoderm (Schubert et al. 2005). Moreover, in the posterior endoderm, RA signaling controls the expression of two *parahox* genes, *xlox* and *cdx*, hence mediating AP patterning of the developing amphioxus hindgut (Osborne et al. 2009). While it is very likely that RA signaling directly activates expression of amphioxus *hox* genes (Koop et al. 2010), further work is needed to assess in more detail the regulation of amphioxus *parahox* genes by RA (Osborne et al. 2009).

Contrasting the situation in amphioxus, RA signaling in tunicates has secondarily been modified in different lineages (Campo-Paysaa et al. 2008, Cañestro et al. 2006, Koop and Holland 2008). For example, although the genome of the ascidian tunicate *Ciona intestinalis* contains the basic molecular components of the RA signaling cascade, that is a single receptor heterodimer (RXR/RAR) and enzymes potentially capable of RA synthesis (RALDH) and degradation (CYP26), the regulation of at least some of these components in this tunicate has been modified (Campo-Paysaa et al. 2008, Cañestro et al. 2007, Koop and Holland 2008). Transcription of RAR in this ascidian tunicate is not regulated by RA, contrasting with amphioxus and vertebrates, where RAR expression is directly regulated by RA (Nagatomo et al. 2003). Moreover, the larvacean tunicate *Oikopleura dioica* has lost most of the molecular

components defining the canonical RA pathway, including RAR, RALDH, and CYP26 (Campo-Paysaa et al. 2008, Cañestro and Postlethwait 2007).

Along these lines, the roles for RA signaling during tunicate development seem to be limited. For example, although RA signaling might function during development of the ascidian tunicate CNS (De Bernardi et al. 1994, Nagatomo et al. 2003), RA probably only has a limited role in the regulation of *hox* gene expression (Nagatomo and Fujiwara 2003, Natale et al. 2011, Wada et al. 2006). Moreover, given that tunicates in general have lost several *hox* genes and that both cluster organization and collinear expression have at least partially been lost (Ikuta et al. 2004, Seo et al. 2004), the contribution of an RA-regulated *hox* code to tunicate CNS development is very limited (Ikuta et al. 2010).

Interestingly, exogenous RA has been shown to activate and shift anteriorly the expression of ascidian *hox1* in both the CNS and ectoderm (Kanda et al. 2009, Katsuyama et al. 1995, Katsuyama and Saiga 1998, Nagatomo and Fujiwara 2003) and, at least in the ectoderm, this activation is directly mediated by the ascidian RXR/RAR heterodimer, which represents a very rare example for RA-dependent gene expression in ascidians (Kanda et al. 2009). This RA-*hox1* network in the ectoderm is required for formation of the ascidian atrial siphon placode, which is considered homologous to the vertebrate otic placode (Sasakura et al. 2012). The *hox1* gene is also expressed in the ascidian endoderm (Ikuta et al. 2004), where exogenous RA leads to a loss of expression of pharyngeal markers and to a posteriorization of the endoderm leading to a complete loss of the branchial basket (Hinman and Degnan 1998, 2000; Ogasawara et al. 1999). Intriguingly, these effects of RA are reminiscent of the phenotypes observed in amphioxus and vertebrates following treatment with exogenous RA (Campo-Paysaa et al. 2008).

Taken together, at the base of the chordate lineage the basic molecular components constituting the RA signaling cascade were most likely already present. Moreover, the RA pathway was probably involved in the specification of neuronal cell populations in both CNS and ectoderm, and RA signals also contributed to the regional patterning along the AP axis of the CNS, ectoderm, and endoderm. In all three tissue layers, this RA-dependent AP regionalization process was probably directly mediated by *hox* genes (Campo-Paysaa et al. 2008, Marletaz et al. 2006). In the tunicate lineage, the functions of the RA signaling cascade have been secondarily modified leading to diverged functions, such as the partial loss of RA-dependent regulation of the *hox* code in ascidians, and culminating in the loss of key components of the RA pathway, including RAR, RALDH, and CYP26 in larvaceans (Cañestro et al. 2007).

1.6.2 Non-Chordate RA Signaling

Outside the chordate lineage, evidence for functional roles of retinoids is scarce (Figure 1.4). Although genes encoding orthologs of vertebrate RAR and RXR as well as orthologs of the basic vertebrate components for endogenous RA synthesis (RALDH) and degradation (CYP26) are present in ambulacrarians (such as hemichordates and echinoderms) and lophotrochozoans (such as annelids and mollusks) (Albalat and Cañestro 2009, Campo-Paysaa et al. 2008, Theodosiou et al. 2010), extensive experimental evidence for retinoid and/or RXR/RAR heterodimer

functions in these lineages is unfortunately still lacking. It has been suggested that RA treatments result in a delay of embryonic development in echinoderms, and, in mollusks, retinoids (in particular 9-*cis* RA) disrupt embryogenesis and are also involved in neuronal differentiation, neuron outgrowth, and growth cone guidance (Farrar et al. 2009). This role of retinoids in mollusk neurogenesis is apparently independent of a functional RXR/RAR heterodimer, but instead requires a cytoplasmic localization of RXR (Carter et al. 2010). Also, in mollusks, exogenous RA affects eye formation and leads to arrested development at the trochophore larval stage (Créton et al. 1993).

Intriguingly, the available *in silico* data suggest that ecdysozoans (which include insects, crustaceans, and nematodes) might have secondarily lost both the canonical retinoid receptor RAR and a CYP26-based RA degrading machinery (Albalat and Cañestro 2009, Campo-Paysaa et al. 2008, Theodosiou et al. 2010), although retinoid functions have been described in tissue regeneration of both insects and crustaceans (Halme et al. 2010, Hopkins 2001). Other functions of retinoids in animals, whose genomes do, *a priori*, not encode the RAR receptor include, for example, hydroid specification, cell proliferation, and neuronal differentiation in cnidarians (Estephane and Anctil 2010, Müller 1984), and tissue regression and spicule formation in sponges (Müller et al. 2011, Wiens et al. 2003). Collectively, these data suggest that in the course of metazoan evolution the functions for retinoids might have originated before a functional RXR/RAR heterodimer and that the presence of both RAR and RXR in a genome does not automatically imply a functional RXR/RAR heterodimer capable of binding and being activated by retinoids.

1.7 CONCLUSION

Considerable time and effort have been invested by the scientific community in the analysis of metabolism and functions of RA during development. It is clear that the signaling cascade controlled by this morphogen intervenes in various tissues at different time points during embryogenesis. Because of this multitude of biological functions, we are still lacking a comprehensive understanding of the developmental roles played by the RA pathway. Moreover, we are only starting to understand the complex regulation during development of the RA signaling activity, which depends not only on a very fine regulation of endogenous RA availability, but also on the complex expression of receptors and cofactors, all of which are controlled by intricate interactions of RA signaling with other signaling cascades creating regulatory loops of various sorts.

Further work is thus needed to unite the various facets of the RA signaling network from metabolism to signal transduction and from interactions with other signaling cascades to developmental functions into a coherent ensemble representing the regulatory network as a whole. In this context, model organisms with representative, vertebrate-like RA signaling cascades that lack genetic redundancy and secondary modifications, such as the cephalochordate amphioxus, will be an important asset for providing the blueprint of an ancestral RA network that was elaborated in vertebrates following the genomic expansion by two rounds of whole genome duplications (Van de Peer et al. 2009).

The usefulness of expanding the scope of RA-centered research to non-vertebrate models is not only limited to instructing the complexity of vertebrate systems. As a matter of fact, data from various organisms might have the potential to reveal the evolutionary origins of the RA signaling network and the ancestral functions that retinoids might have assumed during development of the first metazoan animals. The results that have already been obtained in several invertebrate taxa suggest that retinoids might not necessarily require the canonical retinoid receptors RAR and RXR to elicit a biological function and that the presence in a genome of RAR and RXR does thus not automatically imply that the heterodimer formed by these receptors is reactive to retinoids. In this context, it is interesting to note that retinoids have very ancient roles in light reception and in the visual system, which represents a pool of retinoids that, in the course of evolution, might have been co-opted for roles in morphogen-dependent signaling and that has ultimately been elaborated into one of the most important intercellular signaling cascades in development.

ACKNOWLEDGEMENTS

The authors are indebted to Ricardo Lara-Ramírez, Florent Campo-Paysaa, Juliana Gutierrez-Mazariegos, and Eric Samarut for critical reading of the manuscript, and to Vincent Laudet for having hosted our group in his laboratory. This work was supported by funds from ANR (ANR-09-BLAN-0262-02 and ANR-11-JSV2-002-01) and CNRS to Michael Schubert.

REFERENCES

Abu-Abed, S., P. Dollé, D. Metzger, et al. 2001. The Retinoic Acid-Metabolizing Enzyme, Cyp26a1, Is Essential for Normal Hindbrain Patterning, Vertebral Identity, and Development of Posterior Structures. *Genes Dev* 15, no 2: 226–40.

Aggarwal, S., S. W. Kim, K. Cheon, et al. 2006. Nonclassical Action of Retinoic Acid on the Activation of the Camp Response Element-Binding Protein in Normal Human Bronchial Epithelial Cells. *Mol Biol Cell* 17, no 2: 566–75.

Albalat, R., F. Brunet, V. Laudet, and M. Schubert. 2011. Evolution of Retinoid and Steroid Signaling: Vertebrate Diversification from an Amphioxus Perspective. *Genome Biol Evol* 3: 985–1005.

Albalat, R., and C. Cañestro. 2009. Identification of Aldh1a, Cyp26 and Rar Orthologs in Protostomes Pushes Back the Retinoic Acid Genetic Machinery in Evolutionary Time to the Bilaterian Ancestor. *Chem-Biol Interact* 178, no 1-3: 188–96.

Avantaggiato, V., D. Acampora, F. Tuorto, and A. Simeone. 1996. Retinoic Acid Induces Stage-Specific Repatterning of the Rostral Central Nervous System. *Dev Biol* 175, no 2: 347–57.

Balmer, J. E., and R. Blomhoff. 2005. A Robust Characterization of Retinoic Acid Response Elements Based on a Comparison of Sites in Three Species. *J Steroid Biochem* 96, no 5: 347–54.

Barron, M., D. McAllister, S. M. Smith, and J. Lough. 1998. Expression of Retinol Binding Protein and Transthyretin During Early Embryogenesis. *Dev Dynam* 212, no 3: 413–22.

Bayha, E., M. C. Jørgensen, P. Serup, and A. Grapin-Botton. 2009. Retinoic Acid Signaling Organizes Endodermal Organ Specification Along the Entire Antero-Posterior Axis. *Plos One* 4, no 6: e5845.

Begemann, G., T. F. Schilling, G. J. Rauch, R. Geisler, and P. W. Ingham. 2001. The Zebrafish Neckless Mutation Reveals a Requirement for Raldh2 in Mesodermal Signals That Pattern the Hindbrain. *Development* 128, no 16: 3081–94.

Bempong, D. K., I. L. Honigberg, and N. M. Meltzer. 1995. Normal-Phase Lc-Ms Determination of Retinoic Acid Degradation Products. *J Pharmaceut Biomed* 13, no 3: 285–91.

Berry, D. C., and N. Noy. 2009. All-*trans*-Retinoic Acid Represses Obesity and Insulin Resistance by Activating Both Peroxisome Proliferation-Activated Receptor β/δ and Retinoic Acid Receptor. *Mol Cell Biol* 29, no 12: 3286–96.

Berry, D. C., and N. Noy. 2012. Signaling by Vitamin A and Retinol-Binding Protein in Regulation of Insulin Responses and Lipid Homeostasis. *Biochim Biophys Acta* 1821, no 1: 168–76.

Blomhoff, R., and H. K. Blomhoff. 2006. Overview of Retinoid Metabolism and Function. *J Neurobiol* 66, no 7: 606–30.

Bollig, F., B. Perner, B. Besenbeck, et al. 2009. A Highly Conserved Retinoic Acid Responsive Element Controls *Wt1a* Expression in the Zebrafish Pronephros. *Development* 136, no 17: 2883–92.

Brent, A. E., and C. J. Tabin. 2002. Developmental Regulation of Somite Derivatives: Muscle, Cartilage and Tendon. *Curr Opin Genet Dev* 12, no 5: 548–57.

Cai, A. Q., K. Radtke, A. Linville, et al. 2012. Cellular Retinoic Acid-Binding Proteins Are Essential for Hindbrain Patterning and Signal Robustness in Zebrafish. *Development* 139, no 12: 2150–55.

Campo-Paysaa, F., F. Marlétaz, V. Laudet, and M. Schubert. 2008. Retinoic Acid Signaling in Development: Tissue-Specific Functions and Evolutionary Origins. *Genesis* 46, no 11: 640–56.

Cañestro, C., L. Godoy, R. Gonzàlez-Duarte, and R. Albalat. 2003. Comparative Expression Analysis of *Adh3* During Arthropod, Urochordate, Cephalochordate, and Vertebrate Development Challenges Its Predicted Housekeeping Role. *Evol Dev* 5, no 2: 157–62.

Cañestro, C., L. Hjelmqvist, R. Albalat, et al. 2000. Amphioxus Alcohol Dehydrogenase Is a Class 3 Form of Single Type and of Structural Conservation but with Unique Developmental Expression. *Eur J Biochem* 267, no 22: 6511–18.

Cañestro, C., and J. H. Postlethwait. 2007. Development of a Chordate Anterior-Posterior Axis without Classical Retinoic Acid Signaling. *Dev Biol* 305, no 2: 522–38.

Cañestro, C., J. H. Postlethwait, R. Gonzàlez-Duarte, and R. Albalat. 2006. Is Retinoic Acid Genetic Machinery a Chordate Innovation? *Evol Dev* 8, no 5: 394–406.

Cañestro, C., H. Yokoi, and J. H. Postlethwait. 2007. Evolutionary Developmental Biology and Genomics. *Nat Rev Genet* 8, no 12: 932–42.

Carter, C. J., N. Farrar, R. L. Carlone, and G. E. Spencer. 2010. Developmental Expression of a Molluscan Rxr and Evidence for Its Novel, Nongenomic Role in Growth Cone Guidance. *Dev Biol* 343, no 1-2: 124–37.

Cartry, J., M. Nichane, V. Ribes, et al. 2006. Retinoic Acid Signalling Is Required for Specification of Pronephric Cell Fate. *Dev Biol* 299, no 1: 35–51.

Chambers, D., L. Wilson, M. Maden, and A. Lumsden. 2007. Raldh-Independent Generation of Retinoic Acid During Vertebrate Embryogenesis by Cyp1b1. *Development* 134, no 7: 1369–83.

Chambon, P. 1996. A Decade of Molecular Biology of Retinoic Acid Receptors. *Faseb J* 10, no 9: 940–54.

Chazaud, C., P. Chambon, and P. Dollé. 1999. Retinoic Acid Is Required in the Mouse Embryo for Left-Right Asymmetry Determination and Heart Morphogenesis. *Development* 126, no 12: 2589–96.

Chen, Y. L., N. Pollet, C. Niehrs, and T. Pieler. 2001. Increased Xraldh2 Activity Has a Posteriorizing Effect on the Central Nervous System of *Xenopus* Embryos. *Mech Develop* 101, no 1-2: 91–103.

Collins, M. D., and G. E. Mao. 1999. Teratology of Retinoids. *Annu Rev Pharmacol* 39: 399–430.

Corcoran, J., B. Shroot, J. Pizzey, and M. Maden. 2000. The Role of Retinoic Acid Receptors in Neurite Outgrowth from Different Populations of Embryonic Mouse Dorsal Root Ganglia. *J Cell Sci* 113: 2567–74.

Créton, R., G. Zwaan, and R. Dohmen. 1993. Specific Developmental Defects in Mollusks after Treatment with Retinoic Acid During Gastrulation. *Dev Growth Differ* 35, no 3: 357–64.

Dalfó, D., N. Marqués, and R. Albalat. 2007. Analysis of the Nadh-Dependent Retinaldehyde Reductase Activity of Amphioxus Retinol Dehydrogenase Enzymes Enhances Our Understanding of the Evolution of the Retinol Dehydrogenase Family. *Febs J* 274, no 14: 3739–52.

D'Ambrosio, D. N., R. D. Clugston, and W. S. Blaner. 2011. Vitamin A Metabolism: An Update. *Nutrients* 3, no 1: 63–103.

De Bernardi, F., C. Sotgia, and G. Ortolani. 1994. Retinoic Acid Treatment of Ascidian Embryos: Effects on Larvae and Metamorphosis. *Anim Dev* 3: 75–81.

del Corral, R. D., and K. G. Storey. 2004. Opposing Fgf and Retinoid Pathways: A Signalling Switch That Controls Differentiation and Patterning Onset in the Extending Vertebrate Body Axis. *Bioessays* 26, no 8: 857–69.

Dong, D., S. E. Ruuska, D. J. Levinthal, and N. Noy. 1999. Distinct Roles for Cellular Retinoic Acid-Binding Proteins I and Ii in Regulating Signaling by Retinoic Acid. *J Biol Chem* 274, no 34: 23695–98.

Dubrulle, J., and O. Pourquié. 2004. Coupling Segmentation to Axis Formation. *Development* 131, no 23: 5783–93.

Duester, G. 2007. Retinoic Acid Regulation of the Somitogenesis Clock. *Birth Defects Res C Embryo Today* 81, no 2: 84–92.

Duester, G. 2008. Retinoic Acid Synthesis and Signaling During Early Organogenesis. *Cell* 134, no 6: 921–31.

Duester, G., F. A. Mic, and A. Molotkov. 2003. Cytosolic Retinoid Dehydrogenases Govern Ubiquitous Metabolism of Retinol to Retinaldehyde Followed by Tissue-Specific Metabolism to Retinoic Acid. *Chem-Biol Interact* 143: 201–10.

Duong, V., and C. Rochette-Egly. 2011. The Molecular Physiology of Nuclear Retinoic Acid Receptors. From Health to Disease. *Biochem Biophys Acta* 1812, no 8: 1023–31.

Dupé, V., N. B. Ghyselinck, O. Wendling, P. Chambon, and M. Mark. 1999. Key Roles of Retinoic Acid Receptors Alpha and Beta in the Patterning of the Caudal Hindbrain, Pharyngeal Arches and Otocyst in the Mouse. *Development* 126, no 22: 5051–59.

Durston, A. J., J. P. M. Timmermans, W. J. Hage, et al. 1989. Retinoic Acid Causes an Anteroposterior Transformation in the Developing Central Nervous System. *Nature* 340, no 6229: 140–44.

Engberg, N., M. Kahn, D. R. Petersen, M. Hansson, and P. Serup. 2010. Retinoic Acid Synthesis Promotes Development of Neural Progenitors from Mouse Embryonic Stem Cells by Suppressing Endogenous, Wnt-Dependent Nodal Signaling. *Stem Cells* 28, no 9: 1498–509.

Escriva, H., S. Bertrand, P. Germain, et al. 2006. Neofunctionalization in Vertebrates: The Example of Retinoic Acid Receptors. *Plos Genet* 2, no 7: 955–65.

Escriva, H., N. D. Holland, H. Gronemeyer, C. Laudet, and L. Z. Holland. 2002. The Retinoic Acid Signaling Pathway Regulates Anterior/Posterior Patterning in the Nerve Cord and Pharynx of Amphioxus, a Chordate Lacking Neural Crest. *Development* 129, no 12: 2905–16.

Estephane, D., and M. Anctil. 2010. Retinoic Acid and Nitric Oxide Promote Cell Proliferation and Differentially Induce Neuronal Differentiation *in vitro* in the Cnidarian *Renilla koellikeri*. *Dev Neurobiol* 70, no 12: 842–52.

Farjo, K. M., G. Moiseyev, O. Nikolaeva, et al. 2011. Rdh10 Is the Primary Enzyme Responsible for the First Step of Embryonic Vitamin A Metabolism and Retinoic Acid Synthesis. *Dev Biol* 357, no 2: 347–55.

Farrar, N. R., J. M. Dmetrichuk, R. L. Carlone, and G. E. Spencer. 2009. A Novel, Nongenomic Mechanism Underlies Retinoic Acid-Induced Growth Cone Turning. *J Neurosci* 29, no 45: 14136–42.

Feng, L., R. E. Hernandez, J. S. Waxman, D. Yelon, and C. B. Moens. 2010. Dhrs3a Regulates Retinoic Acid Biosynthesis through a Feedback Inhibition Mechanism. *Dev Biol* 338, no 1: 1–14.

Fraser, P. D., and P. M. Bramley. 2004. The Biosynthesis and Nutritional Uses of Carotenoids. *Prog Lipid Res* 43, no 3: 228–65.

Ghyselinck, N. B., C. Båvik, V. Sapin, et al. 1999. Cellular Retinol-Binding Protein I Is Essential for Vitamin A Homeostasis. *Embo J* 18, no 18: 4903–14.

Gibbs, M. A., and R. G. Northcutt. 2004. Retinoic Acid Repatterns Axolotl Lateral Line Receptors. *Int J Dev Biol* 48, no 1: 63–6.

Gibert, Y., A. Gajewski, A. Meyer, and G. Begemann. 2006. Induction and Prepatterning of the Zebrafish Pectoral Fin Bud Requires Axial Retinoic Acid Signaling. *Development* 133, no 14: 2649–59.

Giguere, V., E. S. Ong, P. Segui, and R. M. Evans. 1987. Identification of a Receptor for the Morphogen Retinoic Acid. *Nature* 330, no 6149: 624–29.

Glover, J. C., J. S. Renaud, and F. M. Rijli. 2006. Retinoic Acid and Hindbrain Patterning. *J Neurobiol* 66, no 7: 705–25.

Grandel, H., and M. Brand. 2011. Zebrafish Limb Development Is Triggered by a Retinoic Acid Signal During Gastrulation. *Dev Dynam* 240, no 5: 1116–26.

Gronemeyer, H., J. A. Gustafsson, and V. Laudet. 2004. Principles for Modulation of the Nuclear Receptor Superfamily. *Nat Rev Drug Discov* 3, no 11: 950–64.

Gutierrez-Mazariegos, J., M. Theodosiou, F. Campo-Paysaa, and M. Schubert. 2011. Vitamin A: A Multifunctional Tool for Development. *Semin Cell Dev Biol* 22, no 6: 603–10.

Hale, F. 1933. Pigs Born Without Eye Balls. *J Heredity* 24, no 3: 105–06.

Hall, B. K. 2009. *The Neural Crest and Neural Crest Cells on Vertebrate Development and Evolution*. 2nd ed. ed: Springer.

Hall, J. A., J. R. Grainger, S. P. Spencer, and Y. Belkaid. 2011. The Role of Retinoic Acid in Tolerance and Immunity. *Immunity* 35, no 1: 13–22.

Halme, A., M. Cheng, and I. K. Hariharan. 2010. Retinoids Regulate a Developmental Checkpoint for Tissue Regeneration in *Drosophila*. *Curr Biol* 20, no 5: 458–63.

Hernandez, R. E., A. P. Putzke, J. P. Myers, L. Margaretha, and C. B. Moens. 2007. Cyp26 Enzymes Generate the Retinoic Acid Response Pattern Necessary for Hindbrain Development. *Development* 134, no 1: 177–87.

Hessel, S., A. Eichinger, A. Isken, et al. 2007. Cmo1 Deficiency Abolishes Vitamin A Production from β-Carotene and Alters Lipid Metabolism in Mice. *J Biol Chem* 282, no 46: 33553–61.

Hinman, V. F., and B. M. Degnan. 1998. Retinoic Acid Disrupts Anterior Ectodermal and Endodermal Development in Ascidian Larvae and Postlarvae. *Dev Genes Evol* 208, no 6: 336–45.

Hinman, V. F., and B. M. Degnan. 2000. Retinoic Acid Perturbs *Otx* Gene Expression in the Ascidian Pharynx. *Dev Genes Evol* 210, no 3: 129–39.

Hochgreb, T., V. L. Linhares, D. C. Menezes, et al. 2003. A Caudorostral Wave of Raldh2 Conveys Anteroposterior Information to the Cardiac Field. *Development* 130, no 22: 5363–74.

Holland, L. Z., R. Albalat, K. Azumi, et al. 2008. The Amphioxus Genome Illuminates Vertebrate Origins and Cephalochordate Biology. *Genome Res* 18, no 7: 1100–11.

Holland, L. Z., and N. D. Holland. 1996. Expression of *Amphihox-1* and *Amphipax-1* in Amphioxus Embryos Treated with Retinoic Acid: Insights into Evolution and Patterning of the Chordate Nerve Cord and Pharynx. *Development* 122, no 6: 1829–38.

Hopkins, P. M. 2001. Limb Regeneration in the Fiddler Crab, *Uca pugilator*: Hormonal and Growth Factor Control. *Am Zool* 41, no 3: 389–98.

Ikuta, T., N. Satoh, and H. Saiga. 2010. Limited Functions of Hox Genes in the Larval Development of the Ascidian *Ciona intestinalis*. *Development* 137, no 9: 1505–13.

Ikuta, T., N. Yoshida, N. Satoh, and H. Saiga. 2004. *Ciona intestinalis* Hox Gene Cluster: Its Dispersed Structure and Residual Colinear Expression in Development. *P Natl Acad Sci USA* 101, no 42: 15118–23.

Irving, C., and I. Mason. 2000. Signalling by Fgf8 from the Isthmus Patterns Anterior Hindbrain and Establishes the Anterior Limit of Hox Gene Expression. *Development* 127, no 1: 177–86.

Jackman, W. R., J. M. Mougey, G. D. Panopoulou, and C. B. Kimmel. 2004. *Crabp* and *Maf* Highlight the Novelty of the Amphioxus Club-Shaped Gland. *Acta Zool-Stockholm* 85, no 2: 91–99.

Kam, R. K., Y. Deng, Y. Chen, and H. Zhao. 2012. Retinoic Acid Synthesis and Functions in Early Embryonic Development. *Cell Biosci* 2, no 1: 11.

Kanda, M., H. Wada, and S. Fujiwara. 2009. Epidermal Expression of *Hox1* Is Directly Activated by Retinoic Acid in the *Ciona intestinalis* Embryo. *Dev Biol* 335, no 2: 454–63.

Kane, M. A. 2012. Analysis, Occurrence, and Function of 9-*cis*-Retinoic Acid. *Biochem Biophys Acta* 1821, no 1: 10–20.

Katsuyama, Y., and H. Saiga. 1998. Retinoic Acid Affects Patterning Along the Anterior-Posterior Axis of the Ascidian Embryo. *Dev Growth Differ* 40, no 4: 413–22.

Katsuyama, Y., S. Wada, S. Yasugi, and H. Saiga. 1995. Expression of the *Labial* Group Hox Gene *Hrhox-1* and Its Alteration Induced by Retinoic Acid in Development of the Ascidian *Halocynthia roretzi*. *Development* 121, no 10: 3197–205.

Kawaguchi, R., J. M. Yu, J. Honda, et al. 2007. A Membrane Receptor for Retinol Binding Protein Mediates Cellular Uptake of Vitamin A. *Science* 315, no 5813: 820–25.

Kiefer, C., S. Hessel, J. M. Lampert, et al. 2001. Identification and Characterization of a Mammalian Enzyme Catalyzing the Asymmetric Oxidative Cleavage of Provitamin A. *J Biol Chem* 276, no 17: 14110–16.

Kim, Y. K., L. Wassef, S. Chung, et al. 2011. β-Carotene and Its Cleavage Enzyme β-Carotene-15,15′-Oxygenase (CmoI) Affect Retinoid Metabolism in Developing Tissues. *Faseb J* 25, no 5: 1641–52.

Koop, D., and L. Z. Holland. 2008. The Basal Chordate Amphioxus as a Simple Model for Elucidating Developmental Mechanisms in Vertebrates. *Birth Defects Res C Embryo Today* 84, no 3: 175–87.

Koop, D., L. Z. Holland, D. Setiamarga, M. Schubert, and N. D. Holland. 2011. Tail Regression Induced by Elevated Retinoic Acid Signaling in Amphioxus Larvae Occurs by Tissue Remodeling, Not Cell Death. *Evol Dev* 13, no 5: 427–35.

Koop, D., N. D. Holland, M. Sémon, et al. 2010. Retinoic Acid Signaling Targets *Hox* Genes During the Amphioxus Gastrula Stage: Insights into Early Anterior-Posterior Patterning of the Chordate Body Plan. *Dev Biol* 338, no 1: 98–106.

Kopinke, D., J. Sasine, J. Swift, W. Z. Stephens, and T. Piotrowski. 2006. Retinoic Acid Is Required for Endodermal Pouch Morphogenesis and Not for Pharyngeal Endoderm Specification. *Dev Dynam* 235, no 10: 2695–709.

Krezel, W., V. Dupé, M. Mark, et al. 1996. Rxrγ Null Mice Are Apparently Normal and Compound $Rxr\alpha^{+/-}/Rxr\beta^{-/-}/Rxr\gamma^{-/-}$ Mutant Mice Are Viable. *P Natl Acad Sci USA* 93, no 17: 9010–4.

Krispin, S., E. Nitzan, Y. Kassem, and C. Kalcheim. 2010. Evidence for a Dynamic Spatiotemporal Fate Map and Early Fate Restrictions of Premigratory Avian Neural Crest. *Development* 137, no 4: 585–95.

Kruse, S. W., K. Suino-Powell, X. E. Zhou, et al. 2008. Identification of Coup-Tfii Orphan Nuclear Receptor as a Retinoic Acid-Activated Receptor. *Plos Biol* 6, no 9: 2002–15.

Lampert, J. M., J. Holzschuh, S. Hessel, et al. 2003. Provitamin A Conversion to Retinal Via the β,β-Carotene-15,15′-Oxygenase (*Bcox*) Is Essential for Pattern Formation and Differentiation During Zebrafish Embryogenesis. *Development* 130, no 10: 2173–86.

Lee, S. J., S. Kim, S. C. Choi, and J. K. Han. 2010. *Xpteg* (*Xenopus Proximal Tubules-Expressed Gene*) Is Essential for Pronephric Mesoderm Specification and Tubulogenesis. *Mech Develop* 127, no 1-2: 49–61.

Li, Z., V. Korzh, and Z. Y. Gong. 2007. Localized *Rbp4* Expression in the Yolk Syncytial Layer Plays a Role in Yolk Cell Extension and Early Liver Development. *Bmc Dev Biol* 7: 117.

Lin, M., M. Zhang, M. Abraham, S. M. Smith, and J. L. Napoli. 2003. Mouse Retinal Dehydrogenase 4 (Raldh4), Molecular Cloning, Cellular Expression, and Activity in 9-*cis*-Retinoic Acid Biosynthesis in Intact Cells. *J Biol Chem* 278, no 11: 9856–61.

Lin, S. C., P. Dollé, L. Ryckebusch, et al. 2010. Endogenous Retinoic Acid Regulates Cardiac Progenitor Differentiation. *P Natl Acad Sci USA* 107, no 20: 9234–39.

Linney, E., S. Donerly, L. Mackey, and B. Dobbs-McAuliffe. 2011. The Negative Side of Retinoic Acid Receptors. *Neurotoxicol Teratol* 33, no 6: 631–40.

Liu, J. P., E. Laufer, and T. M. Jessell. 2001. Assigning the Positional Identity of Spinal Motor Neurons: Rostrocaudal Patterning of Hox-C Expression by Fgfs, Gdf11, and Retinoids. *Neuron* 32, no 6: 997–1012.

Lobo, G. P., J. Amengual, G. Palczewski, D. Babino, and J. von Lintig. 2012. Mammalian Carotenoid-Oxygenases: Key Players for Carotenoid Function and Homeostasis. *Biochem Biophys Acta* 1821, no 1: 78–87.

López-Carballo, G., L. Moreno, S. Masiá, P. Pérez, and D. Barettino. 2002. Activation of the Phosphatidylinositol 3-Kinase/Akt Signaling Pathway by Retinoic Acid Is Required for Neural Differentiation of Sh-Sy5y Human Neuroblastoma Cells. *J Biol Chem* 277, no 28: 25297–304.

MacLean, G., H. Li, D. Metzger, P. Chambon, and M. Petkovich. 2007. Apoptotic Extinction of Germ Cells in Testes of *Cyp26b1* Knockout Mice. *Endocrinology* 148, no 10: 4560–67.

Maden, M. 2002. Retinoid Signalling in the Development of the Central Nervous System. *Nat Rev Neurosci* 3, no 11: 843–53.

Mark, M., N. B. Ghyselinck, and P. Chambon. 2006. Function of Retinoid Nuclear Receptors. Lessons from Genetic and Pharmacological Dissections of the Retinoic Acid Signaling Pathway During Mouse Embryogenesis. *Annu Rev Pharmacol* 46: 451–80.

Mark, M., N. B. Ghyselinck, and P. Chambon. 2009. Function of Retinoic Acid Receptors During Embryonic Development. *Nucl Recept Signal* 7: e002.

Marletaz, F., L. Z. Holland, V. Laudet, and M. Schubert. 2006. Retinoic Acid Signaling and the Evolution of Chordates. *Int J Biol Sci* 2, no 2: 38–47.

Marshall, H., A. Morrison, M. Studer, H. Pöpperl, and R. Krumlauf. 1996. Retinoids and Hox Genes. *Faseb J* 10, no 9: 969–78.

Martínez-Morales, P. L., R. D. del Corral, I. Olivera-Martínez, et al. 2011. Fgf and Retinoic Acid Activity Gradients Control the Timing of Neural Crest Cell Emigration in the Trunk. *J Cell Biol* 194, no 3: 489–503.

Matt, N., V. Dupé, J. M. Garnier, et al. 2005. Retinoic Acid-Dependent Eye Morphogenesis Is Orchestrated by Neural Crest Cells. *Development* 132, no 21: 4789–800.

Matt, N., N. B. Ghyselinck, O. Wendling, P. Chambon, and M. Mark. 2003. Retinoic Acid-Induced Developmental Defects Are Mediated by Rarβ/Rxr Heterodimers in the Pharyngeal Endoderm. *Development* 130, no 10: 2083–93.

Matt, N., C. K. Schmidt, V. Dupé, et al. 2005. Contribution of Cellular Retinol-Binding Protein Type 1 to Retinol Metabolism During Mouse Development. *Dev Dynam* 233, no 1: 167–76.

McCaffery, P., E. Wagner, J. O'Neil, M. Petkovich, and U. C. Dräger. 1999. Dorsal and Ventral Retinal Territories Defined by Retinoic Acid Synthesis, Break-Down and Nuclear Receptor Expression. *Mech Develop* 82, no 1-2: 119–30.

McNamara, P., S. B. Seo, R. D. Rudic, et al. 2001. Regulation of Clock and Mop4 by Nuclear Hormone Receptors in the Vasculature: A Humoral Mechanism to Reset a Peripheral Clock. *Cell* 105, no 7: 877–89.

Mic, F. A., R. J. Haselbeck, A. E. Cuenca, and G. Duester. 2002. Novel Retinoic Acid Generating Activities in the Neural Tube and Heart Identified by Conditional Rescue of *Raldh2* Null Mutant Mice. *Development* 129, no 9: 2271–82.

Mic, F. A., A. Molotkov, D. M. Benbrook, and G. Duester. 2003. Retinoid Activation of Retinoic Acid Receptor but Not Retinoid X Receptor Is Sufficient to Rescue Lethal Defect in Retinoic Acid Synthesis. *P Natl Acad Sci USA* 100, no 12: 7135–40.

Minoux, M., and F. M. Rijli. 2010. Molecular Mechanisms of Cranial Neural Crest Cell Migration and Patterning in Craniofacial Development. *Development* 137, no 16: 2605–21.

Molotkov, A., N. Molotkova, and G. Duester. 2005. Retinoic Acid Generated by Raldh2 in Mesoderm Is Required for Mouse Dorsal Endodermal Pancreas Development. *Dev Dynam* 232, no 4: 950–57.

Molotkova, N., A. Molotkov, I. O. Sirbu, and G. Duester. 2005. Requirement of Mesodermal Retinoic Acid Generated by *Raldh2* for Posterior Neural Transformation. *Mech Develop* 122, no 2: 145–55.

Moreno, T. A., and C. Kintner. 2004. Regulation of Segmental Patterning by Retinoic Acid Signaling During *Xenopus* Somitogenesis. *Dev Cell* 6, no 2: 205–18.

Müller, W. A. 1984. Retinoids and Pattern Formation in a Hydroid. *J Embryol Exp Morph* 81, no Jun: 253–71.

Müller, W. E. G., M. Binder, J. von Lintig, et al. 2011. Interaction of the Retinoic Acid Signaling Pathway with Spicule Formation in the Marine Sponge *Suberites domuncula* Through Activation of Bone Morphogenetic Protein-1. *Biochem Biophys Acta* 1810, no 12: 1178–94.

Nagatomo, K., and S. Fujiwara. 2003. Expression of *Raldh2*, *Cyp26* and *Hox-1* in Normal and Retinoic Acid-Treated *Ciona intestinalis* Embryos. *Gene Expr Patterns* 3, no 3: 273–77.

Nagatomo, K., T. Ishibashi, Y. Satou, N. Satoh, and S. Fujiwara. 2003. Retinoic Acid Affects Gene Expression and Morphogenesis Without Upregulating the Retinoic Acid Receptor in the Ascidian *Ciona intestinalis*. *Mech Develop* 120, no 3: 363–72.

Nagl, F., K. Schönhofer, B. Seidler, et al. 2009. Retinoic Acid-Induced Nnos Expression Depends on a Novel Pi3k/Akt/Dax1 Pathway in Human Tgw-Nu-I Neuroblastoma Cells. *Am J Physiol Cell Physiol* 297, no 5: C1146–56.

Napoli, J. L. 1999. Interactions of Retinoid Binding Proteins and Enzymes in Retinoid Metabolism. *Biochem Biophys Acta* 1440, no 2-3: 139–62.

Napoli, J. L. 2000. A Gene Knockout Corroborates the Integral Function of Cellular Retinol-Binding Protein in Retinoid Metabolism. *Nutr Rev* 58, no 8: 230–6.

Natale, A., C. Sims, M. L. Chiusano, et al. 2011. Evolution of Anterior *Hox* Regulatory Elements among Chordates. *Bmc Evol Biol* 11: 330.

Niederreither, K., and P. Dollé. 2008. Retinoic Acid in Development: Towards an Integrated View. *Nat Rev Genet* 9, no 7: 541–53.

Niederreither, K., V. Subbarayan, P. Dollé, and P. Chambon. 1999. Embryonic Retinoic Acid Synthesis Is Essential for Early Mouse Post-Implantation Development. *Nat Genet* 21, no 4: 444–48.

Niederreither, K., J. Vermot, I. Le Roux, et al. 2003. The Regional Pattern of Retinoic Acid Synthesis by Raldh2 Is Essential for the Development of Posterior Pharyngeal Arches and the Enteric Nervous System. *Development* 130, no 11: 2525–34.

Novitch, B. G., H. Wichterle, T. M. Jessell, and S. Sockanathan. 2003. A Requirement for Retinoic Acid-Mediated Transcriptional Activation in Ventral Neural Patterning and Motor Neuron Specification. *Neuron* 40, no 1: 81–95.

Noy, N. 2000. Retinoid-Binding Proteins: Mediators of Retinoid Action. *Biochem J* 348: 481–95.

Noy, N. 2010. Between Death and Survival: Retinoic Acid in Regulation of Apoptosis. *Annu Rev Nutr* 30: 201–17.

Ochoa, W. F., A. Torrecillas, I. Fita, et al. 2003. Retinoic Acid Binds to the C2-Domain of Protein Kinase Cα. *Biochemistry* 42, no 29: 8774–79.

Ogasawara, M., H. Wada, H. Peters, and N. Satoh. 1999. Developmental Expression of *Pax1/9* Genes in Urochordate and Hemichordate Gills: Insight into Function and Evolution of the Pharyngeal Epithelium. *Development* 126, no 11: 2539–50.

Osborne, P. W., G. Benoit, V. Laudet, M. Schubert, and D. E. K. Ferrier. 2009. Differential Regulation of Parahox Genes by Retinoic Acid in the Invertebrate Chordate Amphioxus (*Branchiostoma floridae*). *Dev Biol* 327, no 1: 252–62.

Parés, X., J. Farrés, N. Kedishvili, and G. Duester. 2008. Medium-Chain and Short-Chain Dehydrogenases/Reductases in Retinoid Metabolism. *Cell Mol Life Sci* 65, no 24: 3936–49.

Petkovich, M., N. J. Brand, A. Krust, and P. Chambon. 1987. A Human Retinoic Acid Receptor Which Belongs to the Family of Nuclear Receptors. *Nature* 330, no 6147: 444–50.

Piskunov, A., and C. Rochette-Egly. 2011. A Retinoic Acid Receptor Rarα Pool Present in Membrane Lipid Rafts Forms Complexes with G Protein Aq to Activate P38mapk. *Oncogene 31, no 28: 3333–45.*

Plant, M. R., M. E. MacDonald, L. I. Grad, S. J. Ritchie, and J. M. Richman. 2000. Locally Released Retinoic Acid Repatterns the First Branchial Arch Cartilages *in vivo*. *Dev Biol* 222, no 1: 12–26.

Pourquié, O. 2003. The Segmentation Clock: Converting Embryonic Time into Spatial Pattern. *Science* 301, no 5631: 328–30.

Quadro, L., L. Hamberger, M. E. Gottesman, et al. 2004. Transplacental Delivery of Retinoid: The Role of Retinol-Binding Protein and Lipoprotein Retinyl Ester. *Am J Physiol-Endoc M* 286, no 5: E844–E51.

Quadro, L., L. Hamberger, M. E. Gottesman, et al. 2005. Pathways of Vitamin A Delivery to the Embryo: Insights from a New Tunable Model of Embryonic Vitamin A Deficiency. *Endocrinology* 146, no 10: 4479–90.

Reijntjes, S., E. Gale, and M. Maden. 2004. Generating Gradients of Retinoic Acid in the Chick Embryo: *Cyp26c1* Expression and a Comparative Analysis of the Cyp26 Enzymes. *Dev Dynam* 230, no 3: 509–17.

Rhinn, M., and P. Dollé. 2012. Retinoic Acid Signalling During Development. *Development* 139, no 5: 843–58.

Ribes, V., I. Le Roux, M. Rhinn, B. Schuhbaur, and P. Dollé. 2009. Early Mouse Caudal Development Relies on Crosstalk Between Retinoic Acid, Shh and Fgf Signalling Pathways. *Development* 136, no 4: 665–76.

Richardson, S. J. 2009. Evolutionary Changes to Transthyretin: Evolution of Transthyretin Biosynthesis. *Febs J* 276, no 19: 5342–56.

Rijli, F. M., A. Gavalas, and P. Chambon. 1998. Segmentation and Specification in the Branchial Region of the Head: The Role of the *Hox* Selector Genes. *Int J Dev Biol* 42, no 3: 393–401.

Rosenthal, N., and J. Xavier-Neto. 2000. From the Bottom of the Heart: Anteroposterior Decisions in Cardiac Muscle Differentiation. *Curr Opin Cell Biol* 12, no 6: 742–46.

Ross, A. C. 2003. Retinoid Production and Catabolism: Role of Diet in Regulating Retinol Esterification and Retinoic Acid Oxidation. *J Nutr* 133, no 1: 291s–96s.

Ross, A. C., and R. Zolfaghari. 2011. Cytochrome P450s in the Regulation of Cellular Retinoic Acid Metabolism. *Annu Rev Nutr* 31, no 1: 65–87.

Ross, S. A., P. J. McCaffery, U. C. Drager, and L. M. De Luca. 2000. Retinoids in Embryonal Development. *Physiol Rev* 80, no 3: 1021–54.

Rosselot, C., L. Spraggon, I. Chia, et al. 2010. Non-Cell-Autonomous Retinoid Signaling Is Crucial for Renal Development. *Development* 137, no 2: 283–92.

Samarut, E., and C. Rochette-Egly. 2012. Nuclear Retinoic Acid Receptors: Conductors of the Retinoic Acid Symphony During Development. *Mol Cell Endocrinol* 348, no 2: 348–60.

Sandell, L. L., B. W. Sanderson, G. Moiseyev, et al. 2007. Rdh10 Is Essential for Synthesis of Embryonic Retinoic Acid and Is Required for Limb, Craniofacial, and Organ Development. *Genes Dev* 21, no 9: 1113–24.

Sasakura, Y., M. Kanda, T. Ikeda, et al. 2012. Retinoic Acid-Driven *Hox1* Is Required in the Epidermis for Forming the Otic/Atrial Placodes During Ascidian Metamorphosis. *Development* 139, no 12: 2156–60.

Schlosser, G. 2010. Making Senses: Development of Vertebrate Cranial Placodes. *Int Rev Cell Mol Biol* 283: 129–234.

Schubert, M., H. Escriva, J. Xavier-Neto, and V. Laudet. 2006. Amphioxus and Tunicates as Evolutionary Model Systems. *Trends Ecol Evol* 21, no 5: 269–77.

Schubert, M., N. D. Holland, H. Escriva, L. Z. Holland, and V. Laudet. 2004. Retinoic Acid Influences Anteroposterior Positioning of Epidermal Sensory Neurons and Their Gene Expression in a Developing Chordate (Amphioxus). *P Natl Acad Sci USA* 101, no 28: 10320–25.

Schubert, M., N. D. Holland, V. Laudet, and L. Z. Holland. 2006. A Retinoic Acid-*Hox* Hierarchy Controls Both Anterior/Posterior Patterning and Neuronal Specification in the Developing Central Nervous System of the Cephalochordate Amphioxus. *Dev Biol* 296, no 1: 190–202.

Schubert, M., J. K. Yu, N. D. Holland, et al. 2005. Retinoic Acid Signaling Acts Via Hox1 to Establish the Posterior Limit of the Pharynx in the Chordate Amphioxus. *Development* 132, no 1: 61–73.

Schug, T. T., D. C. Berry, N. S. Shaw, S. N. Travis, and N. Noy. 2007. Opposing Effects of Retinoic Acid on Cell Growth Result from Alternate Activation of Two Different Nuclear Receptors. *Cell* 129, no 4: 723–33.

Seo, H. C., R. B. Edvardsen, A. D. Maeland, et al. 2004. Hox Cluster Disintegration with Persistent Anteroposterior Order of Expression in *Oikopleura dioica*. *Nature* 431, no 7004: 67–71.

Serluca, F. C., and M. C. Fishman. 2001. Pre-Pattern in the Pronephric Kidney Field of Zebrafish. *Development* 128, no 12: 2233–41.

Shaw, N., M. Elholm, and N. Noy. 2003. Retinoic Acid Is a High Affinity Selective Ligand for the Peroxisome Proliferator-Activated Receptor β/δ. *J Biol Chem* 278, no 43: 41589–92.

Sheng, N. Y., Z. H. Xie, C. Wang, et al. 2010. Retinoic Acid Regulates Bone Morphogenic Protein Signal Duration by Promoting the Degradation of Phosphorylated Smad1. *P Natl Acad Sci USA* 107, no 44: 18886–91.

Shirai, H., K. Oishi, and N. Ishida. 2006. Bidirectional Clock/Bmal1-Dependent Circadian Gene Regulation by Retinoic Acid *in vitro*. *Biochem Bioph Res Co* 351, no 2: 387–91.

Simões-Costa, M. S., A. P. Azambuja, and J. Xavier-Neto. 2008. The Search for Non-Chordate Retinoic Acid Signaling: Lessons from Chordates. *J Exp Zool Part B* 310B, no 1: 54–72.

Sirbu, I. O., and G. Duester. 2006. Retinoic-Acid Signalling in Node Ectoderm and Posterior Neural Plate Directs Left-Right Patterning of Somitic Mesoderm. *Nat Cell Biol* 8, no 3: 271–77.

Sirbu, I. O., L. Gresh, J. Barra, and G. Duester. 2005. Shifting Boundaries of Retinoic Acid Activity Control Hindbrain Segmental Gene Expression. *Development* 132, no 11: 2611–22.

Sirbu, I. O., X. L. Zhao, and G. Duester. 2008. Retinoic Acid Controls Heart Anteroposterior Patterning by Down-Regulating *Isl1* Through the *Fgf8* Pathway. *Dev Dynam* 237, no 6: 1627–35.

Sobreira, T. J. P., F. Marlétaz, M. Simões-Costa, et al. 2011. Structural Shifts of Aldehyde Dehydrogenase Enzymes Were Instrumental for the Early Evolution of Retinoid-Dependent Axial Patterning in Metazoans. *P Natl Acad Sci USA* 108, no 1: 226–31.

Song, Y., J. N. Hui, K. K. Fu, and J. M. Richman. 2004. Control of Retinoic Acid Synthesis and Fgf Expression in the Nasal Pit Is Required to Pattern the Craniofacial Skeleton. *Dev Biol* 276, no 2: 313–29.

Spiegler, E., Y. K. Kim, L. Wassef, V. Shete, and L. Quadro. 2012. Maternal-Fetal Transfer and Metabolism of Vitamin A and Its Precursor β-Carotene in the Developing Tissues. *Biochem Biophys Acta* 1821, no 1: 88–98.

Stafford, D., A. Hornbruch, P. R. Mueller, and V. E. Prince. 2004. A Conserved Role for Retinoid Signaling in Vertebrate Pancreas Development. *Dev Genes Evol* 214, no 9: 432–41.

Stainier, D. Y. R., and M. C. Fishman. 1992. Patterning the Zebrafish Heart Tube: Acquisition of Anteroposterior Polarity. *Dev Biol* 153, no 1: 91–101.

Stehlin-Gaon, C., D. Willmann, D. Zeyer, et al. 2003. All-*trans* Retinoic Acid Is a Ligand for the Orphan Nuclear Receptor Rorβ. *Nat Struct Biol* 10, no 10: 820–25.

Sun, H. 2012. Membrane Receptors and Transporters Involved in the Function and Transport of Vitamin A and Its Derivatives. *Biochem Biophys Acta* 1821, no 1: 99–112.

Suruga, K., T. Goda, M. Igarashi, et al. 1997. Cloning of Chick Cellular Retinol-Binding Protein, Type II and Comparison to That of Some Mammals: Expression of the Gene at Different Developmental Stages, and Possible Involvement of Rxrs and Ppar. *Comp Biochem Physiol A Physiol* 118, no 3: 859–69.

Tan, N. S., N. S. Shaw, N. Vinckenbosch, et al. 2002. Selective Cooperation Between Fatty Acid Binding Proteins and Peroxisome Proliferator-Activated Receptors in Regulating Transcription. *Mol Cell Biol* 22, no 14: 5114–27.

Tanaka, Y., Y. Okada, and N. Hirokawa. 2005. Fgf-Induced Vesicular Release of Sonic Hedgehog and Retinoic Acid in Leftward Nodal Flow Is Critical for Left-Right Determination. *Nature* 435, no 7039: 172–77.

Theodosiou, M., V. Laudet, and M. Schubert. 2010. From Carrot to Clinic: An Overview of the Retinoic Acid Signaling Pathway. *Cell Mol Life Sci* 67, no 9: 1423–45.

Tickle, C., B. Alberts, L. Wolpert, and J. Lee. 1982. Local Application of Retinoic Acid to the Limb Bond Mimics the Action of the Polarizing Region. *Nature* 296, no 5857: 564–66.

Trainor, P. A., and R. Krumlauf. 2001. *Hox* Genes, Neural Crest Cells and Branchial Arch Patterning. *Curr Opin Cell Biol* 13, no 6: 698–705.

Uehara, M., K. Yashiro, S. Mamiya, et al. 2007. Cyp26a1 and Cyp26c1 Cooperatively Regulate Anterior-Posterior Patterning of the Developing Brain and the Production of Migratory Cranial Neural Crest Cells in the Mouse. *Dev Biol* 302, no 2: 399–411.

Van de Peer, Y., S. Maere, and A. Meyer. 2009. The Evolutionary Significance of Ancient Genome Duplications. *Nat Rev Genet* 10, no 10: 725–32.

Vermot, J., and O. Pourquié. 2005. Retinoic Acid Coordinates Somitogenesis and Left-Right Patterning in Vertebrate Embryos. *Nature* 435, no 7039: 215–20.

Vilhais-Neto, G. C., and O. Pourquié. 2008. Retinoic Acid. *Curr Biol* 18, no 7: 550–52.

von Lintig, J., S. Hessel, A. Isken, et al. 2005. Towards a Better Understanding of Carotenoid Metabolism in Animals. *Biochem Biophys Acta* 1740, no 2: 122–31.

Wada, H., H. Escriva, S. C. Zhang, and V. Laudet. 2006. Conserved Rare Localization in Amphioxus *Hox* Clusters and Implications for *Hox* Code Evolution in the Vertebrate Neural Crest. *Dev Dynam* 235, no 6: 1522–31.

Wang, Z. X., P. Dollé, W. V. Cardoso, and K. Niederreither. 2006. Retinoic Acid Regulates Morphogenesis and Patterning of Posterior Foregut Derivatives. *Dev Biol* 297, no 2: 433–45.

Wasiak, S., and D. Lohnes. 1999. Retinoic Acid Affects Left-Right Patterning. *Dev Biol* 215, no 2: 332–42.

Waxman, J. S., and D. Yelon. 2007. Comparison of the Expression Patterns of Newly Identified Zebrafish Retinoic Acid and Retinoid X Receptors. *Dev Dynam* 236, no 2: 587–95.

White, J. C., M. Highland, M. Kaiser, and M. Clagett-Dame. 2000. Vitamin A Deficiency Results in the Dose-Dependent Acquisition of Anterior Character and Shortening of the Caudal Hindbrain of the Rat Embryo. *Dev Biol* 220, no 2: 263–84.

White, R. J., and T. F. Schilling. 2008. How Degrading: Cyp26s in Hindbrain Development. *Dev Dynam* 237, no 10: 2775–90.

Wiens, M., R. Batel, M. Korzhev, and W. E. G. Müller. 2003. Retinoid X Receptor and Retinoic Acid Response in the Marine Sponge *Suberites domuncula*. *J Exp Biol* 206, no 18: 3261–71.

Wilson, J. G., C. B. Roth, and J. Warkany. 1953. An Analysis of the Syndrome of Malformations Induced by Maternal Vitamin A Deficiency. Effects of Restoration of Vitamin A at Various Times During Gestation. *Am J Anat* 92, no 2: 189–217.

Wilson, J. G., and J. Warkany. 1949. Aortic-Arch and Cardiac Anomalies in the Offspring of Vitamin A Deficient Rats. *Am J Anat* 85, no 1: 113–55.

Wingert, R. A., R. Selleck, J. Yu, et al. 2007. The *Cdx* Genes and Retinoic Acid Control the Positioning and Segmentation of the Zebrafish Pronephros. *Plos Genet* 3, no 10: 1922–38.

Xavier-Neto, J., N. Rosenthal, F. A. Silva, et al. 2001. Retinoid Signaling and Cardiac Anteroposterior Segmentation. *Genesis* 31, no 3: 97–104.

Yamauchi, K., and A. Ishihara. 2009. Evolutionary Changes to Transthyretin: Developmentally Regulated and Tissue-Specific Gene Expression. *Febs J* 276, no 19: 5357–66.

Yashiro, K., X. L. Zhao, M. Uehara, et al. 2004. Regulation of Retinoic Acid Distribution Is Required for Proximodistal Patterning and Outgrowth of the Developing Mouse Limb. *Dev Cell* 6, no 3: 411–22.

Yasuda, Y., N. Nishi, J. A. Takahashi, et al. 1992. Induction of Avascular Yolk Sac Due to Reduction of Basic Fibroblast Growth Factor by Retinoic Acid in Mice. *Dev Biol* 150, no 2: 397–413.

Zanotto-Filho, A., M. Cammarota, D. P. Gelain, et al. 2008. Retinoic Acid Induces Apoptosis by a Non-Classical Mechanism of Erk1/2 Activation. *Toxicol in vitro* 22, no 5: 1205–12.

Zhang, Z. Y., J. E. Balmer, A. Lovlie, S. H. Fromm, and R. Blomhoff. 1996. Specific Teratogenic Effects of Different Retinoic Acid Isomers and Analogs in the Developing Anterior Central Nervous System of Zebrafish. *Dev Dynam* 206, no 1: 73–86.

Zhao, X. L., T. Brade, T. J. Cunningham, and G. Duester. 2010. Retinoic Acid Controls Expression of Tissue Remodeling Genes *Hmgn1* and *Fgf18* at the Digit–Interdigit Junction. *Dev Dynam* 239, no 2: 665–71.

Zhao, X. L., I. O. Sirbu, F. A. Mlc, et al. 2009. Retinoic Acid Promotes Limb Induction Through Effects on Body Axis Extension but Is Unnecessary for Limb Patterning. *Curr Biol* 19, no 12: 1050–57.

Zhou, X. E., K. M. Suino-Powell, Y. Xu, et al. 2011. The Orphan Nuclear Receptor Tr4 Is a Vitamin A-Activated Nuclear Receptor. *J Biol Chem* 286, no 4: 2877–85.

Appendix 5

Published article

Evolution of bilaterian central nervous systems: a single origin?

Holland LZ, Carvalho JE, Escriva H, Laudet V, Schubert M, Shimeld SM, Yu J-K.

EvoDevo. 2013;4(1):27.

(20 pages)

Holland et al. EvoDevo 2013, 4:27
http://www.evodevojournal.com/content/4/1/27

EvoDevo

REVIEW Open Access

Evolution of bilaterian central nervous systems: a single origin?

Linda Z Holland[1*], João E Carvalho[2], Hector Escriva[3], Vincent Laudet[4], Michael Schubert[2], Sebastian M Shimeld[5] and Jr-Kai Yu[6]

Abstract

The question of whether the ancestral bilaterian had a central nervous system (CNS) or a diffuse ectodermal nervous system has been hotly debated. Considerable evidence supports the theory that a CNS evolved just once. However, an alternative view proposes that the chordate CNS evolved from the ectodermal nerve net of a hemichordate-like ancestral deuterostome, implying independent evolution of the CNS in chordates and protostomes. To specify morphological divisions along the anterior/posterior axis, this ancestor used gene networks homologous to those patterning three organizing centers in the vertebrate brain: the anterior neural ridge, the zona limitans intrathalamica and the isthmic organizer, and subsequent evolution of the vertebrate brain involved elaboration of these ancestral signaling centers; however, all or part of these signaling centers were lost from the CNS of invertebrate chordates. The present review analyzes the evidence for and against these theories. The bulk of the evidence indicates that a CNS evolved just once – in the ancestral bilaterian. Importantly, in both protostomes and deuterostomes, the CNS represents a portion of a generally neurogenic ectoderm that is internalized and receives and integrates inputs from sensory cells in the remainder of the ectoderm. The expression patterns of genes involved in medio/lateral (dorso/ventral) patterning of the CNS are similar in protostomes and chordates; however, these genes are not similarly expressed in the ectoderm outside the CNS. Thus, their expression is a better criterion for CNS homologs than the expression of anterior/posterior patterning genes, many of which (for example, *Hox* genes) are similarly expressed both in the CNS and in the remainder of the ectoderm in many bilaterians. The evidence leaves hemichordates in an ambiguous position – either CNS centralization was lost to some extent at the base of the hemichordates, or even earlier, at the base of the hemichordates + echinoderms, or one of the two hemichordate nerve cords is homologous to the CNS of protostomes and chordates. In any event, the presence of part of the genetic machinery for the anterior neural ridge, the zona limitans intrathalamica and the isthmic organizer in invertebrate chordates together with similar morphology indicates that these organizers were present, at least in part, at the base of the chordates and were probably elaborated upon in the vertebrate lineage.

Keywords: Central nervous system evolution, Hemichordate, Urbilaterian, Amphioxus, Tunicate, Vertebrate brain, Nerve cord

Review

Introduction

There is general agreement that the relatively complex central nervous system (CNS) characterizing most higher metazoan animals can be traced back through evolution to a nerve net in a cnidarian-like ancestor. However, it is highly controversial whether the nervous system of the next evolutionary stage (the urbilaterian) still consisted solely of a nerve net or included a CNS. If the urbilaterian had only a nerve net, then the CNSs of protostomes and deuterostomes likely evolved independently. In contrast, the view that the urbilaterian had a CNS is consistent with the view that the CNSs of all metazoans are homologous. At present, opinion is still divided, with the majority advocating a single evolutionary origin for the CNS [1-8] and the minority favoring an urbilaterian with a nerve net [9,10].

* Correspondence: lzholland@ucsd.edu
[1]Marine Biology Research Division, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093-0202, USA
Full list of author information is available at the end of the article

This controversy about CNS evolution is intimately related to issues of homology, and it is useful to outline current thinking about homology at the outset. It is important not to conflate the concepts and recognition criteria for homology. There are currently three concepts underlying homology – biological [11], taxic/cladistic [12,13] and historical [14-16]. What matters for the last, which is the most familiar and most germane for the nervous system controversy, is the historical continuity of descent from a common ancestor. The historical concept requires one to be explicit about what is being compared [17]; for example, bird wings and bat wings are homologous as vertebrate forelimbs, but not as wings. Importantly, historical homologies can become very different through divergence. The three chief criteria for recognizing homology are relative position to other body parts, special quality, and transitional stages [18]. A developmental criterion, introduced by Haeckel [19], proved difficult to apply and has been largely submerged into the previous three criteria. Importantly, there are differing views concerning the hierarchical distribution of homology across levels of biological organization. In the view of Striedter and Northcutt, homology at one level (say behavior) does not necessarily connote homology at another (say morphology) [20]. In contrast, Wagner argued that structures descended from a common ancestor are homologous even if they have diverged and have no clear morphological similarity [16]. The problems raised by the hierarchical nature of homology have been heightened by the discovery of developmental gene conservation [21] and are especially noticeable in the discussion of CNS evolution.

In the past 20 years, it has been found that developmental genes and core signaling pathways are typically conserved across phyla and that gene expression patterns during development can often be used as characters for inferring homologies. Thus, although the majority view had been that the urbilaterian had a nerve net, the balance was tipped towards an urbilaterian with a CNS by the discovery that *bone morphogenic protein* (*BMP*)*/decapentaplegic* genes were expressed dorsally in *Drosophila* and ventrally in vertebrates with the BMP antagonists *chordin/short gastrulation* expressed on the opposite side [22] (Figure 1). In this view, which is consistent with the CNSs of all higher metazoans being homologous, a dorso/ventral (D/V) inversion occurred either in basal protostomes or in the deuterostome lineage [3,5,23]. However, in the last 10 years, studies of gene expression and function in an enteropneust (acorn worm; phylum Hemichordata) have been interpreted as evidence that the ancestral deuterostome and, by extension, the urbilaterian had a nerve net and no CNS [9,24,25]. Thus, while a CNS would have arisen close to the base of the protostomes, the evolution of a CNS in deuterostomes did not occur until the base of the chordates. In the present review, we examine the detailed

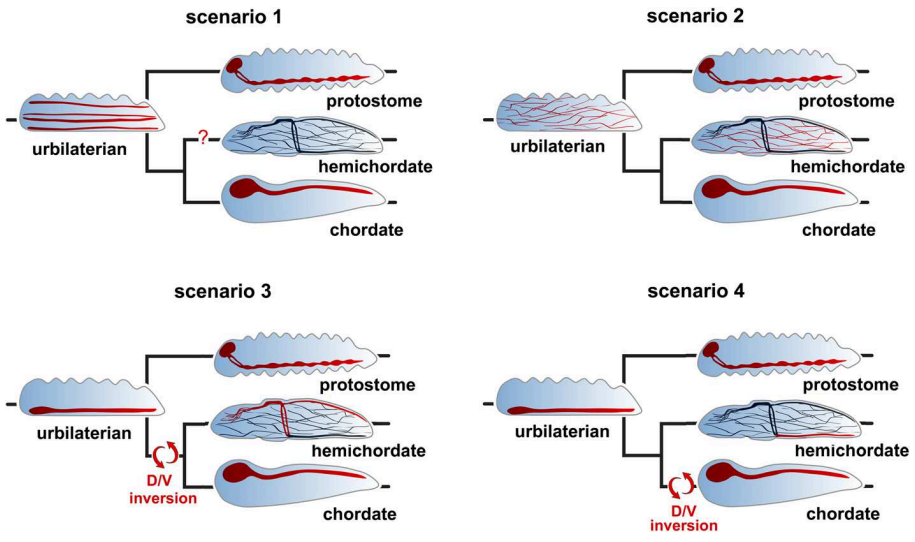


Figure 1 Four scenarios for evolution of central nervous systems in bilaterians. In scenario 1, the urbilaterian had multiple nerve cords, one of which evolved into the dorsal central nervous system (CNS) of chordates, while another nerve cord evolved into the ventral CNS of protostomes. In scenario 2, the CNSs of protostomes and deuterostomes evolved independently from an ectodermal nerve net in the bilaterian ancestor. In scenario 3, the chordate and protostome nerve cords evolved from a ventral nerve cord in the urbilaterian ancestor. A dorso/ventral (D/V) inversion occurred at the base of the deuterostomes; the dorsal nerve cord of hemichordates is thus homologous to the chordate CNS and to the protostome ventral nerve cord. In scenario 4, the protostome and chordate nerve cords evolved from the CNS of an urbilaterian ancestor, but a D/V inversion occurred at the base of the chordates. Thus, the ventral nerve cord of a hemichordate is homologous to the chordate and protostome CNSs. Scenarios after [1,3,7,23,26-29].

evidence on both sides of the controversy and evaluate its interpretations. We conclude that a stronger case can be made for the initial appearance of the CNS at the level of the urbilaterian than for independent evolution of the CNS in more than one line of metazoan descent.

Reconstructing the ancestral bilaterian

Although several features of the ancestral bilaterian in addition to the presence or absence of a CNS are widely debated, a range of molecular, developmental and comparative morphological evidence indicates that this animal was bilaterally symmetrical, with distinct anterior and posterior ends, dorsal and ventral surfaces, and left and right sides. It almost certainly had defined muscle, derived from mesoderm, allowing active locomotion and a gut with either a single opening or a separate mouth and anus [30]. Whether or not this animal had a CNS, an ectodermal nerve net or some combination of the two has been hotly debated (reviewed in [31]) (Figure 1).

One difficulty in deciding whether the ancestral bilaterian had a CNS is that the ectoderm in bilaterians is broadly neurogenic. Therefore, the distinction between the CNS and the remainder of the relatively neurogenic ectoderm is not always clear-cut. In chordates, arthropods and annelids, the distinction is most clear as there is a fully internalized concentration of neurons, axons and supporting cells along the anterior/posterior (A/P) axis (that is, a CNS) that integrates information from sensory cells both associated with the CNS (for example, eyes) and with other portions of the ectoderm and coordinates behavior. Importantly, the CNS in these organisms has an anterior concentration of discrete neural centers or "brain", which coordinates sensory inputs and responses. At the other extreme are "diffuse ectodermal nerve nets" such as in cnidarians. However, such nerve nets are not uniform; specific types of neurons may be regionally localized [32]. An additional problem in understanding the evolution of CNSs comes with the Ambulacraria (echinoderms and hemichordates), as they have both ectodermal nerve nets and nerve cords. It is controversial whether echinoderm and/or hemichordate nerve cords, neither of which has a concentration of neurons that could be termed a brain, and the CNS of chordates have a common evolutionary origin [33,34]. Here we will use the term CNS for a nervous system that is derived from ectoderm, includes both axons and neurons and is specialized along the A/P axis with an anterior concentration of neural centers (brain), and the term "nerve cord" more broadly to include axonal tracts with few or no neurons and lacking a discrete brain. The diversity of animal nervous systems and paucity of data from some species may blur this distinction on occasion; however, we will be explicit in such instances.

What is the evidence for a CNS in the ancestral bilaterian?

It is generally agreed that bilaterians evolved from radially or bi-radially symmetrical animals, comparable in some ways to modern cnidarians. Adult cnidarians have an ectodermal nerve net with a concentration of neurons around the single gut opening (Figure 2). Therefore, if the ancestral bilaterian had already evolved a CNS, it would presumably have arisen as a concentration or amplification of neurons along one side of this nerve net, perhaps together with a reduction in numbers of neurons elsewhere in the ectoderm.

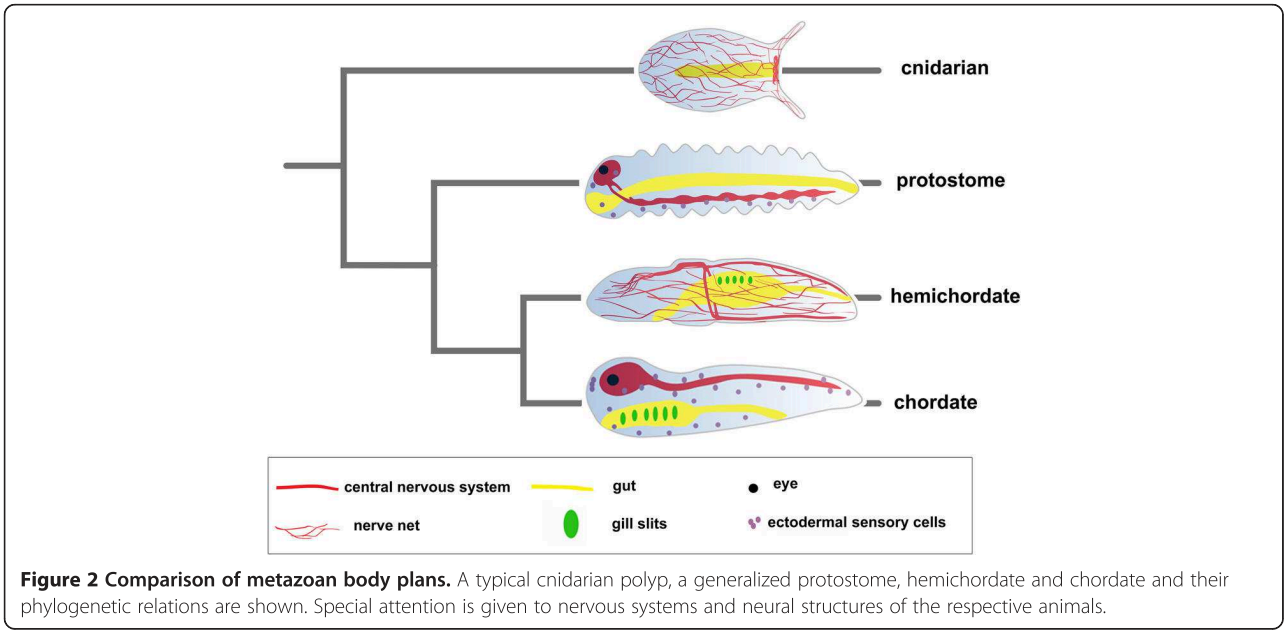


Figure 2 Comparison of metazoan body plans. A typical cnidarian polyp, a generalized protostome, hemichordate and chordate and their phylogenetic relations are shown. Special attention is given to nervous systems and neural structures of the respective animals.

Unfortunately, no extant animal is a good stand-in for this ancestral bilaterian. Extant animals that are thought to have diverged from the bilaterian lineage before it radiated into the protostomes (Ecdysozoa and Lophotrochozoa) and the deuterostomes do not show an intermediate condition between a nerve net and a CNS (Figure 2). The best candidates for such early bilaterian offshoots are the acoel and nemertodermatid flatworms, and the xenoturbellids, which in some studies have been placed basal to the deuterostomes plus protostomes but in others are placed basal in the deuterostome lineage [35,36]. Acoels have a concentration of neurons, or a "brain", anteriorly with up to six tracts of axons extending posteriorly [37]. In contrast, xenoturbellids have an intraepithelial nerve net that lacks aggregations of neurons or axonal tracts [38]. As a result of the lack of a clear intermediate, scenarios for evolution of CNSs are necessarily based on similarities in gene expression and neuroanatomy in the two main lineages of bilaterians: protostomes (Ecdysozoa plus Lophotrochozoa) and deuterostomes (Figure 1).

Regionalization of nerve cords in protostomes and chordates

Because the CNSs in protostomes and deuterostomes are in different positions, develop rather differently and are morphologically somewhat diverse, possible homologies between them have been highly contentious. Complicating the picture is that some of the genetic mechanisms for specifying A/P positions in the CNS are common to the entire organism, including the general ectoderm exterior to the CNS, and are therefore not entirely useful for inferring homologies of CNSs. For example, some genetic mechanisms mediating A/P patterning in the CNS were clearly inherited from a cnidarian-like ancestor in which they patterned the entire body axis. Thus, *Six3/6* and *Irx* are expressed in the aboral region of the planula larva of the sea anemone *Nematostella vectensis*, opposite the blastopore [39] and in the anterior end of the brain of both protostomes and deuterostomes – *Six3/6* in the anterior tip of the CNS and *Irx* genes a little more posteriorly [24,40-44]. In *N. vectensis* the domains of these two genes are initially congruent, while in the CNS of bilaterians the *Six3/6* domain is anterior to that of *Irx*. Therefore, although it is most parsimonious to propose that these genes were coopted into the CNS of an ancestral bilaterian, it cannot be ruled out that they were coopted independently into the CNS of protostomes and deuterostomes.

Hox genes are another example of A/P patterning genes that are not entirely useful for inferring homologies between the protostome and chordate CNS. The problem is that although they do mediate A/P patterning of the CNS in bilaterians [45,46], they mediate A/P patterning of other tissues as well [47-54]. Thus, while their expression patterns have been used to infer homologies between the CNS in insects and vertebrates, it remains possible that they patterned the entire body axis of the Urbilaterian and were independently coopted into the CNSs of protostomes and deuterostomes. It is not clear when a role for nested expression of *Hox* genes in regionalization of the A/P axis evolved. They do not appear to be involved in A/P patterning in cnidarians [55,56]. Comparisons of *Hox* genes in protostomes with up to 10 or 11 *Hox* genes and invertebrate deuterostomes with up to 15 indicate that the ancestral bilaterian had at least eight to 10 *Hox* genes [57], while cnidarians have up to six depending on the species and acoel flatworms have three, which are more or less regionally expressed in the surface ectoderm along the A/P axis [58,59] with later expression in putative neural precursors [49]. Thus, acoels either arose before a large *Hox* cluster evolved or they lost some *Hox* genes. Even so, a role for an expanded array of *Hox* genes in specification of A/P positions in the ectoderm was evidently present in the ancestral bilaterian. Thus, although expression of the *Drosophila melanogaster Hox1* gene *labial* in a stripe at the posterior end of the tritocerebrum within the *unpg* (*Gbx*) domain has been likened to nested expression of *Hox* genes in the vertebrate hindbrain, with *Hoxb1* being expressed in a stripe in rhombomere 4, the possibility that *Hox* genes were independently coopted into the CNS in protostomes and deuterostomes cannot be ruled out.

Stronger support for a single origin of the CNS comes from similar expression in the CNSs of protostomes and chordates of genes that are not expressed in comparable patterns in other tissues. Thus, Reichert and colleagues used gene expression patterns to support the perhaps surprising idea that the three parts of the *Drosophila* brain – protocerebrum, deutocerebrum and tritocerebrum [5,45] – are homologous to the forebrain, midbrain and hindbrain of vertebrates (Figure 2) [60]. For example, in *D. melanogaster*, the *Otx* homolog *Otd* is expressed throughout the protocerebrum and deutocerebrum, while *unpg* (homologous to *Gbx*) is expressed in the tritocerebrum, the subesophageal ganglion and the ventral nerve cord [5]. The domains of the two abut at the boundary between the deutocerebrum and tritocerebrum, similar to the abutting domains of *Otx2* and *Gbx2* at the midbrain/hindbrain boundary (MHB) in vertebrates [5,61]. In addition, although some domains of *Pax2/5/8* genes are not similar between the CNS of flies and chordates, *Pax2/5/8* is expressed at high levels in the posterior part of the deutocerebrum (just anterior to the deutocerebrum/tritocerebrum boundary) in *D. melanogaster*, while the three vertebrate *Pax2/5/8* genes are expressed at high levels at the MHB [5,62]. Moreover, in third instar

larvae of *D. melanogaster*, the *earmuff* gene (homologous to *Fezf*) is broadly expressed in the anterior brain with a posterior boundary at the protocerebrum/deutocerebrum boundary [63]. The domain is just anterior to that of *mirror*, one of the three *Irx* homologs. Similarly, Irimia and colleagues showed that in chordates, the posterior limits of *Fez* genes (*Fez* and *Fez-like*) abut the anterior limit of *Irx1* in the forebrain [64]. In vertebrates, this is the zona limitans intrathalamica (ZLI) [65].

Compatible with a single origin of the CNS, expression of the genes mediating D/V patterning within the CNS is also conserved between protostomes and deuterostomes [66] (Figure 3). These genes are not comparably expressed in cnidarians, suggesting that they were recruited for roles in D/V patterning the CNS of an ancestral bilaterian. Notably, homologs of some key genes expressed mediolaterally in the neuroectoderm of *D. melanogaster* embryos are expressed in comparable domains in the vertebrate CNS. Thus, the *msh* gene is expressed laterally in the *D. melanogaster* neuroectoderm, with *ind* expressed in an intermediate longitudinal domain and *vnd* expressed in a medial stripe of neuroblasts (reviewed in [7,67]). Vertebrate homologs of these three homeobox genes are comparably expressed in the developing neural tube. Two of the three *msh* orthologs (*Msx1, Msx2, Msx3*) are expressed dorsally (that is, laterally) in the roof plate of the CNS, one of the two *ind* orthologs (*Gsh1*) is expressed in the adjacent zone (alar plate), and one of the two *vnd* orthologs (*Nkx2.2*) is expressed more ventrally (that is, medially) in the basal plate.

Additional evidence for homology of protostome and chordate nerve cords, and thus a bilaterian ancestor with a CNS, comes from neuroanatomy, neuronal function and gene expression. Strausfeld and Hirth found striking parallels between the central complex in the arthropod protocerebrum and the basal ganglia in the ventral forebrain of vertebrates [3]. In particular, the vertebrate striatum and pallidum have similar organization as, respectively, the insect fan-shaped body and ellipsoid body. Both the types of neurons and their connections and the functions of these regions are similar in the two organisms. Taken together, the data from comparative gene expression and anatomy provide relatively strong support for a single origin of the CNS in insects and chordates.

Parallels between the brains of annelids and vertebrates

Additional evidence for a single origin of the CNS comes from comparisons between annelids and vertebrates. Not only have parallels been drawn between patterning the *Drosophila* and vertebrate brains, but Arendt and colleagues have also noted similarities between the genetic mechanisms patterning the nervous systems of the annelid *Platynereis dumerilii* and vertebrates [2,81] (reviewed in [4]). The annelid brain varies from species to species, with the brains of some species lacking clear compartments but many others having such features as

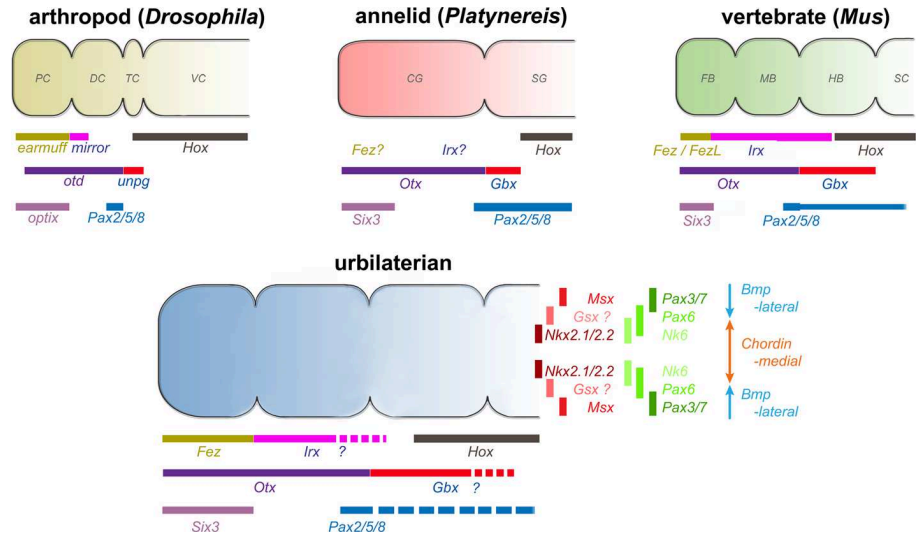


Figure 3 Anterior–posterior gene expression in central nervous systems of three extant bilaterians and the urbilaterian. Anterior–posterior regionalization of gene expression in the central nervous systems of three extant bilaterians (an arthropod, an annelid and a vertebrate) and inferred expression in the last common bilaterian ancestor, the urbilaterian. Expression of *Fez* and *Irx* in the annelid *Platynereis* is unknown. For the urbilaterian, both anterior–posterior and medio-lateral gene expression domains are shown. Hypothetical posterior limits of *Irx* and *Gbx* domains in the urbilaterian brain are highlighted by a "?" and dashed lines. PC, protocerebrum; DC, deutocerebrum; TC, tritocerebrum; VC, ventral nerve cord; CG, cerebral ganglion; SG, segmental ganglia; FB, forebrain; MB, midbrain; HB, hindbrain; SC, spinal cord. Gene expression domains based on [1,2,9,24,29,34,42,64,68-80].

complex, neuron-rich mushroom bodies (a comparatively large part of the brain in insects and annelids that integrates olfactory information) [82]. Extensive comparisons of gene expression have been used to argue for homology between the mushroom bodies and the pallium of the vertebrate brain [2,81]. For example, *Bf-1* (*FoxG1*) is expressed in the anterior part of the vertebrate telencephalon and the pallium as well as in the tip of the annelid brain, while *Wnt5/8* is expressed in the vertebrate pallium and in the annelid mushroom bodies, flanking more medial expression of *Hh* in both [2].

Furthermore, in *P. dumerilii Six3* and *Otx* are expressed anteriorly in the CNS (the peristomium) with the *Six3* domain extending anterior to that of *Otx* [42]. The posterior limit of the *Otx* domain abuts that of *Gbx* in the first larval segment, while the anterior boundaries of *Hox1* and *Hox4* are in the second and third larval segments [48]. *Six3/6* and *Otx* are similarly expressed in acoel flatworms [26,83,84], and in *D. melanogaster* all three genes are expressed in similar patterns as in *P. dumerilii*. Therefore, the annelid cerebral ganglion has been homologized with the insect protocerebrum. In addition, similar to the CNS in *Drosophila* and vertebrates, the neuroectoderm in *P. dumerilii* is divided into a series of domains with outer/dorsal expression of *Msx* and *Pax3/7* (*gooseberry*), intermediate expression of *Nk6* and *Pax6*, and medial expression of *Nkx2.1*/*Nkx2.2* [1].

Together with anatomical similarities, these data showing distinct similarities in expression of genes patterning the CNS both anteriorly/posteriorly and medio/laterally between both major lineages of protostomes (Ecdysozoa and Lophotrochozoa) and vertebrates support a single origin of the CNS in the bilaterian ancestor (Figure 2). The counterargument would be that the CNS in protostomes evolved independently coopting A/P and D/V patterning mechanisms from an ancestor that used them to pattern a body axis. However, this would mean that the extensive similarities in neuronal architecture between chordates, arthropods and annelids would have been convergently evolved, which seems most unlikely.

Was there a dorso/ventral inversion, and if so, when did it occur?

If the CNS evolved just once, then a D/V inversion must have occurred during evolution of either protostomes or deuterostomes (reviewed in [27,85]). At present, the chief theories are as follows. The first is Anton Dohrn's idea that a D/V inversion occurred either at the base of the protostomes or within the deuterostomes [86]. The second is the idea most recently articulated by John Gerhart, Christopher Lowe and colleagues that the ancestral deuterostome was hemichordate-like with dorsal and ventral nerve cords and an ectodermal nerve net, with the chordate CNS arising directly from the nerve net [9,24,27] or alternatively, as proposed by van Wijhe [87] and more recently by Nomaksteinsky and colleagues [88], from the dorsal nerve cord. A third theory that the ancestral bilaterian had multiple nerve cords, with one evolving into the protostome CNS and another into the deuterostome CNS, was suggested by Gerhart [27] but has received little attention.

Major evidence supporting a D/V inversion in either basal deuterostomes, basal protostomes or basal chordates is that genes involved specifying polarity of the D/V body axis are expressed in opposite orientations in protostomes and chordates. Sasai, de Robertis and colleagues found that in both groups, BMP signaling is involved in establishing D/V polarity and in neural specification, with suppression of BMP signaling being a prerequisite for formation of a CNS [22,23,89]. In agreement with a D/V inversion having occurred in either the deuterostome or protostome lineages [23], *BMPs* are expressed dorsally in protostomes and hemichordates and the *BMP* antagonist *short gastrulation* (= *chordin* in deuterostomes) is expressed ventrally, while in chordates it is the opposite – *BMPs* are expressed ventrally and *chordin* dorsally. In most bilaterians, D/V orientation of the body and position of the nerve cord are coupled; however, Hejnol and Martindale have noted that expression of *BMP2/4* dorsally (opposite the future mouth) in an acoel with neurite tract(s) dorsally as well as laterally [26] supports the idea that a role for *BMP*/*chordin* in axial patterning may have preceded a role in neural patterning. Another line of evidence supporting D/V inversion comes from analysis of genes involved in left–right patterning. For example, two key regulators of this distinction, *Nodal* and *Pitx*, are expressed on the left side of chordates, but on the right in echinoderms and in some molluscs [90,91].

In summary, conserved expression of some genes along the longitudinal axis of cnidarians and in the CNS and general ectoderm of bilaterians indicates likely cooption of roles for these genes in patterning the CNS. However, similar expression of genes involved in both D/V patterning of the CNS and in A/P regionalization of the brain in chordates and protostomes together with neuroanatomical parallels provides considerable support for the idea that the bilaterian ancestor had a CNS, which was modified or possibly lost in various protostome and deuterostome lineages.

Hemichordate theories

Despite considerable evidence in support of a single origin of the CNS, data from hemichordates have been interpreted as indicating that the ancestral deuterostome had a nerve net, and therefore the CNSs in chordates and protostomes evolved independently. Hemichordates and echinoderms form a clade, the Ambulacraria, which

branched off the deuterostome tree as a sister group to chordates. Indirect developing members of both groups have similar pelagic larvae with an apical tuft of cilia. Echinoderms, which have pentamerous symmetry, typically have an ectodermal nerve net plus radial nerves and a circumoral nerve ring. While Haag proposed that the sea urchin radial nerves are homologous to the chordate CNS [33], most authors disagree [92,93]. Importantly, echinoderm nerve cords do not express *Hox* genes [51,94,95], and an extensive screen by Sly and colleagues for expression of neural patterning genes in a juvenile sea urchin failed to find evidence that the nerve ring or radial nerves are homologous to any part of the brain or nerve cord in bilaterians [93]. Moreover, *Engrailed* is very broadly expressed in the nervous systems and other tissues of the juvenile starfish and not in localized domains as in the chordate CNS [94]. Therefore, echinoderms are currently not considered relevant to the question of evolution of chordate nerve chords. In contrast, the worm-like enteropneust hemichordates, which have longitudinal nerve cords as well as a nerve net, have figured prominently in discussions of the evolution of chordates [10,96-98].

Inferring homologies between chordate nerve cords and hemichordate nervous tissues has been complicated by large differences in morphology. Adult hemichordates have three distinct regions: proboscis, collar and trunk. There are two classes of hemichordates – enteropneusts and pterobranchs. All four families of enteropneusts (Harrimaniidae, Spengelidae, Ptychoderidae, Torquaratoridae) have an ectodermal nerve net, located in all three regions, plus dorsal and ventral nerve cords, suggesting that this organization is a basal hemichordate characteristic. In contrast, the sessile pterobranchs have anterior tentacles and a concentration of neurons at the base of the tentacles that has been termed a brain, as well as several concentrations of neurites and associated neurons extending into the tentacles, the stalk and between the gill slits [99]. Although Romer and others argued that pterobranchs were basal hemichordates [100,101], recent molecular phylogenetic analyses do not distinguish which family is basal [102,103], leaving open the possibility that pterobranchs are derived. Indeed, fossil tube-dwelling enteropneusts from the Cambrian were recently discovered [104].

Most of the work on neural development in hemichordates concerns indirectly developing ptychoderids and the direct developing harrimaniid *Saccoglossus kowalevskii* (reviewed by Röttinger and Lowe [105]). Miyamoto and colleagues showed that the larval nervous system in indirect ptychoderids does not carry over into the adult; in late larvae, the larval nervous system is gradually replaced by the adult one [106]. Therefore, it is the development of the adult nervous system that is pertinent for understanding evolution of the CNS.

Hemichordates and the argument of an ectodermal nerve net versus a CNS: theory one

There are two competing theories concerning the evolutionary relationship between the nerve net and nerve cords of hemichordates and the chordate CNS (Figure 1). One theory, most recently articulated by Kaul and Stach [107], proposes that one of the hemichordate nerve cords, typically the dorsal one, is homologous to the chordate CNS. This theory implies that the ancestral deuterostome and perhaps also the ancestral bilaterian had a CNS. The chief basis for this idea is that the collar nerve cord neurulates, suggestive of neurulation in vertebrates [88,108]. Anterior and posterior to the collar, the nerve cord is continued by basiepithelial tracts of neurites, which are concentrated dorsally [99]. However, there is nothing that resembles a brain. In the direct developing S. *kowalevskii*, neurulation in the collar nerve cord progresses from posterior to anterior, and there are posterior and anterior neuropores [107]. The nerve cord continues posteriorly as a superficial tract of nerve cell bodies overlying nerve cell fibers and rostral of the anterior neuropore as a wide, superficial tract of both neurons and nerve fibers [88]. In addition to the longitudinal nerve cords, there is a peribranchial nerve ring, which develops from ventral to dorsal, as well as a collar nerve ring at the collar–trunk boundary. Although initially neither nerve cord was thought to contain nerve cell bodies, studies with electron microscopy and with specific nerve cell markers have demonstrated nerve cell bodies and glia in the dorsal nerve cord and at least some neurons associated with the ventral one [107,109,110]. Ventrally in the dorsal cord, there is a neuropil. Bullock [111] and Brown and colleagues [110] have suggested that the large neurons may be homologous to Mauthner cells of the lamprey and Rhode cells of amphioxus. The developing collar and ventral nerve cords as well as the peripharyngeal cord of both *Ptychodera flava* and *S. kowalevskii* express nerve cell-specific genes including *Elav*, *synaptogamin* and also genes for peptides and proteins specific for subsets of nerve cells including *VAChT*, *serotonin*, *Hb9*, *Drg11* and *GABA*. Serotonergic neurons are restricted to the peripheral nervous system, while those labeling with *Drg11*, *Hb9* and cholinergic neurons are preferentially in the collar nerve cord [88,112].

Although most of the ptychoderid ectoderm is non-neural [88], basiepithelial nerve cells are moderately numerous in the proboscis [106]. Nomaksteinsky and colleagues suggested that the more even distribution of neurons in the basiepithelial nerve net of *S. kowalevskii* might represent a transient larval nervous system and

that part of the diffuse nervous system of developing *S. kowalevskii* larvae, especially in the proboscis, will become the peripheral nervous system [88]. Based on the morphological and gene expression data, they concluded that it was 'implausible that the enteropneust skin is homologous to the chordate CNS'. Instead, they argued that the relatively few neurons in the adult "non-neural" ectoderm constituted a peripheral nervous system, and that either the dorsal or ventral nerve cord (they could not decide which one) was homologous to that of chordates [88]. A comprehensive study of developmental gene expression in the developing nerve cords of hemichordates is sorely needed; to date only a few pictures showing expression of genes including *Dlx*, several *Hox* genes, *Tbx2/3*, *PoxN*, *Pitx* and *Olig* in the dorsal and/or ventral midline of embryos of *S. kowalevskii* have been published [9,24,25,113], but whether the tissue expressing these genes is the developing nerve cord or overlying ectoderm is not clear. There are no studies of developmental gene expression in indirectly developing species such as *P. flava* due to a long pelagic larval period [114,115].

A problem with homologizing the dorsal nerve cord of hemichordates with the chordate CNS is the finding by Lowe and colleagues [113] that, as in protostomes, *BMP2-4* and *BMP5-8* are expressed dorsally in *S. kowalevskii*, while *chordin* is expressed ventrally. Consistent with a role in D/V patterning, excess BMP4 protein radializes the embryos and eliminates *chordin* expression, indicating an evolutionarily conserved role of BMP in D/V patterning. This suggests that if either nerve cord in hemichordates is homologous to the chordate CNS, it is the ventral nerve cord, which does not neurulate, and a D/V inversion occurred at the base of the chordates. This is consistent with the gill slits being dorsal and the stomochord, a dorsal/anterior extension of the gut, having been shown to be unrelated to the notochord [116]. Confusing the issue further, in amphioxus and vertebrates, *Nodal* expression dorsally acts in opposition to *BMP* expression ventrally [117], while in sea urchin embryos, BMPs and Nodal oppose each other in patterning the oral/aboral axis (Nodal ventralizes; BMP dorsalizes), suggesting that a role for Nodal in opposing BMPs was present at the base of the deuterostomes and that a D/V inversion occurred in chordates. To some extent this is similar in *S. kowalevskii*, in that perturbation of Nodal signaling results in D/V patterning defects [118]. However, treatment with the Nodal inhibitor SB431542 eliminates expression of both *BMP2/4* and *chordin* and anteriorizes embryos, indicating that Nodal posteriorizes embryos, the opposite of the situation in chordates [118]. These results suggest that the role of Nodal [119] may have been altered in *S. kowalevskii*. Whether a role for BMP/Nodal opposition in D/V patterning was present in the ancestral bilaterian is uncertain. Nodal is involved in left/right patterning in a mollusk [120], but possible roles in D/V or A/P patterning in protostomes have apparently not been investigated. In summary, since it neurulates, the dorsal nerve cord of hemichordates has been proposed as homologous to the chordate CNS. However, the rather scanty data on gene expression are more compatible with homology of the ventral nerve cord and the chordate CNS. More data are clearly needed on both anatomy and gene expression in the hemichordate nerve cords.

Hemichordates and the argument of an ectodermal nerve net versus a CNS: theory two

In spite of the evidence supporting the idea that the ancestral bilaterian had a CNS, there is an alternative theory – namely that the ancestral bilaterian and the ancestral deuterostome had ectodermal nerve nets, from which the chordate CNS evolved. The dorsal and ventral nerve cords of hemichordates are therefore not only a hemichordate invention, but are unrelated to the chordate CNS (reviewed in [34]). This theory, most recently articulated by Lowe and colleagues [9,24,113], is based on developmental gene expression and gene interactions in the direct-developing hemichordate, *S. kowalevskii*. It proposes that the chordate CNS evolved from the ectodermal nerve net of a hemichordate-like ancestral deuterostome and maintains that the hemichordate nerve net contains signaling centers evolutionarily related to the anterior neural ridge (ANR), the ZLI and the isthmic organizer (ISO) in vertebrates. As a corollary, part or all of these signaling centers have been lost in the invertebrate chordates (amphioxus and tunicates) [9]. This idea deserves careful consideration because it not only argues that the considerable similarities of gene expression in protostome and chordate nerve cords represent convergent evolution, but it assigns a key position to hemichordates in evolution of the vertebrate CNS.

de Beer [121] was one of the first to recognize the hierarchical nature of homology when he noted similar morphological features in two different animals could develop under the control of different genes; a phenomenon now known as genetic piracy [122]. The converse is also known – where parts of homologous gene networks are involved in the development of apparently nonhomologous structures [123]. When such disconnects are discovered, some would pay more attention to structure [124], and others would pay more attention to the genes (as a deep homology) [125]. In discussing neural evolution in higher deuterostomes, Lowe and colleagues strongly favor genes over morphological features as arbiters of homology. Thus, they maintain that structures with very different morphology (for example, the proboscis of a hemichordate and the forebrain of a

vertebrate) are not morphologically homologous [25,126] even though conserved gene expression patterns may indicate a common evolutionary origin. Thus, while those authors find that homologous genetic programs operate at the anterior end of the hemichordate proboscis and vertebrate ANR, at the boundary between the hemichordate proboscis and collar and at the vertebrate ZLI, and at the boundary between the hemichordate collar and trunk and at the vertebrate ISO located at the MHB, they do not refer to these three regions of the hemichordate ectoderm and vertebrate brain as homologs [9]. Even so, as discussed by Wagner [16], these regions could be homologous even though morphologically divergent.

Lowe and colleagues, in their series of papers concerning genes and structures involved in neural evolution of deuterostomes, have been somewhat inconsistent in their treatment of the subject of homology. In their initial work, which concerned expression of 22 genes with restricted ectodermal domains along the A/P axis (Table 1), they concluded that the surface ectoderm of *S. kowalevskii* and the chordate CNS have a common ancestry [24]. These authors noted that *S. kowalevskii* 'shows pervasive neurogenesis with no large, contiguous non-neurogenic subregion, as occurs in chordates' and concluded that the deuterostome ancestor had a diffuse ectodermal nerve net that evolved into the vertebrate CNS [24]. One difficulty with this argument is that, as Aronowicz and Lowe [25] later noted, the surface ectoderm of invertebrate chordates outside the neural tube contains widespread ectodermal sensory neurons (Figure 1) [127,128], while in vertebrates the large panplacodal region outside the neural plate is highly neurogenic [68]. In addition, although the authors maintained that most of 22 genes they studied are not expressed in

Table 1 Gene expression in nervous tissues of -*Saccoglossus kowalevskii*, *Branchiostoma floridae*, *Ciona intestinalis* and vertebrates

Gene	*S. kowalevskii*		*B. floridae*		*C. intestinalis*		Vertebrate		References
	s.e.	CNC[a]	s.e.	CNS	Ectoderm	CNS	Placode	CNS	
SoxB	+	+	+	+	+	+	+	+	[131]
Hu/Elav	+	+	+	+	?	?	+	+	
Nrp/Musashi	+	+	+	+	?	+	+	+	[141]
Vax	+	–	?	?	gene	lost	–	retina, forebrain	[142]
Rx	+	–	–	+	–	+	–	+	[69,76]
Six3	+	+/–	+	+	–	+	–	+	[43,131,143]
Nkx2.1	+	–	–	+	?	?	–	+	[140,144]
Bf-1	+	+/–	+	+	+	?	+	+	[131,145,146]
Dlx	+	+	+	+	+	+	+	+	[77]
Pax6	+	+	+	+	–	+	+	+	[75]
Tll/Tlx	+	+/–	+	+	?	?	+	+	[147,148]
BarH	+	+	?	?	?	?	+	+	[149]
Emx	+	+	?	?	?	?	+	+	[150]
Otp	+	?	?	?	+	+	?	+	[131,151]
Dbx	+	+	?	?	?	?	–	+	[152]
Lim1/5	+	+	+	+	?	?	+	+	[153]
Irx	+	+	+	+	?	?	+	+	[154]
Otx	+	+	–	+	+	+	+	+	[78,137]
En	+	+	+	+	–	+	+ lamprey	+	[78,155]
Gbx	+	+	+	+	gene	lost	+	+	[142]
Hox1	+	+	+	+	+	+	+	+	[156-158]
Hox3	+	–	+	+	+	+	+	+	[134,156,158]
Hox4	+	?	+	+	–	–	–	+	[156]
Hox7/8	+	?	?	?	gene	Lost	–	+	[156]
Hox11/13	+	–	?	?	Hox12+	Hox12+	Hox11+	+	[156,159]

Gene expression in ectoderm and nervous tissue of the hemichordate *Saccoglossus kowalevskii*, the cephalochordate *Branchiostoma floridae*, the tunicate *Ciona intestinalis* and vertebrates. Note: En expression in placodes only documented for lamprey among the vertebrates. CNC, collar nerve cord; CNS, central nervous system; s.e., surface ectoderm. [a]In the absence of sections, which would distinguish between expression in the surface ectoderm and in the collar nerve cord, all genes with ectodermal expression in the region of the dorsal portion of the collar are listed.

the "epidermal ectoderm" in chordates [24], in the 10 years since this paper was published, it has become clear that all except possibly *Vax*, *Rx* and *Nkx2-1* are expressed in ectodermal sensory cells in chordates (Table 1) [68,129-131]; expression of *Vax* in amphioxus and of *Nkx2-1* in the tunicate *Ciona intestinalis* have not been determined. For example, *Sox1/2/3* (*SoxB*) and *Hu/Elav* are expressed in ectodermal sensory cells in chordates [129,130], *Hox1*, *Hox3* and *Hox4* are expressed in nested patterns in the amphioxus ectoderm and at especially high levels in ectodermal sensory cells [132], ascidian *Hox1* is expressed in development of ectodermal sensory cells [133], and some *Hox* genes are expressed in placode derivatives in vertebrates [134]. In addition, *Gbx* is expressed in the "non-neural" ectoderm in amphioxus [135] and in developing placodes in vertebrates [136], while *Gbx* and *Otx2* mutually repress one another in development of the otic placode as they do in A/P patterning of the CNS [137]. Therefore, as noted above, the ectodermal expression patterns of genes along the A/P axis do not clearly distinguish CNS from ectoderm outside the CNS. Indeed, Aronowicz and Lowe noted that genes such as *Otx* and *Hox* appear to be involved in patterning neurogenic tissues generally [25]. Thus, expression of these A/P patterning genes alone does not distinguish between theories that the ancestral bilaterian and ancestral deuterostome had an ectodermal nerve net or whether they both had a CNS. As noted above, genes expressed exclusively in the CNS of bilaterians (for example, *earmuff/Fezf/Fezl* and D/V patterning genes) are generally more informative for inferring the course of evolution of the CNS [64]. Indeed, domains of homologs of three genes mediating lateral to medial (D/V) patterning in the chordate and protostome CNSs (*Hh*, *Nkx2.2*, *Msx*) [138-140] are not expressed in comparable patterns in the hemichordate ectoderm [113]. Thus, if the chordate CNS evolved from a nerve net in the ancestral deuterostome, expression of these lateral to medial genes would represent convergent evolution in the CNSs of protostomes and chordates.

The similarities in patterning the ANR and the anterior part of the hemichordate proboscis, the ZLI and the boundary between the proboscis and collar, and the ISO and the boundary between the collar and trunk lead to the conclusion that in the ancestral deuterostome the role for these signaling centers, which ultimately gave rise to the vertebrate ANR, ZLI and ISO, was to regionalize the general body plan [9]. However, Pani and colleagues also proposed that amphioxus partially lost the ANR, and completely lost both the ZLI and ISO, while the tunicate *C. intestinalis* lost the ZLI and partially lost both the ANR and ISO [9]. They conclude that in 'certain cases hemichordates will be a more informative group than basal chordates for reconstructing stem chordate characters and understanding the origins of vertebrate developmental genetic processes' [9]. As this is quite an extreme view and is at odds with conclusions based on morphology, gene expression and gene function in amphioxus, tunicates and vertebrates, the evidence merits close examination.

How much of the ANR gene network is present in S. kowalevskii? Is the evidence sufficient that the vertebrate ANR evolved from the anteriormost ectoderm of the ancestral deuterostome?

The vertebrate ANR is characterized by expression of *Fgf8*, *Six3*, *Pax6*, *Otx2*, *Sox2* with *Dlx5* expressed in adjacent non-neural ectoderm and *Bf-1* (*FoxG-1*) expressed in the rostral forebrain (Figure 4) [160], while *Sfrp1a* is expressed in the anterior/ventral part of the developing neural tube [161]. In addition, transplantation of the ANR laterally expanded the telencephalon and promoted expression of *Bf-1* (*FoxG1*), demonstrating that the ANR is an organizer [162]. In *S. kowalevskii*, homologs of these genes are expressed in the proboscis ectoderm, although their relative domains are rather divergent from those of their homologs in the vertebrate forebrain. For example, *Six3*, *Sfrp1/5* and *Fgf8/17/18* are strongly expressed in the anterior proboscis ectoderm, and *Sox1/2/3* is broadly expressed in the proboscis and anterior part of the collar [9]. However, *Pax6* is not expressed at all in the anteriormost proboscis ectoderm but more posteriorly throughout much of the proboscis and collar, while *Otx* is chiefly expressed in the collar ectoderm with only very weak expression at the anterior end of the proboscis. In addition, *Dlx* is very patchily expressed in the proboscis ectoderm, but appears to be highly expressed where the dorsal nerve cord will form. *Hh* is expressed in ectoderm at the tip of the proboscis in *S. kowalevskii* [9], while in vertebrates it is expressed in the basal plate of the forebrain and midbrain and in the floor plate of the anterior hindbrain as well as in the prechordal plate [163].

Experimental evidence for homology of gene networks patterning the ANR and anterior proboscis ectoderm in *S. kowalevskii* came from manipulation of Fgf and Hh signaling. Although inhibition of *Fgf* or Fgf signaling scarcely affected expression of the proboscis domain of the anterior marker *Rx*, it did eliminate *FoxG1* (*BF-1*) expression in the proboscis. In addition, knockdown of *Hh* eliminated expression of the anterior marker *Fgf-Sk1*. Thus, inhibition of Fgf or Hh signaling affects development of the proboscis and expression of some anterior markers. However, Green and colleagues found that the major role of *Fgf8/17/18* in development appears to be in mesoderm induction [164]. Therefore, the effects of inhibition of Fgf signaling on anterior development could be secondary to those on mesoderm.

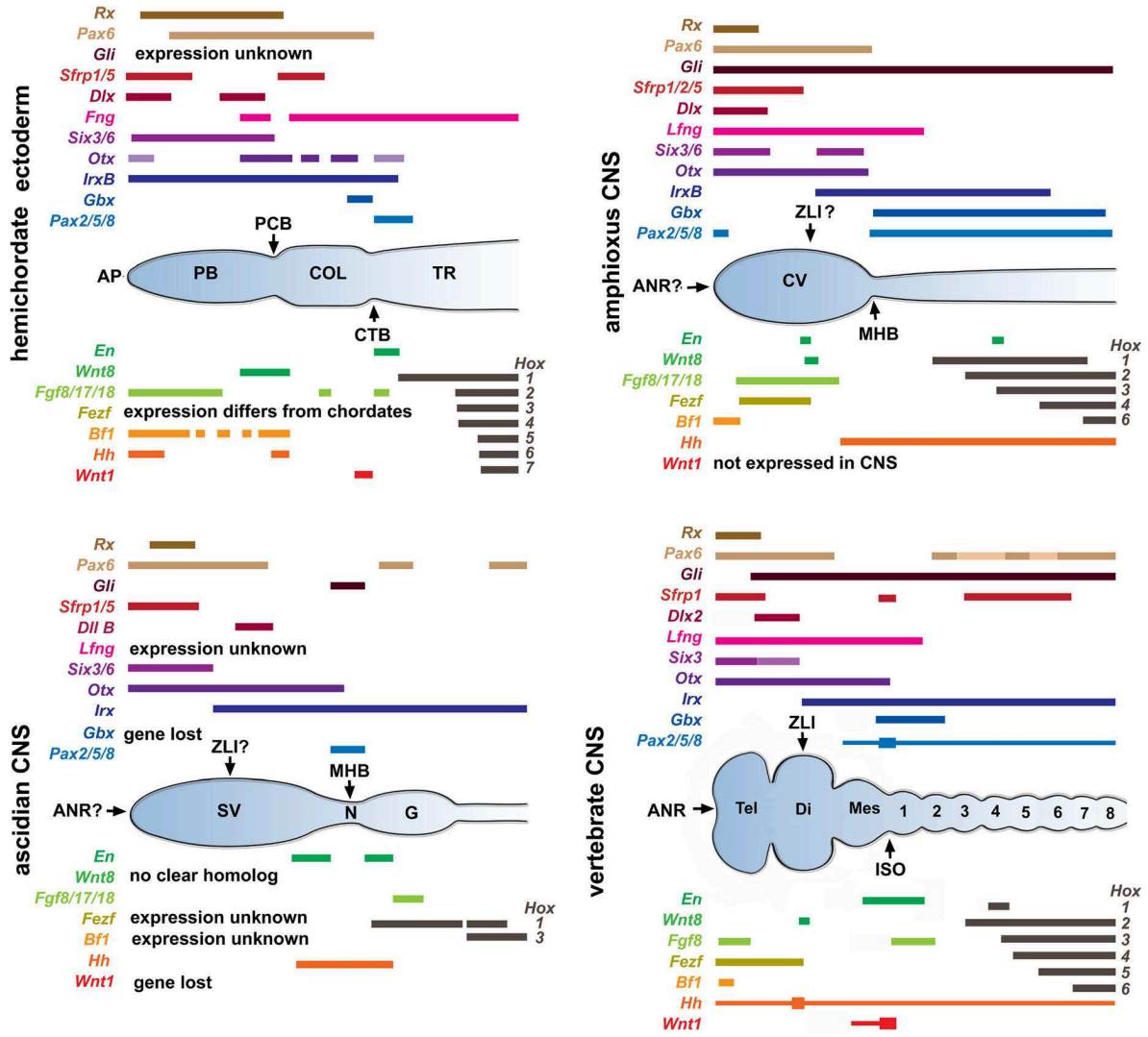


Figure 4 Anterior–posterior gene expression in hemichordate ectoderm and central nervous systems of three chordate subphyla. Anterior–posterior regionalization of gene expression domains in the ectoderm of the hemichordate *Saccoglossus kowalevskii* as well as in the central nervous system (CNS) of representatives of the three chordate subphyla (that is, amphioxus, ascidian tunicates and vertebrates). Question marks on the diagrams of the amphioxus and ascidian CNS indicate that organizer properties of the regions marked 'ZLI?' and 'ANR?' have not been tested. AP, anterior proboscis; PB, proboscis; PCB, proboscis/collar boundary; COL, collar; CTB, collar/trunk boundary; TR, trunk; ANR, anterior neural ridge; ZLI, zona limitans intrathalamica; MHB, midbrain/hindbrain boundary; SV, sensory vesicle; N, neck; G, ganglion; ISO, isthmic organizer; Tel, telencephalon; Di, diencephalon; Mes, mesencephalon. Gene expression domains based on [9,24,34,64,68-72,74-80].

Whether this evidence is sufficient to indicate that the vertebrate ANR evolved from the anterior part of a diffuse ectodermal nerve net in the deuterostome ancestor is open to question. A major difficulty in identifying homologous gene networks is deciding how much of two gene networks must be conserved for them to be considered homologous [165]. This is a particular problem when the morphology is not conserved. Gene networks can include several thousand genes, and it is well known that core parts of signaling pathways are often coopted for patterning nonhomologous structures [166]. Thus, there are several alternative explanations for similarities in gene expression between the hemichordate proboscis ectoderm and the vertebrate ANR. One is that similar expression of a suite of genes including *Fgfs* in these two regions may simply be indicative of ancient roles in specification of the anterior end of embryos in general. For example, Sinigaglia and colleagues reported that *Six3/6*, *FoxQ2*, *Irx*, *SoxB1* and *Fgf* are expressed in the aboral region in the cnidarian *N. vectensis*, while *Wnts* are expressed at the opposite end of the embryo [39]. *Six3/6* is initially involved in specification of the anterior end of the embryo and later in neurogenesis. Fgf signaling is required for development of the apical tuft

of cilia. Therefore, roles for these genes in patterning the hemichordate proboscis may simply reflect an inheritance from prebilaterians. Another possibility is that if, as comparisons of protostomes and chordates indicate, these genes were coopted for patterning the anterior CNS of the ancestral bilaterian, then the ancestral deuterostome might have had a more extensive CNS than modern hemichordates. If so, the forebrain may have been lost as the proboscis evolved, and the genetic pathways for anterior neural development coopted into the proboscis ectoderm. Expression of *Rx*, *Fgf8/17/18* and other genes in the hemichordate proboscis is compatible with either theory. In chordates, *Rx* expression is apparently restricted to the anterior end of the forebrain (it is not expressed in sensory ectoderm) [69], while *Fgf8/17/18* is expressed in the ANR and telencephalon of vertebrates as well as throughout the forebrain of amphioxus. In contrast, *Hh* is expressed in the anterior tip of the notochord in amphioxus and in the floorplate but not in the anterior CNS. Given that *Fgf8* and *Hh* have been coopted in vertebrate limbs, which evolved in gnathostomes, for patterning the apical ectodermal ridge (AER) and zone of polarizing activity (ZPA) [167], it is certainly possible that these genes plus *Rx* and other genes involved in neuronal specification may have been coopted for patterning the proboscis ectoderm of hemichordates.

Another possibility, proposed by Nomaksteinsky and colleagues [88], is that the ectoderm of developing *S. kowalevskii* may represent a transient larval nervous system unrelated to that of the adult. Evidence is that *SoxB* genes, which mark the neural plate in chordates, are broadly expressed in the ectoderm of indirectly developing embryos of the hemichordate *Ptychodera flava*, with expression becoming localized during development to the stomodaeum, ciliary bands and apical ciliary tuft [114]. *Dlx* is expressed in the aboral ectoderm, with a concentration towards the apical tuft, and *Fz5/8* and *FoxQ2* are expressed in the apical tuft [168]. Similarly in *S. kowalevskii*, *Six3, Fz5/8* and *Dlx* are broadly expressed in the proboscis ectoderm and *Bf-1* (*FoxG1*) in the apical region [9,24]. The tornaria larva of indirect developing hemichordates has much in common with the dipleurula type of echinoderm larvae, which also have an apical tuft and a band of cilia around the mouth. Moreover, gene expression in the apical tuft and surrounding ectoderm is highly conserved in the sea urchin and hemichordate larvae, and there is considerable similarity with expression in the proboscis ectoderm of *S. kowalevskii* embryos [169]. In sea urchins, *FoxQ2*, *Fzl5/8*, *Sfrp1/5*, *Dkk3* and *Six3* are expressed at the apical end of the larva, with Wnt signaling acting to restrict expression of these genes to the anterior end of the larva [169,170]. Similarly, in *S. kowalevskii* knocking down the Wnt receptor *Fz5/8* results in a posterior expansion of the domains of anterior markers [9], indicating that Wnt signaling is also regulating expression of genes in the proboscis ectoderm as it is in the apical ectoderm of sea urchin larvae.

How much of the ZLI gene network is present in S. kowalevskii Does the evidence support evolution of the vertebrate ZLI from a boundary partitioning the anterior/posterior axis of the ancestral deuterostome?

The vertebrate ZLI is positioned and regulated, at least in part, by anterior expression of *Otx* and *Fezf* abutting a posterior domain of *Irx* [171,172]. Organizer properties are conferred by *Hh*, which is expressed at the ZLI and which is regulated by *Wnt8b* (Figure 4) [173]. In fish, knockdown of *Fezf2*, which is expressed anterior to *Irx*, eliminates the prethalamus and causes mis-specification of the ZLI. Correspondingly, knockdown of *Otx* inhibits expression of *Shh* at the ZLI and reduces expression of *Ptc1* and *Wnt8b* [172]. Evidence for organizer properties of the vertebrate ZLI comes from grafting experiments and implants of Fgf8-coated beads, which induce the rostral neuroepithelium to develop an ectopic and polarized mesencephalon/metencephalon (reviewed in [174]).

The region of the hemichordate ectoderm expressing homologs of ZLI genes lies at the boundary between the proboscis and the collar [9] and is marked by the posterior limit of *Rx* expression and bands of *Fgf8/17/18*, *FoxG*, *Otx*, *Wnt8*, *Hh*, *Ptch*, *FoxA* and *Dlx*. Congruent expression of *Wnt8* and *Hh* in *S. kowalevskii* is consistent with a possible role of *Wnt8* regulating *Hh* as it does at the ZLI in vertebrates. However, domains of *Fezf* and *Irx*, which abut at the ZLI in vertebrates, do not abut at the proboscis/collar boundary in *S. kowalevskii*. Expression of *Fezf* was not shown, but the major domain of *Irx* appears to be congruent with that of *engrailed* in the posterior part of the collar [24].

For *S. kowalevskii*, the experimental evidence presented by Pani and colleagues for the proboscis/collar boundary region acting as a signaling center was that knockdown of *Otx* expression reduced the intensity of the stripe of *Hh* expression just posterior to the proboscis/collar boundary, while inhibition of Hh signaling downregulated *Dlx* both at the proboscis/collar boundary and more anteriorly and also reduced the size of the proboscis [9]. These results show that *Otx* may act upstream of *Hh* at the proboscis/collar boundary, while *Hh* is vital for normal development. However, *Otx* and *Shh* also interact in patterning the vertebrate midbrain, not just the ZLI, and knockdown of *Otx* results in dorsal expansion of the *Shh* domain and a dorsal and anterior rotation of the MHB [175]. Thus, the relationship between *Otx* and *Hh* and perhaps some other parts of the gene network may be evolutionarily conserved between *S. kowalevskii* and vertebrates, but whether tissue at the

proboscis/collar boundary in an ancestral deuterostome evolved into the ZLI is open to question given the markedly different expression of key genes such as *Fezf* and *Irx* as well as the lack of anatomical similarity.

How much of the ISO gene network is present in S. kowalevskii Does the evidence support evolution of the vertebrate ISO from a boundary partitioning the anterior/posterior axis of the ancestral deuterostome?

In the vertebrate CNS, the MHB or isthmus functions as an organizer, and is therefore termed the isthmic organizer or ISO. The MHB is positioned by opposition of anterior *Otx* and posterior *Gbx*. Organizer properties are conferred by the action of a suite of genes including *En*, *Fgf8/17/18*, *Wnt1* and *Pax2/5/8* (Figure 4) [176]. *Otx* and *Fgf8* are expressed in the midbrain and *Gbx* in the anterior hindbrain, with the *Wnt1*, *Engrailed* and *Pax2/5/8* domains spanning the boundary; mutual repression of *Otx* and *Gbx* positions the ISO, and Wnt signaling is required for expression of *Engrailed*. Pani and colleagues also presented evidence for a homologous gene network in *S. kowalevskii* operating at the boundary between the collar and trunk [9]. *Engrailed*, *Fgf8/17/18*, *Wnt1* and *Gbx* are expressed near this boundary. In addition, clonal suppression of *β-catenin* effectively inhibited *Engrailed* expression, suggesting that Wnt signaling regulates *Engrailed*, as it does at the ISO in vertebrates, while suppression of *Fgf8/17/18* reduced *Engrailed* expression at the collar/trunk boundary. However, despite these similarities, there are several problems in interpreting the gene network at the collar trunk boundary in *S. kowalevskii* as homologous to that patterning the vertebrate ISO. Importantly, as the authors note, the 'spatial arrangements of *Otx* and *Wnt1*, and *Gbx* and *Fgf8/17/18* are reversed in *S. kowalevskii* compared to the ISO in vertebrates' [9]. Moreover, suppression of Fgf signaling had no effect on expression of *Pax2/5/8*. A problem in interpreting these experiments is that in vertebrates, *Fgf8* has an early role in induction of neural tissue, while patterning the ISO is a relatively late role. In *S. kowalevskii, Fgf8/17/18* is expressed from the blastula stage, and by the neurula stage expression is restricted to anterior ectoderm. Knockdown experiments show that *Fgf8/17/18* is required for mesoderm induction [164]. Therefore, effects of *Fgf8/17/18* inhibition in *S. kowalevskii* on *Engrailed* expression may be secondary to the loss of mesoderm and not directly related to effects on A/P patterning. Additional evidence for the common ancestry of the collar/trunk boundary and the vertebrate ISO was that *Hox* genes are expressed in nested patterns in the trunk ectoderm. However, it appears that at least *Hox1* and perhaps *Hox4* are also expressed in the dorsal part of the collar (Figure five in [24]), indicating that Hox expression in the ectoderm is probably not congruent with that in the collar, and raising the question of the genetic mechanisms patterning the collar nerve cord and how they might compare with those in chordates.

As for the ANR and ZLI, drawbacks to interpreting the boundary of the collar and trunk in hemichordates with the vertebrate ISO as having common ancestry not only include differences in expression of key genes expressed in these regions and also the general phenomenon of cooption of parts of gene networks for new functions. For example, in the AER of the vertebrate limb bud, *Wnt3a* induces expression of *Fgfs*, while ectopic *Engrailed* (*En-1*) induces ectopic *Fgf8* expression [177,178]. Moreover, in amphioxus, *engrailed* is coexpressed with *Wnt8*, but not at the MHB, suggesting that coexpression of *Wnt* and *engrailed* is not wedded to the MHB. The questions therefore become: how conserved must gene networks be in order for one to be reasonably certain that two morphologically rather different structures have a common ancestry; and if they do share an ancestry, can one distinguish whether portions of the gene networks operating in the CNS of a deuterostome ancestor were transferred to the hemichordate ectoderm as the hemichordate CNS became reduced or whether the ancestral deuterostome lacked a CNS and used these gene networks to partition the A/P axis?

How much of the ANR, ZLI and ISO do invertebrate chordates have?

In addition to asserting that the *S. kowalevskii* ectoderm is not only homologous to the vertebrate CNS but also has homologs of the ANR, ZLI and ISO, Pani and colleagues maintain that the invertebrate chordates, amphioxus and tunicates, have lost all or part of these three regions [9]. Here we draw attention to data that are not consistent with such a view (Figure 4).

Although transplantation experiments in vertebrates demonstrated that the ANR functions as an organizer [163], such transplantation experiments are not feasible for hemichordate, amphioxus or *C. intestinalis* embryos due to their small size. However, it is clear that much of the gene network for specification of the ANR and conferring organizer properties upon it is present in amphioxus. *Fgf8/17/18* is expressed in the entire forebrain of amphioxus, and *Bf-1* (*FoxG1*) is expressed at the tip of the forebrain [145] as are *Pax2/5/8* and *Six3/6* [43]. *Otx* and *Pax6* are expressed in comparable patterns with strong expression in the anterior forebrain [135,179], and the Wnt antagonist *Sfrp1/2/5* is expressed in the anteriormost dorsal ectoderm, including the most anterior neuroectoderm [180]. Moreover, as in vertebrates, *Dlx* is expressed in ectoderm outside the neural tube as well as in the edges of the anterior neural plate [70], while Hh is expressed in the underlying

tip of the notochord [181,182]. Taken together, these expression patterns indicate that amphioxus has most of the components for an ANR comparable to that in vertebrates.

Tunicates have undergone some radical changes to their genome and their anatomy in evolution. For example, the *Gbx* gene has been lost, and there are relatively few neurons – none in the tail nerve cord of ascidians. *Otx* is expressed anteriorly in the CNS, but the evidence for an ANR in *C. intestinalis* is relatively weak. In addition to *Otx*, *Pax6* is expressed in the anterior CNS, but not *Fgf8/17/18* and *Pax2/5/8*, *Hh*, or *Gli* [183]. Expression of *Bf-1* is not known. Therefore, it may be that as tunicates adopted early decision of cell fates and decreased the number of cells in the nerve cord, the need for an anterior brain organizer diminished.

For the ZLI, patterns of gene expression indicate that amphioxus probably has much of the genetic mechanism in place. In amphioxus, Irimia and colleagues showed that anterior expression of *Fezf* abuts posterior expression of *IrxB* about the midpoint of the forebrain [64] – approximately where *Wnt8* is expressed [184]. Similarly, *Wnt8b* expression at the ZLI in vertebrates is flanked by *Lfng*, while in amphioxus the *Fng* domain appears to be posterior to and possibly abutting that of the *Wnt8* domain at the late neurula stage [184,185]. *Dlx*, *Nkx2-2* and *Gli*, which mediates Hh signaling, are also expressed in this region, while *Fgf8/17/18* is expressed throughout the forebrain [70,71,186]. *Hh* is expressed in the floor plate as it is in vertebrates, but it is unclear at the early neurula stage whether it is expressed congruently with *Engrailed* and *Wnt8* or a little more posteriorly. Later expression is limited to a zone posterior to the forebrain [181]. Inhibition of Fgf signaling at the late blastula does not inhibit expression of neural plate markers, but eliminates that of *Otx* in the cerebral vesicle (forebrain) and reduces its size [187], indicating that Fgf signaling is essential for development of the forebrain. Similarly, upregulation of Wnt/β-catenin signaling reduces expression of *Otx* in the forebrain and eliminates expression of the anterior marker *FoxQ2* [188], showing that suppression of Wnt/β-catenin signaling by inhibitors such as Sfrp1/2/5 [180] is essential for forebrain development.

For ascidians, expression of *Irx* and *Fezf* is not known. *C. intestinalis* has two *Wnt* genes that are possibly related to *Wnt8* but which have long branches in phylogenetic analyses and for which assignation remain unclear [189]. Expression of these genes has not been characterized. However, *Gli*, *Hh* and *Fgf8/17/18* are not expressed in the anterior CNS. Even so, inhibition of Fgf signaling blocks expression of the anterior marker *Six3/6*, indicating that Fgf signaling is required for development of the anterior brain [190]. Thus, while *C. intestinalis* likely lacks most of the gene network specifying the vertebrate ZLI and conferring organizer properties, until the expression patterns of *Irx* and *Fezf* are known the presence of part of the network for specification of the ZLI cannot be ruled out.

Amphioxus also appears to have in place the genetic mechanism for positioning the ISO, but this region does not express homologs of the genes that confer organizer properties on the vertebrate ISO. In the amphioxus brain, as in that of vertebrates, the domain of *Otx* expression in the forebrain/midbrain abuts that of *Gbx* in the hindbrain [135]. In vertebrates, organizer properties are conferred on the ISO by expression of *Fgf8*, *Wnt1*, *Engrailed* and *Pax2/5/8* genes at this boundary. In amphioxus, *Wnt1* and *Engrailed* are not expressed between the *Otx* and *Gbx* domains, although *Fgf8/17/18* is expressed anterior to and abutting this boundary, while *Pax2/5/8* is expressed posterior to and abutting it [187,191,192]. *Engrailed* is expressed together with *Wnt8* approximately at the boundary between the *Fezf* and *IrxB* domains. Interestingly, there is also a stripe of *Engrailed* in the general ectoderm that is in register with that in the CNS [193].

The tunicate *C. intestinalis* appears to have more of the ISO network in place than amphioxus does, but it has lost a key gene – *Gbx*. *Pax2/5/8* and *Fgf8/17/18* are expressed in the neck region in a few cells just posterior to the posterior limit of *Otx*. *Engrailed* is expressed in two domains, one near the *Pax2/5/8* domain and the other overlapping with that of *Otx* [194].

The emerging picture is that amphioxus has most, if not all of the genetic mechanism for the ANR in place and part of that for the ZLI and ISO while the ancestral tunicate probably added to these gene networks but subsequently lost parts of them as the CNS became simplified. The most parsimonious explanation is that the ancestral bilaterian had a CNS in which portions of these gene networks were used to establish divisions in the CNS and perhaps to some extent in the ectoderm generally, and these gene networks became modified in the protostome and deuterostome lineages. Under this scenario, the basic gene networks for specifying the ANR, ZLI and MHB began to be established in the ancestral chordate or even earlier in the ancestral deuterostome or ancestral bilaterian, and additional genes were recruited to these networks before tunicates branched off the lineage leading to vertebrates. In the tunicate lineage, in conjunction with shrinking genomes, a switch from regulative to determinate development and a reduction of cell numbers in the CNS, some parts of the ancestral tunicate networks for these brain regions were lost. Modifications occurring in the hemichordate lineage may have involved changes in the CNS and/or cooption of additional genes and portions of gene networks to the surface ectoderm.

Conclusions

The bilaterian CNS typically develops from a longitudinal strip of ectoderm that is then internalized. Although the remainder of the ectoderm is usually termed "non-neural", it also contains numerous neurons – often sensory neurons – which send signals to the CNS. Therefore, the ectoderm really should be thought of as "more neural" and "less neural" rather than "neural" and "non-neural". This is highly relevant for discussion of how a CNS evolved from an ectodermal nerve net, since part of the genetic mechanism(s) for patterning the bilaterian CNS are shared with the ectoderm as a whole.

The basic A/P patterning mechanism in bilaterians with *Six3/6* expressed anteriorly and *Wnt* genes posteriorly was inherited from an ancestral cnidarian-like animal. Additional genes (for example, *Hox* genes) were recruited to pattern the A/P axis of the ancestral bilaterian. Consequently, even though the ectoderm in annelids and arthropods is segmented, and the ectoderm of hemichordates and chordates is not, several genes that mediate A/P patterning of the CNS are expressed in similar patterns along the A/P axis of the non-neural ectoderm in early embryos/larvae of both protostomes, chordates and hemichordates. For example, the patterns of *Otx*, *Gbx* and *Hox* genes in the early ectoderm of the annelid *P. dumerilii* are conserved in the neuroectoderm of amphioxus embryos with an anterior *Otx* domain abutting that of *Gbx*, anterior to nested domains of *Hox* expression [48]. Expression of *Hox* genes and *Gbx* in the "non-neural ectoderm" in amphioxus is similar to that in the CNS [131,134,192]. In contrast, in the hemichordate ectoderm, while *Otx* and *Gbx* domains lie anterior to nested *Hox* domains, that of *Gbx* lies in between two domains of *Otx* [9].

More critical for distinguishing a CNS from the remainder of the ectoderm is the expression of the lateral/medial patterning genes (*msh/Msx*, *ind/Gsh1*, *vnd/Nkx2.2*), because expression of these genes is similar in the CNS of protostomes and chordates, but they are not expressed in comparable patterns in the "non-neural" ectoderm of either group.

The idea that the hemichordate ectoderm has organizing centers patterned by gene networks equivalent to those operating at the ANR, ZLI and ISO in the vertebrate CNS is based on similarities in gene expression as well as in some conserved gene interactions. However, some key genes involved in establishing these organizing centers in the vertebrate CNS are not similarly expressed in the hemichordate ectoderm. For example, in the *S. kowalevskii* ectoderm, the *Fezf* and *Irx* domains, which mark the vertebrate ZLI, do not abut at the collar/proboscis interface, while patterns of *Otx* and *Wnt1* as well as those of *Gbx* and *Fgf8/17/18*, which mark the ISO, are reversed. Moreover, since parts of gene networks are often coopted for roles in different tissues, the fact that two genes interact does not suffice to demonstrate that a given region acts as an organizer. Thus, without transplantation experiments or at least the demonstration that a given morphogen such as *Fgf8/17/18* can induce a change in cell fate when ectopically expressed, conclusions that homologs of these organizing centers are present in hemichordates are premature.

In addition, since several genes involved in specification of the ISO and ZLI are expressed in comparable positions in the amphioxus and vertebrate CNS as well as in the protostome CNS, claims that amphioxus has completely lost these regions are unwarranted. A more parsimonious explanation is that amphioxus has part of the machinery in place, upon which tunicates and vertebrates elaborated. Tunicates do appear to have more of the genetic machinery for specification of the ISO than amphioxus does, but have lost some key components, such as the *Gbx* gene.

The Ambulacraria (hemichordates and echinoderms), which are the sister group to chordates in the deuterostomes, are interesting in regard to what evolution can do. Both hemichordates and echinoderms have ectodermal nerve nets plus nerve cords. Gene expression indicates that the nerve cords of echinoderms, which have evolved pentamerous symmetry, are probably not homologous to chordate or protostome nerve cords. *Hox* genes, for example, are expressed in the developing coeloms of the adult, but not in the nerve cords or ectoderm [48,51]. It has been suggested that one or the other nerve cord of hemichordates is evolutionarily related to the chordate CNS, with historical opinions generally favoring the dorsal cord. However, dorsal expression of *BMPs* suggests that this homolog would be the ventral nerve cord even though it is the dorsal nerve cord that neurulates. If this is true, then a D/V inversion would have occurred at the base of the chordate lineage.

In sum, similar expression patterns of developmental genes involved in both A/P and D/V patterning in the protostome and chordate nerve cords as well as anatomical and functional similarities support the view that the ancestral bilaterian had a CNS. In the Ambulacraria, echinoderms may have lost this CNS, while it remains to be seen whether either of the hemichordate nerve cords is homologous to that of protostomes and chordates. Similarities of gene expression in the hemichordate ectoderm and chordate CNS may reflect conservation of A/P patterning in the ectoderm as a whole and in the CNS of bilaterians and/or a merging of the anterior portion of the brain and the proboscis ectoderm during evolution of hemichordates. A comprehensive study of expression of developmental genes in the hemichordate nerve cords would help to resolve these questions. In any event, similarities in gene expression between the CNS of amphioxus and vertebrates indicate that amphioxus has

at least part of the genetic mechanisms for the ANR, ZLI and MHB in place and suggest that vertebrates subsequently elaborated upon these gene networks.

Abbreviations

ANR: Anterior neural ridge; A/P: Anterior/posterior; BMP: Bone morphogenetic protein; CNS: Central nervous system; D/V: Dorso/ventral; ISO: Isthmic organizer; MHB: Midbrain/hindbrain boundary; ZLI: Zona limitans intrathalamica.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LZH coordinated and wrote much of the manuscript. MS, HE, SMS, JKY and VL contributed to writing the manuscript. JEC and MS prepared the figures. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Peter WH Holland, ND Holland and Nacho Maeso for critically reading the manuscript and for useful suggestions. JKY was supported by the National Science Council, Taiwan (NSC99-2627-B-001-003, NSC101-2923-B-001-004-MY2), and by the Career Development Award from Academia Sinica, Taiwan (AS-98-CDA-L06). MS was supported by funds from the ANR (ANR-09-BLAN-0262-02 and ANR-11-JSV2-002-01), CNRS and UPMC.

Author details

[1]Marine Biology Research Division, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093-0202, USA. [2]Laboratoire de Biologie du Développement de Villefranche-sur-Mer (UMR 7009 – CNRS/UPMC), Observatoire Océanologique de Villefranche-sur-Mer, 181 Chemin du Lazaret, B.P. 28, 06230 Villefranche-sur-Mer, France. [3]CNRS, UMR 7232, BIOM, Université Pierre et Marie Curie Paris 06, Observatoire Océanologique, 66650 Banyuls-sur-Mer, France. [4]Institut de Génomique Fonctionnelle de Lyon (CNRS UMR5242, UCBL, ENS, INRA 1288), Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France. [5]Department of Zoology, University of Oxford, The Tinbergen Building, South Parks Road, Oxford OX1 3PS, UK. [6]Institute of Cellular and Organismic Biology, Academia Sinica, Taipei 11529, Taiwan.

Received: 8 April 2013 Accepted: 14 August 2013
Published: 7 October 2013

References

1. Arendt D, Denes AS, Jékely G, Tessmar-Raible K: **The evolution of nervous system centralization.** *Philos Trans R Soc Lond B Biol Sci* 2008, **363:**1523–1528.
2. Tomer R, Denes AS, Tessmar-Raible K, Arendt D: **Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium.** *Cell* 2010, **142:**800–809.
3. Strausfeld NJ, Hirth F: **Deep homology of arthropod central complex and vertebrate basal ganglia.** *Science* 2013, **340:**157–161.
4. Strausfeld NJ: **Brain homology: Dohrn of a new era?** *Brain Behav Evol* 2010, **76:**165–167.
5. Hirth F, Kammermeier L, Frei E, Walldorf U, Noll M, Reichert H: **An urbilaterian origin of the tripartite brain: developmental genetic insights from *Drosophila*.** *Development* 2003, **130:**2365–2373.
6. Sen S, Reichert H, VijayRaghavan K: **Conserved roles of *ems/Emx* and *otd/Otx* genes in olfactory and visual system development in Drosophila and mouse.** *Open Biol* 2013, **3:**120177.
7. Bailly X, Reichert H, Hartenstein V: **The urbilaterian brain revisited: novel insights into old questions from new flatworm clades.** *Dev Genes Evol* 2013, **223:**149–157.
8. De Robertis EM: **The molecular ancestry of segmentation mechanisms.** *Proc Natl Acad Sci U S A* 2008, **105:**16411–16412.
9. Pani AM, Mullarkey EE, Aronowicz J, Assimacopoulos S, Grove EA, Lowe CJ: **Ancient deuterostome origins of vertebrate brain signalling centres.** *Nature* 2012, **483:**289–294.
10. Gerhart J, Lowe C, Kirschner M: **Hemichordates and the origin of chordates.** *Curr Opin Genet Dev* 2005, **15:**461–467.
11. Wagner GP: **The biological homology concept.** *Annu Rev Ecol Sys* 1989, **20:**51–69.
12. Cracraft J: **Phylogeny and evo-devo: characters, homology, and the historical analysis of the evolution of development.** *Zoology* 2005, **108:**345–356.
13. Richter S: **Homologies in phylogenetic analyses – concept and tests.** *Theory Biosci* 2005, **124:**105–120.
14. Jenner RA: **Unburdening evo-devo: ancestral attractions, model organisms, and basal baloney.** *Dev Genes Evol* 2006, **216:**385–394.
15. Ghiselin MT: **The nomenclature of correspondence: a new look at 'homology' and 'analogy'.** In *Evolution, Brain and Behavior: Persistent Problems.* 2nd edition. Edited by Masterton RB, Hodos W, Jerison H. New York: Wiley; 1976.
16. Wagner GP: **The developmental genetics of homology.** *Nat Rev Genet* 2007, **8:**473–479.
17. Hodos W: ***The concept of homology and the evolution of behavior.*** *In Evolution, Brain and Behavior: Persistent Problems. Volume 2. Edited by Masterton RB, Hodos W, Jerison H.* New York: Wiley; 1976.
18. Remane A: **Methodological problems of hominid phylogeny. 3. Phylogeny of lifestyles and the origin of upright gait.** *Z Morphol Anthropol* 1956, **48:**28–54.
19. Haeckel E: **The gastraea-theory, the phylogenetic classification of the animal kindgom and the homology of the germ-lamellae.** *Q J Microsc Sci* 1874, **14:**142–165. 223–247.
20. Striedter GF, Northcutt RG: **Biological hierarchies and the concept of homology.** *Brain Behav Evol* 1991, **38:**177–189.
21. Bolker JA, Raff RA: **Developmental genetics and traditional homology.** *Bioessays* 1996, **18:**489–494.
22. Sasai Y, Lu B, Steinbeisser H, De Robertis EM: **Regulation of neural induction by the Chd and Bmp-4 antagonistic patterning signals *Xenopus* in.** *Nature* 1995, **376:**333–336.
23. De Robertis EM: **Evo-devo: variations on ancestral themes.** *Cell* 2008, **132:**185–195.
24. Lowe CJ, Wu M, Salic A, Evans L, Lander E, Stange-Thomann N, Gruber CE, Gerhart J, Kirschner M: **Anteroposterior patterning in hemichordates and the origins of the chordate nervous system.** *Cell* 2003, **113:**853–865.
25. Aronowicz J, Lowe CJ: ***Hox* gene expression in the hemichordate *Saccoglossus kowalevskii* and the evolution of deuterostome nervous systems.** *Integr Comp Biol* 2006, **46:**890–901.
26. Hejnol A, Martindale MQ: **Acoel development indicates the independent evolution of the bilaterian mouth and anus.** *Nature* 2008, **456:**382–386.
27. Gerhart J: **Inversion of the chordate body axis: are there alternatives?** *Proc Natl Acad Sci U S A* 2000, **97:**4445–4448.
28. Gerhart J: **The deuterostome ancestor.** *J Cell Physiology* 2006, **209:**677–685.
29. Sprecher SG, Reichert H: **The urbilaterian brain: developmental insights into the evolutionary origin of the brain in insects and vertebrates.** *Arthropod Struct Dev* 2003, **32:**141–156.
30. Martindale Mark Q: **Evolution of development: the details are in the entrails.** *Curr Biol* 2013, **23:**R25–R28.
31. Northcutt RG: **Evolution of centralized nervous systems: two schools of evolutionary thought.** *Proc Natl Acad Sci USA* 2012, **109**(Suppl 1):10626–10633.
32. Koizumi O, Sato N, Goto C: **Chemical anatomy of hydra nervous system using antibodies against hydra neuropeptides: a review.** *Hydrobiologia* 2004, **530–531:**41–47.
33. Haag ES: **Echinoderm rudiments, rudimentary bilaterians, and the origin of the chordate CNS.** *Evol Dev* 2005, **7:**280–281.
34. Holland ND: **Early central nervous system evolution: an era of skin brains?** *Nat Rev Neurosci* 2003, **4:**617–627.
35. Philippe H, Brinkmann H, Copley RR, Moroz LL, Nakano H, Poustka AJ, Wallberg A, Peterson KJ, Telford MJ: **Acoelomorph flatworms are deuterostomes related to *Xenoturbella*.** *Nature* 2011, **470:**255–258.
36. Hejnol A, Obst M, Stamatakis A, Ott M, Rouse GW, Edgecombe GD, Martinez P, Baguñà J, Bailly X, Jondelius U, Wiens M, Müller WE, Seaver E, Wheeler WC, Martindale MQ, Giribet G, Dunn CW: **Assessing the root of bilaterian animals with scalable phylogenomic methods.** *Proc Roy Soc B: Biol Sci* 2009, **276:**4261–4270.
37. Bery A, Cardona A, Martinez P, Hartenstein V: **Structure of the central nervous system of a juvenile acoel, Symsagittifera roscoffensis.** *Dev Genes Evol* 2010, **220:**61–76.

38. Raikova OI, Reuter M, Jondelius U, Gustafsson MKS: **An immunocytochemical and ultrastructural study of the nervous and muscular systems of *Xenoturbella westbladi* (Bilateria inc. sed.).** *Zoomorphology* 2000, **120:**107–118.
39. Sinigaglia C, Busengdal H, Leclère L, Technau U, Rentzsch F: **The bilaterian head patterning gene *six3/6* controls aboral domain development in a cnidarian.** *PLoS Biol* 2013, **11:**e1001488.
40. Yaguchi S, Yaguchi J, Angerer RC, Angerer LM: **A Wnt–FoxQ2–Nodal pathway links primary and secondary axis specification in sea urchin embryos.** *Dev Cell* 2008, **14:**97–107.
41. Yu J-K, Holland N, Holland L: ***AmphiFoxQ2*, a novel winged helix/forkhead gene, exclusively marks the anterior end of the amphioxus embryo.** *Dev Genes Evol* 2003, **213:**102–105.
42. Steinmetz PR, Urbach R, Posnien N, Eriksson J, Kostyuchenko RP, Brena C, Guy K, Akam M, Bucher G, Arendt D: ***Six3* demarcates the anterior-most developing brain region in bilaterian animals.** *EvoDevo* 2010, **1:**14.
43. Kozmik Z, Holland ND, Kreslova J, Oliveri D, Schubert M, Jonasova K, Holland LZ, Pestarino M, Benes V, Candiani S: ***Pax–Six–Eya–Dach* network during amphioxus development: conservation *in vitro* but context specificity *in vivo*.** *Dev Biol* 2007, **306:**143–159.
44. Santagata S, Resh C, Hejnol A, Martindale MQ, Passamaneck YJ: **Development of the larval anterior neurogenic domains of *Terebratalia transversa* (Brachiopoda) provides insights into the diversification of larval apical organs and the spiralian nervous system.** *EvoDevo* 2012, **3:**3.
45. Hirth F, Reichert H: **Conserved genetic programs in insect and mammalian brain development.** *BioEssays* 1999, **21:**677–684.
46. Lichtneckert R, Reichert H: **Anteroposterior regionalization of the brain: genetic and comparative aspects.** *Adv Exp Med Biol* 2008, **628:**32–41.
47. Sasakura Y, Mita K, Ogura Y, Horie T: **Ascidians as excellent chordate models for studying the development of the nervous system during embryogenesis and metamorphosis.** *Dev Growth Differ* 2012, **54:**420–437.
48. Steinmetz PRH, Kostyuchenko RP, Fischer A, Arendt D: **The segmental pattern of *otx*, *gbx*, and *Hox* genes in the annelid *Platynereis dumerilii*.** *Evol Dev* 2011, **13:**72–79.
49. Lee PN, Callaerts P, de Couet HG, Martindale MQ: **Cephalopod *Hox* genes and the origin of morphological novelties.** *Nature* 2003, **424:**1061–1065.
50. Tihanyi BVT, Regos A, Ari E, Müller F, Takács-Vellai K: **The *C. elegans Hox* gene *ceh-13* regulates cell migration and fusion in a non-colinear way. Implications for the early evolution of *Hox* clusters.** *BMC Dev Biol* 2010, **10:**78.
51. Arenas-Mena C, Cameron AR, Davidson EH: **Spatial expression of Hox cluster genes in the ontogeny of a sea urchin.** *Development* 2000, **127:**4631–4643.
52. Barak H, Preger-Ben Noon E, Reshef R: **Comparative spatiotemporal analysis of *Hox* gene expression in early stages of intermediate mesoderm formation.** *Dev Dynam* 2012, **241:**1637–1649.
53. Hara Y, Yamaguchi M, Akasaka K, Nakano H, Nonaka M, Amemiya S: **Expression patterns of *Hox* genes in larvae of the sea lily *Metacrinus rotundus*.** *Dev Genes Evol* 2006, **216:**797–809.
54. Seo H-C, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaat M, Weissenbach J, Lehrach H, Wincker P, Reinhardt R, Chourrout D: **Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*.** *Nature* 2004, **431:**67–71.
55. Galliot B, Quiquand M, Ghila L, de Rosa R, Miljkovic-Licina M, Chera S: **Origins of neurogenesis, a cnidarian view.** *Dev Biol* 2009, **332:**2–24.
56. Ryan JF, Mazza ME, Pang K, Matus DQ, Baxevanis AD, Martindale MQ, Finnerty JR: **Pre-bilaterian origins of the Hox cluster and the Hox code: evidence from the sea anemone, *Nematostella vectensis*.** *PLoS ONE* 2007, **2:**e153.
57. de Rosa R, Grenier JK, Andreeva T, Cook CE, Adoutte A, Akam M, Carroll SB, Balavoine G: **Hox genes in brachiopods and priapulids and protostome evolution.** *Nature* 1999, **399:**772–776.
58. Hejnol A, Martindale MQ: **Coordinated spatial and temporal expression of Hox genes during embryogenesis in the acoel *Convolutriloba longifissura*.** *BMC Biol* 2009, **7:**65.
59. Moreno E, Nadal M, Baguna J, Martinez P: **Tracking the origins of the bilaterian Hox patterning system: insights from the acoel flatworm *Symsagittifera roscoffensis*.** *Evol Dev* 2009, **11:**574–581.
60. Reichert H: **A tripartite organization of the urbilaterian brain: developmental genetic evidence from *drosophila*.** *Brain Res Bull* 2005, **66:**491–494.
61. Inoue F, Kurokawa D, Takahashi M, Aizawa S: **Gbx2 directly restricts Otx2 expression to forebrain and midbrain, competing with Class III POU factors.** *Mol Cell Biol* 2012, **32:**2618–2627.
62. Püschel AW, Westerfield M, Dressler GR: **Comparative analysis of Pax-2 protein distributions during neurulation in mice and zebrafish.** *Mech Dev* 1992, **38:**197–208.
63. Pfeiffer BD, Jenett A, Hammonds AS, Ngo T-TB, Misra S, Murphy C, Scully A, Carlson JW, Wan KH, Laverty TR, Mungall C, Svirskas R, Kadonaga JT, Doe CQ, Eisen MB, Celniker SE, Rubin GM: **Tools for neuroanatomy and neurogenetics in *Drosophila*.** *Proc Natl Acad Sci U S A* 2008, **105:**9715–9720.
64. Irimia M, Piñeiro C, Maeso I, Gómez-Skarmeta JL, Casares F, Garcia-Fern àndez J: **Conserved developmental expression of *Fezf* in chordates and *Drosophila* and the origin of the Zona Limitans Intrathalamica (ZLI) brain organizer.** *EvoDevo* 2010, **1:**7.
65. Hirata T, Nakazawa M, Muraoka O, Nakayama R, Suda Y, Hibi M: **Zinc-finger genes *Fez* and *Fez-like* function in the establishment of diencephalon subdivisions.** *Development* 2006, **133:**3993–4004.
66. Mizutani CM, Bier E: **EvoD/Vo: the origins of BMP signalling in the neuroectoderm.** *Nat Rev Genet* 2008, **9:**663–677.
67. Urbach R, Technau GM: **Dorsoventral patterning of the brain: a comparative approach.** In , Volume 628 Edited by Technau GM. New York: Springer; 2008:42–56.
68. Holland LZ: **Chordate roots of the vertebrate nervous system: expanding the molecular toolkit.** *Nat Rev Neurosci* 2009, **10:**736–746.
69. Vopalensky P, Pergner J, Liegertova M, Benito-Gutierrez E, Arendt D, Kozmik Z: **Molecular analysis of the amphioxus frontal eye unravels the evolutionary origin of the retina and pigment cells of the vertebrate eye.** *Proc Natl Acad Sci U S A* 2012, **109:**15383–15388.
70. Holland ND, Panganiban G, Henyey EL, Holland LZ: **Sequence and developmental expression of *AmphiDll*, an amphioxus Distal-less gene transcribed in the ectoderm, epidermis and nervous system: insights into evolution of craniate forebrain and neural crest.** *Development* 1996, **122:**2911–2920.
71. Shimeld SM, van den Heuvel M, Dawber R, Briscoe J: **An amphioxus *Gli* gene reveals conservation of midline patterning and the evolution of Hedgehog signalling diversity in chordates.** *PLoS ONE* 2007, **2:**e864.
72. Robertshaw E, Kiecker C: **Phylogenetic origins of brain organisers.** *Scientifica* 2012, **2012:**1–14.
73. Seimiya M, Gehring WJ: **The *Drosophila* homeobox gene *optix* is capable of inducing ectopic eyes by an eyeless-independent mechanism.** *Development* 2000, **127:**1879–1886.
74. Beccari L, Marco-Ferreres R, Bovolenta P: **The logic of gene regulatory networks in early vertebrate forebrain patterning.** *Mech Dev* 2013, **130:**95–111.
75. Mazet F, Hutt JA, Millard J, Shimeld SM: **Pax gene expression in the developing central nervous system of *Ciona intestinalis*.** *Gene Exp Patt* 2003, **3:**743–745.
76. D'Aniello S, D'Aniello E, Locascio A, Memoli A, Corrado M, Russo MT, Aniello F, Fucci L, Brown ER, Branno M: **The ascidian homolog of the vertebrate homeobox gene *Rx* is essential for ocellus development and function.** *Differentiation* 2006, **74:**222–234.
77. Caracciolo A, Di Gregorio A, Aniello F, Di Lauro R, Branno M: **Identification and developmental expression of three *Distal-less* homeobox containing genes in the ascidian *Ciona intestinalis*.** *Mech Dev* 2000, **99:**173–176.
78. Imai KS, Satoh N, Satou Y: **Region specific gene expressions in the central nervous system of the ascidian embryo.** *Mech Dev* 2002, **119**(Suppl):S275–S277.
79. Islam AFMT, Moly PK, Miyamoto Y, Kusakabe TG: **Distinctive expression patterns of Hedgehog pathway genes in the *Ciona intestinalis* larva: implications for a role of Hedgehog signaling in postembryonic development and chordate evolution.** *Zoolog Sci* 2010, **27:**84–90.
80. Retaux S, Kano S: **Midline signaling and evolution of the forebrain in chordates: a focus on the lamprey hedgehog case.** *Integr Comp Biol* 2010, **50:**98–109.
81. Denes AS, Jékely G, Steinmetz PRH, Raible F, Snyman H, Prud'homme B, Ferrier DEK, Balavoine G, Arendt D: **Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in Bilateria.** *Cell* 2007, **129:**277–288.
82. Heuer CM, Mueller CHG, Todt C, Loesel R: **Comparative neuroanatomy suggests repeated reduction of neuroarchitectural complexity in Annelida.** *Front Zool* 2010, **7:**13.

83. Sikes JM, Bely AE: **Making heads from tails: development of a reversed anterior–posterior axis during budding in an acoel.** *Dev Biol* 2010, **338:**86–97.
84. Achatz JG, Martinez P: **The nervous system of *Isodiametra pulchra* (Acoela) with a discussion on the neuroanatomy of the Xenacoelomorpha and its evolutionary implications.** *Front Zool* 2012, **9:**27.
85. Lacalli TC: **Protochordate body plan and the evolutionary role of larvae: old controversies resolved?** *Can J Zool* 2005, **83:**216–224.
86. Dohrn A: *Der Ursprung der Wirbelthiere und das Princip des Functionswechsels.* Leipzig: Wilhelm Englmann; 1875.
87. Van-Wijhe JW: **Die Terminalbeuge des Vorderhirns bei den Chordaten und die prootischen Nerven der Kranioten bei Amphioxus. Erste Mitteilung plus Zweite Mitteilung: Schluss des Riechorgans.** *Proc Koninklijke Akademie Van Wetenschappen Te Amsterdam* 1931, **34:**131–141. (Erste Mitteilung); & 636–653 (Zweite Mitteilung).
88. Nomaksteinsky M, Röttinger E, Dufour HD, Chettouh Z, Lowe CJ, Martindale MQ, Brunet J-F: **Centralization of the deuterostome nervous system predates chordates.** *Curr Biol* 2009, **19:**1264–1269.
89. De Robertis EM, Sasai Y: **A common plan for dorsoventral patterning in Bilateria.** *Nature* 1996, **380:**37–40.
90. Duboc V, Röttinger E, Lapraz F, Besnardeau L, Lepage T: **Left-right asymmetry in the sea urchin embryo is regulated by Nodal signaling on the right side.** *Dev Cell* 2005, **9:**147–158.
91. Grande C, Patel NH: **Nodal signalling is involved in left-right asymmetry in snails.** *Nature* 2009, **457:**1007–1011.
92. Nielsen C: **Homology of echinoderm radial nerve cords and the chordate neural tube?** *Evol Dev* 2006, **8:**1–2.
93. Sly BJ, Hazel JC, Popodi EM, Raff RA: **Patterns of gene expression in the developing adult sea urchin central nervous system reveal multiple domains and deep-seated neural pentamery.** *Evol Dev* 2002, **4:**189–204.
94. Cisternas P, Byrne M: **Expression of Hox4 during development of the pentamerous juvenile sea star, *Parvulastra exigua*.** *Dev Genes Evol* 2009, **219:**613–618.
95. Morris VB, Byrne M: **Involvement of two Hox genes and Otx in echinoderm body-plan morphogenesis in the sea urchin *Holopneustes purpurescens*.** *J Exp Zool Part B: Mol Dev Evol* 2005, **304B:**456–467.
96. Lacalli TC: **Head organization and the head/trunk relationship in protochordates: problems and prospects.** *Integr Comp Biol* 2008, **48:**620–629.
97. Nieuwenhuys R: **Deuterostome brains: synopsis and commentary.** *Brain Res Bull* 2002, **57:**257–270.
98. Tagawa K, Humphreys T, Satoh N: ***T-brain* expression in the apical organ of hemichordate tornaria larvae suggests its evolutionary link to the vertebrate forebrain.** *J Exp Zool* 2000, **288:**23–31.
99. Stach T, Gruhl A, Kaul-Strehlow S: **The central and peripheral nervous system of *Cephalodiscus gracilis* (Pterobranchia, Deuterostomia).** *Zoomorphology* 2012, **131:**11–24.
100. Romer AS: **Major steps in vertebrate evolution.** *Science* 1967, **158:**1629–1637.
101. Williams JB: **Sessile lifestyle and origin of chordates.** *New Zealand J Zool* 1996, **23:**111–133.
102. Osborn KJ, Kuhnz LA, Priede IG, Urata M, Gebruk AV, Holland ND: **Diversification of acorn worms (Hemichordata, Enteropneusta) revealed in the deep sea.** *Proc Roy Soc B: Biol Sci* 2012, **279:**1646–1654.
103. Cannon JT, Rychel AL, Eccleston H, Halanych KM, Swalla BJ: **Molecular phylogeny of hemichordata, with updated status of deep-sea enteropneusts.** *Mol Phylog Evol* 2009, **52:**17–24.
104. Caron J-B, Morris SC, Cameron CB: **Tubicolous enteropneusts from the Cambrian period.** *Nature* 2013, **495:**503–506.
105. Röttinger E, Lowe CJ: **Evolutionary crossroads in developmental biology: hemichordates.** *Development* 2012, **139:**2463–2475.
106. Miyamoto N, Nakajima Y, Wada H, Saito Y: **Development of the nervous system in the acorn worm *Balanoglossus simodensis*: insights into nervous system evolution.** *Evol Dev* 2010, **12:**416–424.
107. Kaul S, Stach T: **Ontogeny of the collar cord: neurulation in the hemichordate *Saccoglossus kowalevskii*.** *J Morph* 2010, **271:**1240–1259.
108. Bateson W: **The later stages in the development of *Balanoglossus kowalevskii*, with a suggestion as to the affinities of the Enteropneusta.** *Q J Microsc Sci* 1886, **126:**511–533.
109. Rehkämper G, Welsch U, Dilly PN: **Fine structure of the ganglion of *Cephalodiscus gracilis* (Pterobranchia, Hemichordata).** *J Comp Neurol* 1987, **259:**308–315.
110. Brown FD, Prendergast A, Swalla BJ: **Man is but a worm: chordate origins.** *Genesis* 2008, **46:**605–613.
111. Bullock TH: **The giant nerve fiber system in balanoglossids.** *J Comp Neurol* 1944, **80:**355–367.
112. Nielsen C, Hay-Schmidt A: **Development of the enteropneust *Ptychodera flava*: ciliary bands and nervous system.** *J Morph* 2007, **268:**551–570.
113. Lowe CJ, Terasaki M, Wu M, Freeman RM Jr, Runft L, Kwan K, Haigo S, Aronowicz J, Lander E, Gruber C, Smith M, Kirschner M, Gerhart J: **Dorsoventral patterning in hemichordates: Insights into early chordate evolution.** *PLoS Biol* 2006, **4:**e291.
114. Taguchi S, Tagawa K, Humphreys T, Satoh N: **Group B *Sox* genes that contribute to specification of the vertebrate brain are expressed in the apical organ and ciliary bands of hemichordate larvae.** *Zool Sci* 2002, **19:**57–66.
115. Tagawa K, Satoh N, Humphreys T: **Molecular studies of hemichordate development: a key to understanding the evolution of bilateral animals and chordates.** *Evol Dev* 2001, **3:**443–454.
116. Peterson KJ, Cameron RA, Tagawa K, Satoh N, Davidson EH: **A comparative molecular approach to mesodermal patterning in basal deuterostomes: the expression pattern of *Brachyury* in the enteropneust hemichordate *Ptychodera flava*.** *Development* 1999, **126:**85–95.
117. Onai T, Yu J-K, Blitz IL, Cho KWY, Holland LZ: **Opposing Nodal/Vg1 and BMP signals mediate axial patterning in embryos of the basal chordate amphioxus.** *Dev Biol* 2010, **344:**377–389.
118. Wlizla M: **Evolution of Nodal signaling in deuterostomes: insights from Saccoglossus kowalevskii. PhD dissertation. University of Chicago.** 2011:195.
119. Saudemont A, Haillot E, Mekpoh F, Bessodes N, Quirin M, Lapraz F, Duboc V, Röttinger E, Range R, Oisel A, Besnardeau L, Wincker P, Lepage T: **Ancestral regulatory circuits governing ectoderm patterning downstream of Nodal and BMP2/4 revealed by gene regulatory network analysis in an echinoderm.** *PLoS Genet* 2010, **6:**e1001259.
120. Grande C, Patel NH: **Lophotrochozoa get into the game: the nodal pathway and left/right asymmetry in Bilateria.** *Cold Spring Harb Symp Quant Biol* 2009, **74:**281–287.
121. de Beer G: *Homology, An Unsolved Problem*, Volume 11. Oxford: Oxford University Press; 1971.
122. Roth VL: **The biological basis of homology.** In *Ontogeny and Systematics.* Edited by Humphries CJ. New York: Columbia University Press; 1988.
123. Panganiban G, Irvine SM, Lowe CJ, Roehl HH, Corley LS, Sherbon B, Grenier JK, Fallon JF, Kimble J, Walker M, Wray GA, Swalla BJ, Martindale MQ, Carroll SB: **The origin and evolution of animal appendages.** *Proc Natl Acad Sci USA* 1997, **94:**5162–5166.
124. Northcutt RG: **Changing views of brain evolution.** *Brain Res Bull* 2001, **55:**663–674.
125. Shubin N, Tabin C, Carroll SB: **Deep homology and the origins of evolutionary novelty.** *Nature* 2009, **457:**818–823.
126. Lowe CJ: **Molecular genetic insights into deuterostome evolution from the direct-developing hemichordate *Saccoglossus kowalevskii*.** *Philos Trans R Soc Lond B Biol Sci* 2008, **363:**1569–1578.
127. Stokes MD, Holland ND: **Embryos and larvae of a lancelet, *Branchiostoma floridae*, from hatching through metamorphosis: growth in the laboratory and external morphology.** *Acta Zool* 1995, **76:**105–120.
128. Bone Q, Ryan KP: **Cupular sense organs in *Ciona* (Tunicata: Ascidiacea).** *J Zool, London* 1978, **186:**417–429.
129. Wood HB, Episkopou V: **Comparative expression of the mouse *Sox1*, *Sox2* and *Sox3* genes from pre-gastrulation to early somite stages.** *Mech Dev* 1999, **86:**197–201.
130. Satoh G, Wang Y, Zhang P, Satoh N: **Early development of amphioxus nervous system with special reference to segmental cell organization and putative sensory cell precursors: a study based on the expression of pan-neuronal marker gene *Hu/elav*.** *J Exp Zool* 2001, **291:**354–364.
131. **Ghost Database.** [http://ghost.zool.kyoto-u.ac.jp/SearchGenomekh.html]
132. Schubert M, Holland ND, Escriva H, Holland LZ, Laudet V: **Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and their gene expression in a developing chordate (amphioxus).** *Proc Natl Acad Sci U S A* 2004, **101:**10320–10325.
133. Sasakura Y, Kanda M, Ikeda T, Horie T, Kawai N, Ogura Y, Yoshida R, Hozumi A, Satoh N, Fujiwara S: **Retinoic acid-driven *Hox1* is required in the epidermis for forming the otic/atrial placodes during ascidian metamorphosis.** *Development* 2012, **139:**2156–2160.

134. Watari-Goshima N, Chisaka O: **Chicken HOXA3 gene: its expression pattern and role in branchial nerve precursor cell migration.** *Int J Biol Sci* 2011, **7**:87–101.
135. Castro LFC, Rasmussen SLK, Holland PWH, Holland ND, Holland LZ: **A *Gbx* homeobox gene in amphioxus: insights into ancestry of the ANTP class and evolution of the midbrain/hindbrain boundary.** *Dev Biol* 2006, **295**:40–51.
136. Sánchez-Calderón H, Martín-Partido G, Hidalgo-Sánchez M: **Differential expression of *Otx2, Gbx2, Pax2,* and *Fgf8* in the developing vestibular and auditory sensory organs.** *Brain Res Bull* 2002, **57**:321–323.
137. Steventon B, Mayor R, Streit A: **Mutual repression between Gbx2 and Otx2 in sensory placodes reveals a general mechanism for ectodermal patterning.** *Dev Biol* 2012, **367**:55–65.
138. Sharman AC, Shimeld SM, Holland PWH: **An amphioxus *Msx* gene expressed predominantly in the dorsal neural tube.** *Dev Genes Evol* 1999, **209**:260–263.
139. Osborne PW, Benoit G, Laudet V, Schubert M, Ferrier DE: **Differential regulation of ParaHox genes by retinoic acid in the invertebrate chordate amphioxus (*Branchiostoma floridae*).** *Dev Biol* 2009, **327**:252–262.
140. Venkatesh TV, Holland ND, Holland LZ, Su M-T, Bodmer R: **Sequence and developmental expression of amphioxus *AmphiNk2–1*: insights into the evolutionary origin of the vertebrate thyroid gland and forebrain.** *Dev Genes Evol* 1999, **209**:254–259.
141. Kawashima T, Murakami AR, Ogasawara M, Tanaka KJ, Isoda R, Sasakura Y, Nishikata T, Okano H, Makabe KW: **Expression patterns of *musashi* homologs of the ascidians, *Halocynthia roretzi* and *Ciona intestinalis*.** *Dev Genes Evol* 2000, **210**:162–165.
142. Wada S, Tokuoka M, Shoguchi E, Kobayashi K, Di Gregorio A, Spagnuolo A, Branno M, Kohara Y, Rokhsar D, Levine M, Saiga H, Satoh N, Satou Y: **A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. II. Genes for homeobox transcription factors.** *Dev Genes Evol* 2003, **213**:222–234.
143. Ghanbari H, Seo H-C, Fjose A, Brändli AW: **Molecular cloning and embryonic expression of *Xenopus Six* homeobox genes.** *Mech Dev* 2001, **101**:271–277.
144. Osorio J, Mazan S, Rétaux S: **Organisation of the lamprey (*Lampetra fluviatilis*) embryonic brain: insights from LIM-homeodomain, *Pax* and *hedgehog* genes.** *Dev Biol* 2005, **288**:100–112.
145. Toresson H, Martinez-Barbera JP, Bardsley A, Caubit X, Krauss S: **Conservation of *BF-1* expression in amphioxus and zebrafish suggests evolutionary ancestry of anterior cell types that contribute to the vertebrate telencephalon.** *Dev Genes Evol* 1998, **208**:431–439.
146. Hatini V, Ye X, Balas G, Lai E: **Dynamics of placodal lineage development revealed by targeted transgene expression.** *Dev Dynam* 1999, **215**:332–343.
147. Kaltenbach SL, Yu J-K, Holland ND: **The origin and migration of the earliest-developing sensory neurons in the peripheral nervous system of amphioxus.** *Evol Dev* 2009, **11**:142–151.
148. Logan C, Wingate RJ, McKay IJ, Lumsden A: ***Tlx-1* and *Tlx-3* homeobox gene expression in cranial sensory ganglia and hindbrain of the chick embryo: markers of patterned connectivity.** *J Neurosci* 1998, **18**:5389–5402.
149. Chellappa R, Li S, Pauley S, Jahan I, Jin K, Xiang M: ***Barhl1* regulatory sequences required for cell-specific gene expression and autoregulation in the inner ear and central nervous system.** *Mol Cell Biol* 2008, **28**:1905–1914.
150. Kawahara A, Dawid IB: **Developmental expression of zebrafish *emx1* during early embryogenesis.** *Gene Exp Patt* 2002, **2**:201–206.
151. Bardet SM, Martinez-de-la-Torre M, Northcutt RG, Rubenstein JLR, Puelles L: **Conserved pattern of OTP-positive cells in the paraventricular nucleus and other hypothalamic sites of tetrapods.** *Brain Res Bull* 2008, **75**:231–235.
152. Ma P, Zhao S, Zeng W, Yang Q, Li C, Lv X, Zhou Q, Mao B: ***Xenopus Dbx2* is involved in primary neurogenesis and early neural plate patterning.** *Biochem Biophy Res Com* 2011, **412**:170–174.
153. Langeland JA, Holland LZ, Chastain RA, Holland ND: **An amphioxus LIM-homeobox gene, *AmphiLim1/5*, expressed early in the invaginating organizer region and later in differentiating cells of the kidney and central nervous system.** *Int J Biol Sci* 2006, **2**:110–116.
154. Feijóo CG, Saldias MP, De la Paz JF, Gómez-Skarmeta JL, Allende ML: **Formation of posterior cranial placode derivatives requires the *Iroquois* transcription factor irx4a.** *Mol Cell Neurosci* 2009, **40**:328–337.
155. Matsuura M, Nishihara H, Onimaru K, Kokubo N, Kuraku S, Kusakabe R, Okada N, Kuratani S, Tanaka M: **Identification of four *Engrailed* genes in the Japanese lamprey, Lethenteron japonicum.** *Dev Dyn* 2008, **237**:1581–1589.
156. Ikuta T, Yoshida N, Satoh N, Saiga H: ***Ciona intestinalis Hox* gene cluster: its dispersed structure and residual colinear expression in development.** *Proc Natl Acad Sci U S A* 2004, **101**:15118–15123.
157. Makki N, Capecchi MR: **Hoxa1 lineage tracing indicates a direct role for *Hoxa1* in the development of the inner ear, the heart, and the third rhombomere.** *Dev Biol* 2010, **341**:499–509.
158. Takio Y, Kuraku S, Murakami Y, Pasqualetti M, Rijli FM, Narita Y, Kuratani S, Kusakabe R: ***Hox* gene expression patterns in *Lethenteron japonicum* embryos – insights into the evolution of the vertebrate Hox code.** *Dev Biol* 2007, **308**:606–620.
159. Uchiyama K, Otsuka R, Hanaoka K: ***cHox11L2*, a *Hox11* related gene, is expressed in the peripheral nervous system and subpopulation of the spinal cord during chick development.** *Neurosci Lett* 1999, **273**:97–100.
160. Sanchez-Arrones L, Stern CD, Bovolenta P, Puelles L: **Sharpening of the anterior neural border in the chick by rostral endoderm signalling.** *Development* 2012, **139**:1034–1044.
161. Tendeng C, Houart C: **Cloning and embryonic expression of five distinct sfrp genes in the zebrafish *Danio rerio*.** *Gene Expr Patterns* 2006, **6**:761–771.
162. Eagleson GW, Dempewolf RD: **The role of the anterior neural ridge and Fgf-8 in early forebrain patterning and regionalization in *Xenopus laevis*.** *Comp Biochem Physiol B Biochem Mol Biol* 2002, **132**:179–189.
163. Vieira C, Martinez S: **Sonic hedgehog from the basal plate and the zona limitans intrathalamica exhibits differential activity on diencephalic molecular regionalization and nuclear structure.** *Neuroscience* 2006, **143**:129–140.
164. Green SA, Norris RP, Terasaki M, Lowe CJ: **FGF signaling induces mesoderm in the hemichordate *Saccoglossus kowalevskii*.** *Development* 2013, **140**:1024–1033.
165. Holland ND, Holland LZ: **Amphioxus and the utility of molecular genetic data for hypothesizing body part homologies between distantly related animals.** *Am Zool* 1999, **39**:630–640.
166. Holland LZ: **Evolution of new characters after whole genome duplications: insights from amphioxus.** *Sem Cell Dev Biol* 2013, **24**:101–109.
167. Capdevila J, Belmonte JCI: **Patterning mechanisms controlling vertebrte limb development.** *Ann Rev Cell Dev Biol* 2001, **17**:87–132.
168. Röttinger E, Martindale MQ: **Ventralization of an indirect developing hemichordate by NiCl2 suggests a conserved mechanism of dorso-ventral (D/V) patterning in Ambulacraria (hemichordates and echinoderms).** *Dev Biol* 2011, **354**:173–190.
169. Range R, Angerer R, Angerer L: **Integration of canonical and noncanonical Wnt signaling pathways patterns the neuroectoderm along the anterior-posterior axis of sea urchin embryos.** *PLoS Biol* 2013, **11**:e1001467.
170. Emily-Fenouil F, Ghiglione C, Lhomond G, Lepage T, Gache C: ***GSK3beta/shaggy* mediates patterning along the animal–vegetal axis of the sea urchin embryo.** *Development* 1998, **125**:2489–2498.
171. Jeong J-Y, Einhorn Z, Mathur P, Chen L, Lee S, Kawakami K, Guo S: **Patterning the zebrafish diencephalon by the conserved zinc-finger protein Fezl.** *Development* 2007, **134**:127–136.
172. Scholpp S, Foucher I, Staudt N, Peukert D, Lumsden A, Houart C: ***Otx1l, Otx2* and *Irx1b* establish and position the ZLI in the diencephalon.** *Development* 2007, **134**:3167–3176.
173. Martinez-Ferre A, Navarro-Garberi M, Bueno C, Martinez S: **Wnt signal specifies the intrathalamic limit and its organizer properties by regulating Shh induction in the alar plate.** *J Neurosci* 2013, **33**:3967–3980.
174. Martinez-Ferre A, Martinez S: **Molecular regionalization of the diencephalon.** *Front Neurosci* 2012, **6**:73.
175. Puelles E, Annino A, Tuorto F, Usiello A, Acampora D, Czerny T, Brodski C, Ang S-L, Wurst W, Simeone A: ***Otx2* regulates the extent, identity and fate of neuronal progenitor domains in the ventral midbrain.** *Development* 2004, **131**:2037–2048.
176. Partanen J: **FGF signalling pathways in development of the midbrain and anterior hindbrain.** *J Neurochem* 2007, **101**:1185–1193.
177. Kengaku M, Capdevila J, Rodriguez-Esteban C, De La Peña J, Johnson RL, Belmonte JCI, Tabin CJ: **Distinct WNT pathways regulating AER formation and dorsoventral polarity in the chick limb bud.** *Science* 1998, **280**:1274–1277.
178. Tanaka M, Shigetani Y, Sugiyama S, Tamura K, Nakamura H, Ide H: **Apical ectodermal ridge induction by the transplantation of *En-1*-**

overexpressing ectoderm in chick limb bud. *Dev Growth Differ* 1998, **40**:423–429.
179. Glardon S, Holland LZ, Gehring WJ, Holland ND: **Isolation and developmental expression of the amphioxus Pax-6 gene (*AmphiPax-6*): insights into eye and photoreceptor evolution.** *Development* 1998, **125**:2701–2710.
180. Yu J-K, Satou Y, Holland ND, Shin-I T, Kohara Y, Satoh N, Bronner-Fraser M, Holland LZ: **Axial patterning in cephalochordates and the evolution of the organizer.** *Nature* 2007, **445**:613–617.
181. Shimeld SM: **The evolution of the hedgehog gene family in chordates: insights from amphioxus *hedgehog*.** *Dev Genes Evol* 1999, **209**:40–47.
182. Shimamura K, Rubenstein JL: **Inductive interactions direct early regionalization of the mouse forebrain.** *Development* 1997, **124**:2709–2718.
183. Imai KS, Stolfi A, Levine M, Satou Y: **Gene regulatory networks underlying the compartmentalization of the *Ciona* central nervous system.** *Development* 2009, **136**:285–293.
184. Schubert M, Holland LZ, Panopoulou GD, Lehrach H, Holland ND: **Characterization of amphioxus *AmphiWnt8*: insights into the evolution of patterning of the embryonic dorsoventral axis.** *Evol Dev* 2000, **2**:85–92.
185. Mazet F, Shimeld S: **Characterisation of an amphioxus *Fringe* gene and the evolution of the vertebrate segmentation clock.** *Dev Genes Evol* 2003, **213**:505–509.
186. Holland LZ, Venkatesh TV, Gorlin A, Bodmer R, Holland ND: **Characterization and developmental expression of *AmphiNk2-2*, an NK2 class homeobox gene from amphioxus (Phylum Chordata; Subphylum Cephalochordata).** *Dev Genes Evol* 1998, **208**:100–105.
187. Bertrand S, Camasses A, Somorjai I, Belgacem MR, Chabrol O, Escande M-L, Pontarotti P, Escriva H: **Amphioxus FGF signaling predicts the acquisition of vertebrate morphological traits.** *Proc Natl Acad Sci U S A* 2011, **108**:9160–9165.
188. Onai T, Lin H-C, Schubert M, Koop D, Osborne PW, Alvarez S, Alvarez R, Holland ND, Holland LZ: **Retinoic acid and Wnt/β-catenin have complementary roles in anterior/posterior patterning embryos of the basal chordate amphioxus.** *Dev Biol* 2009, **332**:223–233.
189. Hotta K, Takahashi H, Ueno N, Gojobori T: **A genome-wide survey of the genes for planar polarity signaling or convergent extension-related genes in *Ciona intestinalis* and phylogenetic comparisons of evolutionary conserved signaling components.** *Gene* 2003, **317**:165–185.
190. Wagner E, Levine M: **FGF signaling establishes the anterior border of the *Ciona* neural tube.** *Development* 2012, **139**:2351–2359.
191. Kozmik Z, Holland ND, Kalousova A, Paces J, Schubert M, Holland LZ: **Characterization of an amphioxus paired box gene, *AmphiPax2/5/8*: developmental expression patterns in optic support cells, nephridium, thyroid-like structures and pharyngeal gill slits, but not in the midbrain-hindbrain boundary region.** *Development* 1999, **126**:1295–1304.
192. Meulemans D, Bronner-Fraser M: **The amphioxus SoxB family: implications for the evolution of vertebrate placodes.** *Int J Biol Sci* 2007, **3**:356–364.
193. Holland LZ, Kene M, Williams NA, Holland ND: **Sequence and embryonic expression of the amphioxus engrailed gene (*AmphiEn*): the metameric pattern of transcription resembles that of its segment-polarity homolog in Drosophila.** *Development* 1997, **124**:1723–1732.
194. Ikuta T, Saiga H: **Dynamic change in the expression of developmental genes in the ascidian central nervous system: revisit to the tripartite model and the origin of the midbrain–hindbrain boundary region.** *Dev Biol* 2007, **312**:631–643.

doi:10.1186/2041-9139-4-27
Cite this article as: Holland *et al.*: Evolution of bilaterian central nervous systems: a single origin? *EvoDevo* 2013 **4**:27.

Appendix 6

Published article

Roles of retinoic acid and Tbx1/10 in pharyngeal segmentation: amphioxus and the ancestral chordate condition.

Koop D, Chen J, Theodosiou M, Carvalho JE, Alvarez S, Lera AR de, Holland LZ, Schubert M.

EvoDevo. 2014;5(1):36

(18 pages)

Koop et al. EvoDevo 2014, 5:36
http://www.evodevojournal.com/content/5/1/36

EvoDevo

RESEARCH Open Access

Roles of retinoic acid and Tbx1/10 in pharyngeal segmentation: amphioxus and the ancestral chordate condition

Demian Koop[1], Jie Chen[2], Maria Theodosiou[2], João E Carvalho[3,4], Susana Alvarez[5], Angel R de Lera[5], Linda Z Holland[1*] and Michael Schubert[3,4*]

Abstract

Background: Although chordates descend from a segmented ancestor, the evolution of head segmentation has been very controversial for over 150 years. Chordates generally possess a segmented pharynx, but even though anatomical evidence and gene expression analyses suggest homologies between the pharyngeal apparatus of invertebrate chordates, such as the cephalochordate amphioxus, and vertebrates, these homologies remain contested. We, therefore, decided to study the evolution of the chordate head by examining the molecular mechanisms underlying pharyngeal morphogenesis in amphioxus, an animal lacking definitive neural crest.

Results: Focusing on the role of retinoic acid (RA) in post-gastrulation pharyngeal morphogenesis, we found that during gastrulation, RA signaling in the endoderm is required for defining pharyngeal and non-pharyngeal domains and that this process involves active degradation of RA anteriorly in the embryo. Subsequent extension of the pharyngeal territory depends on the creation of a low RA environment and is coupled to body elongation. RA further functions in pharyngeal segmentation in a regulatory network involving the mutual inhibition of RA- and Tbx1/10-dependent signaling.

Conclusions: These results indicate that the involvement of RA signaling and its interactions with Tbx1/10 in head segmentation preceded the evolution of neural crest and were thus likely present in the ancestral chordate. Furthermore, developmental comparisons between different deuterostome models suggest that the genetic mechanisms for pharyngeal segmentation are evolutionary ancient and very likely predate the origin of chordates.

Keywords: Cephalochordate, Cyp26 function, evolution of developmental mechanisms, evolution of the vertebrate head, functional knockdown, pharmacological treatments, pharyngeal patterning, retinoic acid signaling, Tbx1/10

Background

The evolution of head segmentation in chordates has been controversial for well over 150 years. In all chordates (vertebrates, tunicates and amphioxus), the pharynx is segmented into gill slits (aquatic chordates) or pouches (terrestrial chordates) and pharyngeal arches. The pharyngeal arches have a mesodermal core that derives from head mesoderm (or in amphioxus, the anteriormost somites) plus, in vertebrates only, a neural crest component that gives rise to pharyngeal cartilages. The mesoderm of the most anterior pharyngeal arch, the mandibular arch, gives rise to the velar muscle in the lamprey and to the jaw and other head muscles in gnathostomes. In addition to anatomical and fossil evidence [1-4], domains of gene expression have suggested homologies between the amphioxus and vertebrate pharynx as well as between the anterior somites of amphioxus and the head mesoderm of vertebrates. For example, *engrailed* is expressed in the posterior portion of each of the anterior somites of amphioxus and in the posterior wall of the mandibular head cavity and upper lip in the lamprey [5,6] as well as in the mandibular mesoderm of sharks [7] and the jaw muscles of the zebrafish [8]. Similarly, in amphioxus, *Tbx1/10* is expressed in the somites and in their ventral mesodermal extensions

* Correspondence: lzholland@ucsd.edu; michael.schubert@obs-vlfr.fr
[1]Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093-0202, USA
[3]Sorbonne Universités, UPMC Université Paris 06, CNRS, UMR 7009, Laboratoire de Biologie du Développement de Villefranche-sur-Mer, Observatoire Océanologique de Villefranche-sur-Mer, 06230 Villefranche-sur-Mer, France
Full list of author information is available at the end of the article

as well as in adjacent endoderm. In the lamprey, the gene is expressed in the mesenchyme of the upper lip, velar muscles and pharyngeal arches, while, in gnathostomes, *Tbx1/10* is detectable both in neural crest and head mesenchyme derivatives [9,10]. In addition, in both amphioxus and aquatic vertebrates, *Pax2/5/8* is expressed where the gill slits are forming as well as in the endostyle or its vertebrate homolog, the thyroid gland, and in the central nervous system (CNS), while *Pax1/9* genes are broadly expressed in the pharyngeal endoderm of all chordates [11-13].

We have previously defined an early phase of pharyngeal specification in amphioxus, which occurs during the gastrula stage and is regulated by retinoic acid (RA) signaling [14-17]. RA, a natural morphogen synthesized from vitamin A, binds to heterodimers of the retinoic acid receptor (RAR) and retinoid X receptor (RXR), allowing the complex to bind to regulatory regions of target genes and thereby activate transcription. The gene encoding the RA-degrading enzyme Cyp26-2 is expressed anterior to the anteriormost domain of *Tbx1/10*, indicating considerable reduction of RA signaling in the rostral head [18]. High RA signaling levels in the middle third of the endoderm of early embryos of amphioxus and vertebrates specify the midgut, while low levels of RA signaling in the pharynx specify the pharyngeal endoderm [14-16,19-22]. Thus, in amphioxus, exogenous RA applied during the gastrula stage causes loss of the pharynx and of all pharyngeal structures by respecifying the pharyngeal endoderm as midgut [14-17]. Similarly, in vertebrates, RA signaling is required for the formation of pharyngeal pouches caudal to the second pharyngeal arch. Thus, excess RA leads to a compression of the pharynx in lampreys [23,24] and to a fusion of the first two pharyngeal arches in gnathostomes, while the inhibition of RA signaling in gnathostomes, either genetically or by vitamin A deficiency, results in a loss of posterior pharyngeal structures [19-22,25-27].

In amphioxus, very few direct targets of RA signaling at the gastrula have been identified to date. Of more than 40 genes tested, the only direct targets were *Hox* genes (*Hox1*, *Hox3*), normally expressed in the dorsal/posterior mesendoderm and ectoderm, and *FoxA2-1*, normally expressed in the dorsal/posterior and anterior/ventral mesendoderm. The domains of all three genes are expanded anteriorly in RA-treated embryos [17]. Furthermore, knockdown of *Hox1* showed that it mediates the effect of RA in establishing the posterior limit of the pharynx [16]. Indirect targets include *Otx*, which is normally expressed in the anterior mesendoderm and dorsal/anterior ectoderm at the gastrula stage, and *Pax1/9*, which turns on in the pharyngeal endoderm at the very early neurula stage [11,17].

In normal embryos and larvae of amphioxus, pharyngeal structures are asymmetrically arrayed (Figure 1). By the early larval stage, from anterior to posterior, there is on the right, the endostyle, homologous to the thyroid, the club-shaped gland, a larval secretory structure that undergoes apoptosis at metamorphosis [28], and two gill slits. Additional gill slits form sequentially from anterior to posterior. On the left, the ciliated pit, which will form part of the homolog of the adenohypophysis, forms just anterior to the mouth, thought to be a modified gill slit [29]. The posterior limit of the pharynx in the early larva is at the same anterior/posterior position as the first photoreceptor and associated pigment cell that form in the nerve cord at the level of somite 5. The position of this pigment spot is unaffected by altered RA signaling even though excess RA shifts the *Hox1* domains in the CNS and endoderm anteriorly [30].

Morphological effects of excess RA added at the gastrula stage are first apparent in the early neurula. In untreated embryos, the pharyngeal endoderm marker *Pax1/9* is downregulated where the first gill slit will form [11]. As neurulation proceeds, *Pax1/9* is also downregulated in the primordium of the second gill slit. In embryos treated with RA from the gastrula stage, *Pax1/9* is not downregulated in the gill slit primordia, and the posterior limit of its domain is shifted anteriorly. Similarly, the posterior limit of the normally broad domain of *Otx* in the endoderm shifts anteriorly in RA-treated larvae. Thus, in embryos treated with 10^{-6} M RA at the gastrula stage, much of the pharyngeal endoderm is respecified as midgut [14,16]. Knockdown experiments showed that Hox1 acts upstream of *Otx* and *Pax1/9* in setting the posterior limit of the pharynx [16]. However, gill slits do form in most embryos when RA addition is delayed until the early neurula stage [14]. Even so, in untreated neurulae, the competitive RA signaling inhibitor *TR2/4* turns on in the gill slit primordia just before the gill slits begin to penetrate [15], suggesting that RA signaling may have to be continuously repressed during the neurula stage to ensure proper development of the gill slits.

To determine if in amphioxus (*Branchiostoma floridae*), an animal lacking definitive neural crest, RA affects pharyngeal morphogenesis post-gastrulation, we adopted a multi-pronged approach. We first manipulated RA signaling in amphioxus embryos from neurula through early larval stages and, as RA and Tbx1/10 mutually repress one another in vertebrates [10], compared the effects to those resulting from loss of Tbx1/10 function. In addition, we determined whether RA suffices to inhibit specification of pharyngeal endoderm or the formation of pharyngeal structures in amphioxus embryos and larvae with reduced Hox1 function. Finally, to comprehensively assess the roles of RA degradation in the developing anterior endoderm, we inhibited Cyp26 enzyme function starting at the gastrula stage. Taken together, our results show that RA signaling has both early and late effects on pharyngeal patterning in amphioxus. During the neurula stage, specification of the pharyngeal endoderm gradually becomes

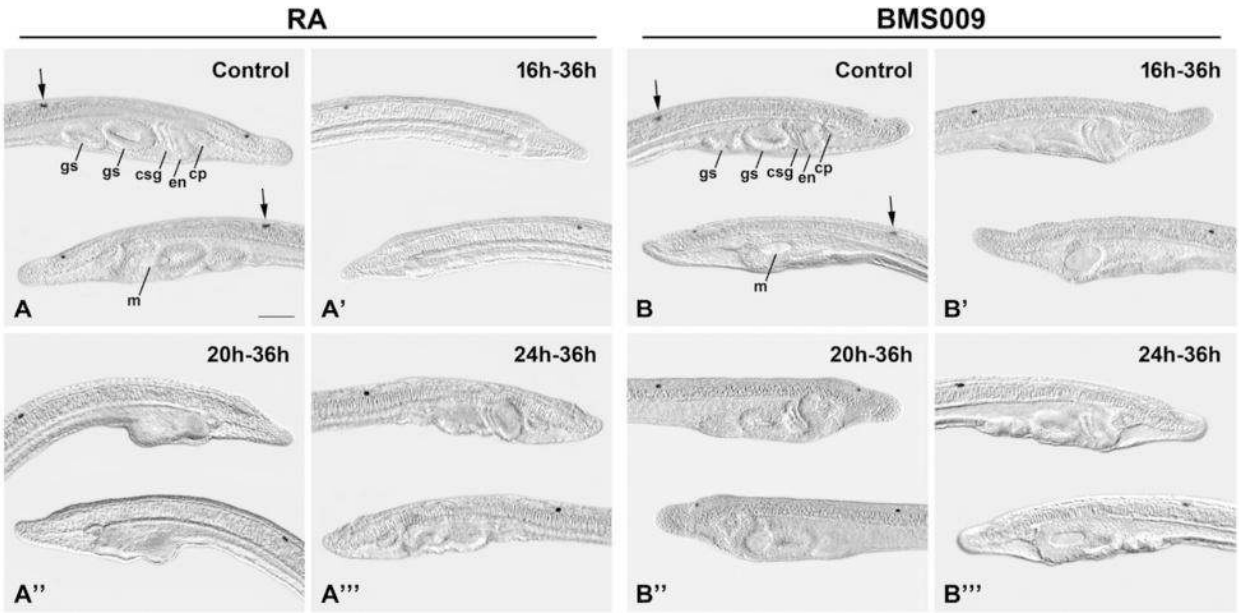


Figure 1 Exposure to retinoic acid (RA) or the RA receptor (RAR) antagonist BMS009 disrupts patterning and formation of the amphioxus pharynx. The anterior portions of amphioxus larvae at 36 hours post fertilization (hpf) are shown in lateral views. The right side of the larva is shown on top (with anterior to the right), with the two gill slits (gs), club-shaped gland (csg), endostyle (en) and ciliated pit (cp) in focus, and the left side is shown below (with anterior to the left), with the open mouth (m) in focus. Arrows highlight the first photoreceptor and associated pigment cell in the central nervous system. The 50-μm scale bar in A is applicable to the entire figure. Control larvae at 36 hpf are shown **(A)**, **(B)**, as are 36-hpf larvae treated with 10^{-6} M RA from 16 hpf **(A')**, 20 hpf **(A")** and 24 hpf **(A"')** or 2×10^{-6} M BMS009 from 16 hpf **(B')**, 20 hpf **(B")** and 24 hpf **(B"')**.

refractory to RA signaling. However, while Cyp26 functions in the anteriormost pharynx to keep RA levels low, partitioning of the pharynx into gill slits is regulated during the neurula stage by fine-tuning of RA levels. Knockdown of Tbx1/10 in the pharyngeal arches has a similar effect as adding RA during the neurula stage, indicating that, in the absence of neural crest, mutual inhibition of RA signaling and Tbx1/10 function is required for the partitioning of the amphioxus pharynx. These results show that the genetic mechanism involving RA and Tbx1/10 for partitioning the pharynx into pharyngeal pouches preceded the evolution of neural crest and was likely present in the ancestral chordate.

Methods

Embryo rearing, RA, RAR antagonist (BMS009) and Cyp26 inhibitor (R115866) treatments

Ripe males and females of the Florida amphioxus (*Branchiostoma floridae*) were collected by shovel and sieve in Tampa Bay, Florida (USA), during the summer breeding season. Spawning was induced electrically, and the embryos and larvae were cultured in the laboratory at 29°C as previously described [31]. Stock solutions of all-*trans* RA or the RAR antagonist BMS009, dissolved in dimethyl sulfoxide (DMSO), were added to cultures at different time points, 16 hours post fertilization (hpf), 20 hpf and 24 hpf, at final concentrations of 10^{-6} M and 2×10^{-6} M, respectively. Control treatments were 1:1,000 dilutions of DMSO alone [14,15]. Dishes were incubated in the dark, because RA and BMS009 are light sensitive. Samples were collected at 36 hpf and fixed for *in situ* hybridization as described below. Cyp26 function was inhibited using R115866 (provided by Janssen Research & Development, a division of Janssen Pharmaceutica NV, Beerse, Belgium). A 10^{-3} M stock solution in DMSO was added to embryo cultures at the onset of gastrulation (3.5 hpf) to a final concentration of 5×10^{-7} M, and embryos were fixed at 36 hpf.

Microinjection-based experiments

Microinjection of amphioxus eggs was performed as previously described [31]. Unfertilized eggs were injected with either a control antisense morpholino oligonucleotide (MO) (5'-CCTCTTACCTCAGTTACAATTTATA-3') or one specific for *AmphiHox1* from *B. floridae* (5'-ATTC TTGCCGTGTCCATTTGCTCCA-3') or *AmphiTbx1/10* from *B. floridae* (5'-ATAGCGGACTGTTGGCTTCCAT GTC-3') (Gene Tools, Philomath, OR, USA). The activity of both MOs was confirmed by *in vitro* translation assays of the *AmphiHox1* [16] and *AmphiTbx1/10* (Additional file 1: Figure S1) coding regions using the TnT Quick Coupled Transcription/Translation System (Promega, Madison, WI, USA) and a detection system based on the Transcend Non-Radioactive

Translation Detection Systems (Promega, Madison, WI, USA) [16]. Approximately 2 pl of a solution containing 15% glycerol, 2 mg/ml Texas Red dextran (Molecular Probes, Eugene, OR, USA) and 500 μM (*Hox1* and control) or 1,000 μM (*Tbx1/10* and control) MO was injected. Following injection, the eggs were fertilized, cultured and fixed at the early larval stage (36 hpf). For control MO/10^{-7} M RA and *Hox1* MO/10^{-7} M RA treatments, injected embryos were treated with RA at final concentrations of 10^{-7} M continuously from the onset of gastrulation (3.5 hpf). Fixed, injected embryos showing clear fluorescence of the Texas Red dextran were analyzed by *in situ* hybridization [32]. Injection of the control MO at both 500 μM and 1000 μM did not induce any abnormalities.

Developmental gene expression analyses using *in situ* hybridization

For *in situ* hybridization, samples were fixed according to established protocols [32]. Effects of treatments and MO injections on pharyngeal development were assayed by *in situ* hybridization with antisense riboprobes synthesized for the following genes: *AmphiPax1/9* (U20167) [11], *AmphiSix1/2* (EF195742) [13], *AmphiTbx1/10* (AF262562) [33], *AmphiPax2/5/8* (AF053762) [12], *AmphiPitx* (AJ438768) [34], *AmphiTR2/4* (AF378828) [15] and *AmphiCyp26-2* (EST clone bfne112a21). After *in situ* hybridization, the embryos were photographed as whole mounts using differential interference contrast (DIC) microscopy [32].

Results

RA signaling is required for both regional specification and morphogenesis of the pharynx

To determine if RA regulates development of the pharynx once it is initially specified, embryos were treated with either RA or the RAR antagonist BMS009 at the early/mid neurula (16 hpf), mid neurula (20 hpf) and late neurula (24 hpf) stage and subsequently fixed at the early larval stage (36 hpf). Given that the position of the first photoreceptor and associated pigment cell in the nerve cord is unaffected by changing RA signaling levels and that its anterior/posterior position coincides with the posterior limit of the pharynx of the early larva [30], we used this first pigment spot as a landmark to assess changes in pharyngeal length resulting from the treatments.

We found that excess RA added at progressively later stages resulted in progressively less severe effects on pharyngeal development (Figure 1). As treated embryos were kept in the dark to avoid degradation of RA, the effects we observed were more severe than in our previous study [14]. RA applied continuously from 16 hpf eliminated all pharyngeal structures (Figure 1A, A'; Figure 2A; Additional file 2: Table S1), indicating that the pharyngeal endoderm had been respecified as midgut (Figure 1A, A'). The ciliated pit was sometimes present (Figure 1A, A'). In embryos treated at 20 hpf, there was an endostyle and sometimes a club-shaped gland as well as a single gill slit primordium lacking a gill slit (Figure 1A, A"; Figure 2A; Additional file 2: Table S1). However, the posterior limit of the pharynx was still far anterior to the first photoreceptor in the CNS (Figure 1A). In embryos treated at 24 hpf, there was a mouth, an endostyle, a club-shaped gland and typically a single gill slit primordium, penetrated in about 50% of the larvae by an abnormal gill slit (Figure 1A, A"'; Figure 2A; Additional file 2: Table S1). A second gill slit primordium was occasionally present, but was less developed compared to controls.

Treatments with the RAR antagonist BMS009 at 16 hpf and 20 hpf resulted in a posterior expansion of the pharynx in 40 to 60% of the larvae. Most larvae had gill slit primordia, but about half of the associated gill slits were smaller than normal (Figure 1B, B', B"; Figure 2B; Additional file 2: Table S1). An endostyle and club-shaped gland were typically present, and, as in larvae treated with BMS009 from the gastrula stage, the mouth was often larger than normal (Figure 1B, B', B") [14]. When applied at 24 hpf, BMS009 did not affect the length of the pharynx. The club-shaped gland and endostyle were normal, and the gill slits and mouth were present, although sometimes the mouth was slightly enlarged (Figures 1B, B"'; Figure 2B; Additional file 2: Table S1).

The expression domains of *Pax1/9* and *Tbx1/10* were eliminated by RA treatment at 16 hpf, as were those of *Six1/2* and *Pax2/5/8*, both of which are normally expressed in the gill slit primordia (Figure 3A, B, D, E, G, H, J, K). Expression of *Pitx* around the mouth, in the ciliated pit and in the club-shaped gland was severely reduced (Figure 3M, N). In contrast, expression of *Cyp26-2* was expanded throughout the pharynx, suggesting that this gene is positively regulated by RA (Figure 3P, Q). When RA was applied at progressively later times, normal expression of all these genes was gradually restored. In embryos treated with RA at 20 hpf, *Pax1/9* and *Six1/2* were expressed in association with the single gill slit primordium, while *Tbx1/10* was expressed in a small area just anterior to the gill slit (Figure 3A, B', D, E', G, H'). Expression of *Pax2/5/8* was limited to the region around the mouth, while expression of *Pitx* in the club-shaped gland was unaffected (Figure 3J, K', M, N'). *Cyp26-2* expression was still expanded, but no longer extended posteriorly beyond the first pigment spot (Figure 3P, Q'). In larvae treated with RA from 24 hpf, most expression domains were normal (Figure 3A, B", D, E", G, H", J, K", M, N", P, Q"). However, *Tbx1/10* expression between the gill slits was still reduced (Figure 3G, H"), while *Pitx* expression remained restricted to the ciliated pit (Figure 3M, N"), and the domain of *Cyp26-2* was still expanded somewhat posteriorly (Figure 3P, Q").

Treatment with the RAR antagonist BMS009 had a milder effect than exogenous RA. Although BMS009 expanded

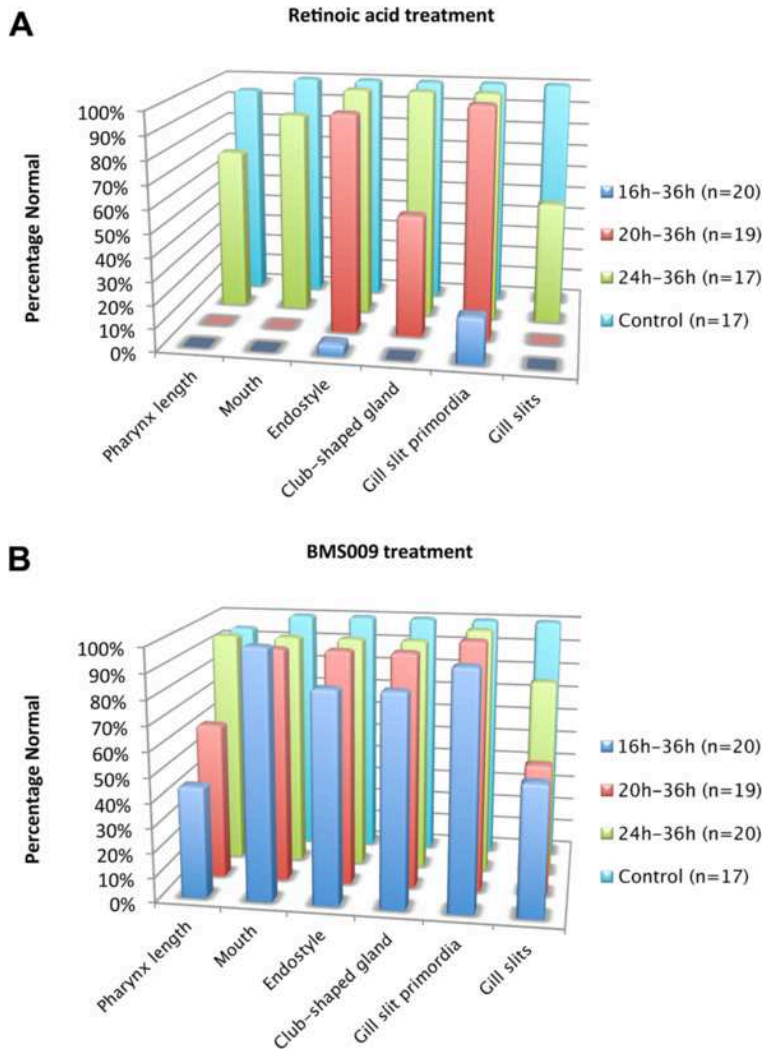


Figure 2 Stage-dependent effects of retinoic acid (RA) and RA receptor (RAR) antagonist (BMS009) on the formation of pharyngeal structures in amphioxus. For each feature, the percentage of normal larvae at 36 hours post fertilization (hpf) is indicated. Animals were treated with RA **(A)** or BMS009 **(B)** from 16 hpf, 20 hpf or 24 hpf.

expression of *Pax1/9* slightly posteriorly in the pharyngeal endoderm at all times of treatment (Figure 3A, C, C', C"), expression of *Six1/2* was largely unaffected (Figure 3D, F, F', F"). *Tbx1/10* expression was abnormal at all time points, in agreement with the gill slits being misshapen (Figure 3G, I, I', I"). Similarly, *Pax2/5/8* expression was reduced in the gill slit primordia of larvae treated at 16 hpf, 20 hpf and 24 hpf, although the domain around the mouth was unaffected (Figure 3J, L, L', L"). Expression of *Pitx* was largely unaffected by BMS009 (Figure 3M, O, O', O"). In contrast, *Cyp26-2* expression was severely reduced, although not completely eliminated, by BMS009, which suggests that RA positively regulates *Cyp26-2* (Figure 3P, R, R', R").

Taken together, these results indicate that RA signaling is not only required in early development for regionalization of the endoderm, but is also necessary during the neurula stage for patterning gill slits and mouth. Thus, there appears to be both an early and a late phase for RA signaling: the first requiring low levels for pharyngeal specification and the second requiring localized regulation of RA for patterning within the pharynx.

Cyp26 activity is required for pharyngeal patterning in amphioxus

Since a low level of RA signaling is required in the amphioxus pharynx for normal specification and patterning, endogenous RA levels must be very tightly regulated. Cyp26 enzymes function as RA sinks by degrading RA into biologically inactive metabolites [35]. Of the three duplicates of amphioxus *Cyp26* [36], one, *Cyp26-2*, is expressed in embryos and early larvae in the anterior CNS, ectoderm, mesoderm and endoderm as well as at the extreme

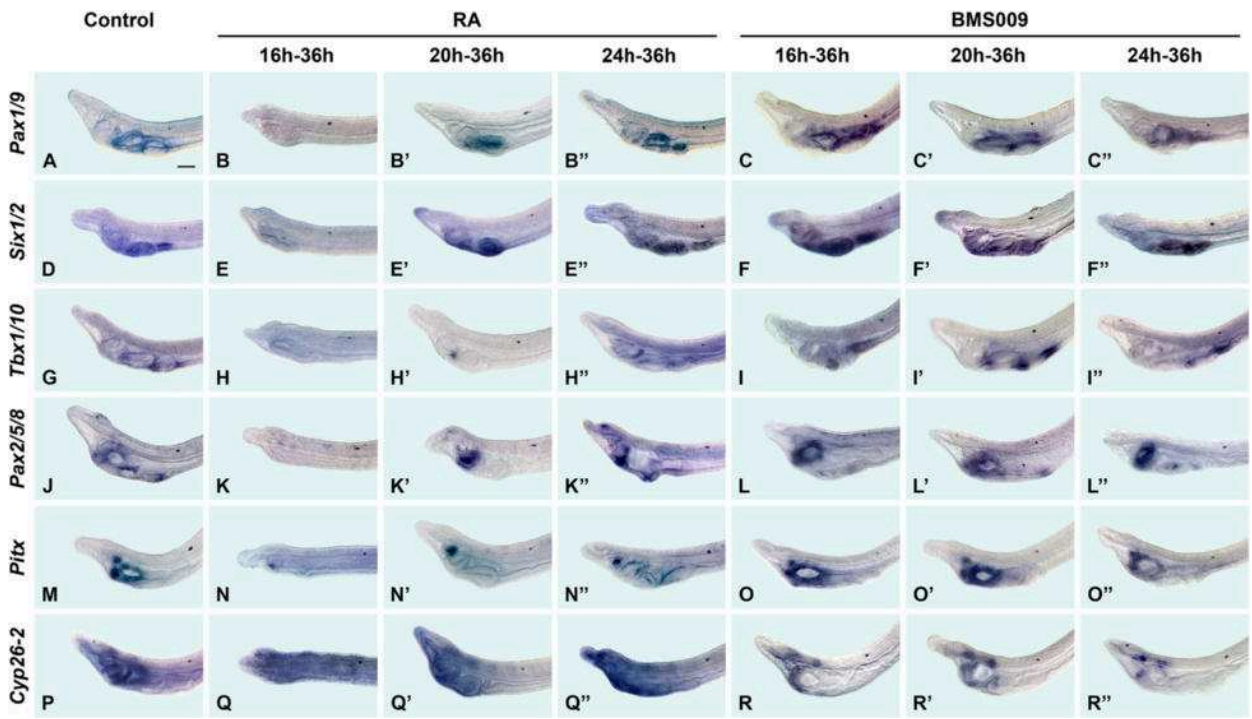


Figure 3 Molecular characterization of retinoic acid (RA) signaling functions during amphioxus pharyngeal development. Lateral views of the anterior portions of amphioxus larvae at 36 hours post fertilization (hpf) are shown with the anterior to the left. The 50-μm scale bar in **A** is applicable to the entire figure. Control larvae are shown as well as those treated with either 10^{-6} M RA or 2×10^{-6} M BMS009, an RAR antagonist, starting at three different developmental stages (16 hpf, 20 hpf and 24 hpf). The assayed genes are amphioxus *Pax1/9* **(A)**, **(B)**, **(B')**, **(B")**, **(C)**, **(C')**, **(C")**, *Six1/2* **(D)**, **(E)**, **(E')**, **(E")**, **(F)**, **(F')**, **(F")**, *Tbx1/10* **(G)**, **(H)**, **(H')**, **(H")**, **(I)**, **(I')**, **(I")**, *Pax2/5/8* **(J)**, **(K)**, **(K')**, **(K")**, **(L)**, **(L')**, **(L")**, *Pitx* **(M)**, **(N)**, **(N')**, **(N")**, **(O)**, **(O')**, **(O")** and *Cyp26-2* **(P)**, **(Q)**, **(Q')**, **(Q")**, **(R)**, **(R')**, **(R")**. RA treatment at 16 hpf eliminates the expression of *Pax1/9*, *Six1/2*, *Tbx1/10* and *Pax2/5/8* and severely reduces that of *Pitx*. In contrast, expression of *Cyp26-2* is expanded. When RA is applied at progressively later stages, normal expression of the genes is gradually restored. Treatments with BMS009 result in an expansion of the *Pax1/9* domain, in abnormal *Tbx1/10* expression, in a reduction of the *Pax2/5/8* signal in gill slit primordia and in a severe reduction of *Cyp26-2* expression. In contrast, the patterns of *Six1/2* and *Pitx* are largely unaffected. The effects of BMS009 are independent of the treatment time point.

posterior of the animal [18]. At 36 hpf, *Cyp26-2* is broadly expressed anteriorly in the anterior somites, the ectoderm surrounding the mouth, the anterior/ventral endoderm, the endostyle and the club-shaped gland (Figure 3P). The domain of *Cyp26-2* (Figure 3P) in the pharyngeal endoderm is just anterior to that of *Tbx1/10* in the first pharyngeal arch (Figure 3G).

To assess the importance of Cyp26 enzymes in amphioxus pharyngeal development, we treated amphioxus embryos from the onset of gastrulation with the Cyp26 inhibitor R115866 [37]. Similar to treatments with RA, inhibition of Cyp26 caused a dramatic shortening of the pharynx and loss of pharyngeal structures: while 21 of 21 control larvae developed normally, 35 of 35 treated larvae showed pharyngeal abnormalities. Lost structures included the mouth, the endostyle, the club-shaped gland and one gill slit; however, a single gill slit primordium was present (Figure 4). *Pax1/9* expression was compressed anteriorly, being restricted chiefly to the level of the gill slit primordium (Figure 4A, A'), as were the domains of *Six1/2* (Figure 4B, B') and *Pax2/5/8* (Figure 4D, D'). *Tbx1/10* was weakly expressed anterior and posterior to the single gill slit primordium (Figure 4C, C'), while *Pitx* was only expressed in the ciliated pit (Figure 4E, E').

Together, these data indicate that inhibition of Cyp26 function leads to the loss of anterior pharyngeal structures, including the mouth, the endostyle, the club-shaped gland and one gill slit. These findings suggest that repression of RA signaling by Cyp26 activity is required to protect the anteriormost part of the amphioxus pharynx from RA teratology.

RA-dependent regionalization of the endoderm requires Hox1 function, while RA-dependent pharyngeal morphogenesis does not

To separate RA signaling functions in endoderm regionalization from those in pharyngeal morphogenesis, we assessed the effects of RA on pharyngeal development in a Hox1-reduced environment. As noted above, Hox1 mediates the roles of RA in establishing the posterior limit of the pharynx [16,17]. Knockdown of Hox1 function results in a posterior expansion of the pharynx, which

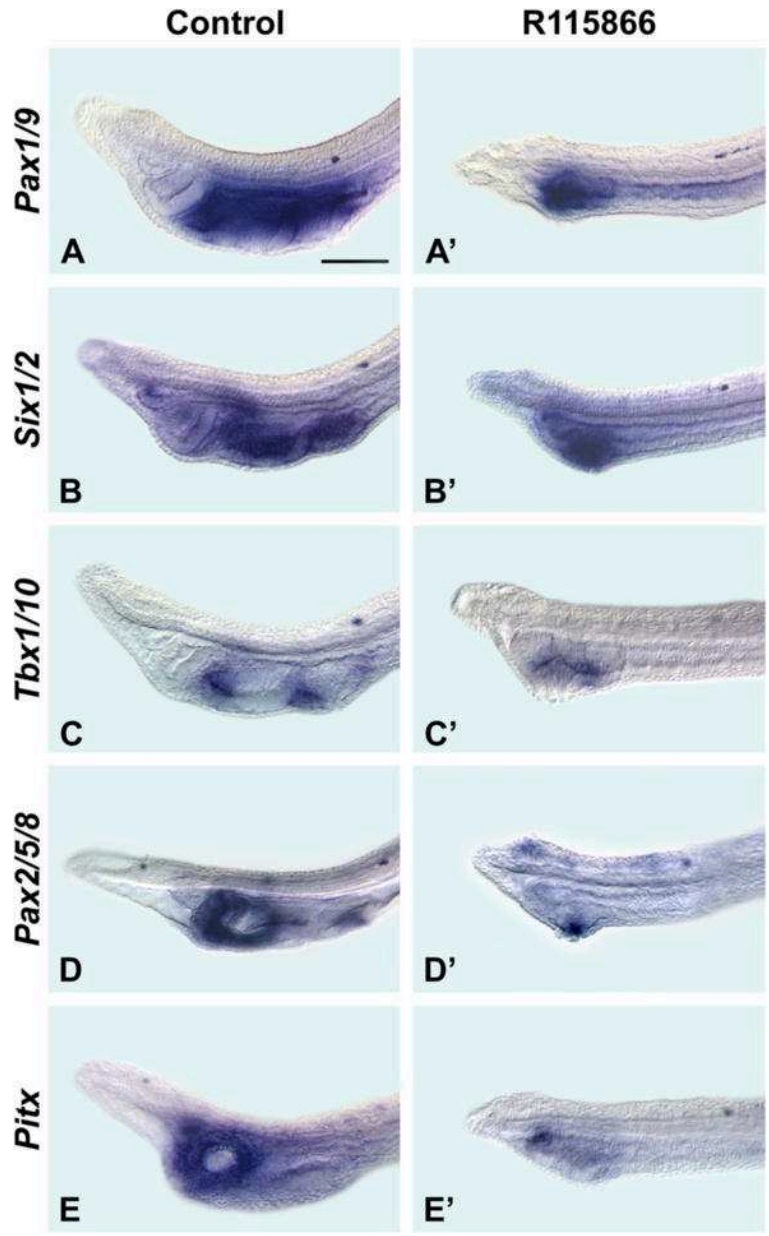


Figure 4 Cyp26 function is required for patterning the amphioxus pharynx. The anterior portions of larvae at 36 hours post fertilization (hpf) are shown in lateral views (with anterior to the left). The 50-μm scale bar in **(A)** is applicable to the entire figure. Control larvae are shown as well as those treated with 5×10^{-7} M R115866 (a Cyp26 inhibitor) at the onset of gastrulation. The assayed genes are amphioxus *Pax1/9* **(A)**, **(A')**, *Six1/2* **(B)**, **(B')**, *Tbx1/10* **(C)**, **(C')**, *Pax2/5/8* **(D)**, **(D')** and *Pitx* **(E)**, **(E')**.

is similar to the effect obtained by treatment with BMS009 at the gastrula stage [16]. Combination of RA treatments with the knockdown of Hox1 function may thus help reveal roles of RA signaling in pharyngeal development, which are independent of its initial, Hox1-mediated role in establishing the posterior limit of the pharynx.

In larvae injected with the control MO and treated with 10^{-7} M RA, the pharynx was evidently specified, but gill slits did not form and *Pax1/9* expression was downregulated only in a small domain ventrally in the pharynx (Figure 5A, A'). In larvae with combined *Hox1* MO injections and 10^{-7} M RA treatment, there were two ventral zones of reduced *Pax1/9* expression, indicating that two gill slit primordia were specified (Figure 5A, A"). In embryos injected with the control MO and treated with 10^{-7} M RA, there was a single domain of *Six1/2* ventrally in the pharyngeal endoderm, suggesting that the two gill slit primordia were fused (Figure 5B, B'). This domain was expanded posteriorly in embryos injected with *Hox1* MO and treated with 10^{-7} M RA. However, in these larvae, the signal

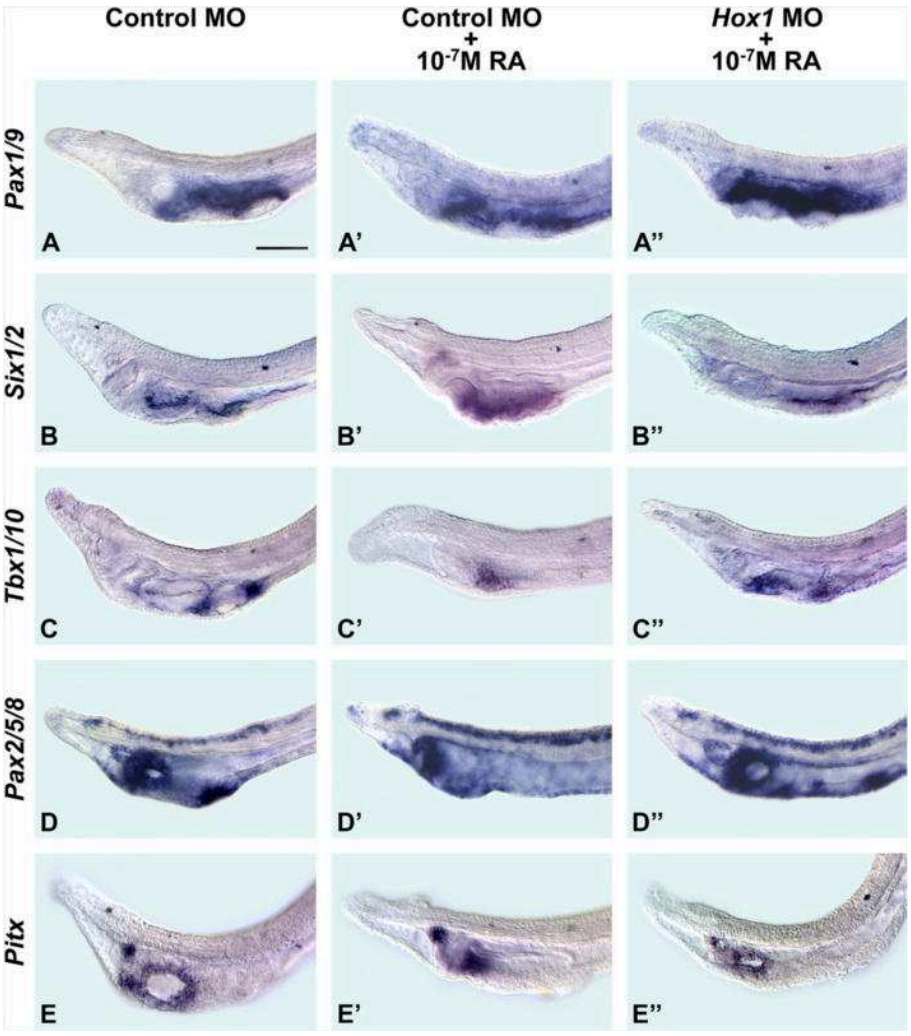


Figure 5 Amphioxus pharyngeal development depends on retinoic acid (RA) signaling for both anterior-posterior regionalization and morphogenesis. The anterior portions of larvae at 36 hours post fertilization (hpf) are shown in lateral views (with anterior to the left). The 50-μm scale bar in **(A)** is applicable to the entire figure. Amphioxus larvae are shown that were injected before fertilization with either 500 μM control antisense morpholino oligonucleotide (MO) or with 500 μM control MO followed by a treatment with 10^{-7} M RA (at the onset of gastrulation) or with 500 μM *Hox1* MO followed by treatment with 10^{-7} M RA (at the onset of gastrulation). The assayed genes are amphioxus *Pax1/9* **(A)**, **(A')**, **(A")**, *Six1/2* **(B)**, **(B')**, **(B")**, *Tbx1/10* **(C)**, **(C')**, **(C")**, *Pax2/5/8* **(D)**, **(D')**, **(D")** and *Pitx* **(E)**, **(E')**, **(E")**.

was still continuous and not in two discrete patches, suggesting that the two gill slit primordia remained fused (Figure 5B"). Expression of *Tbx1/10* between the gill slits was also disrupted by RA (Figure 5C, C'), while injection of *Hox1* MO and treatment with 10^{-7} M RA resulted in ectopic expression of *Tbx1/10* ventrally in the pharynx (Figure 5C, C").

Treatments with 10^{-7} M RA reduced the *Pax2/5/8* signal around the mouth and in the gill slits, but not in the endostyle or CNS (Figure 5D'). Combined *Hox1* MO injection and RA treatment largely restored the mouth- and gill slit-associated domains. In fact, the domain of *Pax2/5/8* in the gill slit region was considerably expanded (Figure 5D"). In RA-treated larvae, expression of *Pitx* was restricted to the ciliated pit plus a small cluster of ectodermal cells, indicating that 10^{-7} M RA is sufficient to suppress mouth formation (Figure 5E, E'), as has previously been described [14,16]. In contrast, in *Hox1* MO/10^{-7} M RA larvae, *Pitx* expression surrounds a relatively small mouth (Figure 5E").

Importantly, in larvae treated with 10^{-7} M RA, which induced severe pharyngeal defects in only 31 of 50 control MO-injected larvae, the injection of *Hox1* MO seemed to preferentially rescue the expression of pharyngeal regionalization markers (in 13 of 13 larvae), while expression of genes marking specific pharyngeal structures tended to be only partially restored (in 23 of 34 larvae). Injection of the control MO alone did not induce any developmental abnormalities (in 45 of 50 larvae). Altogether, these experiments are in agreement with the idea that, in

amphioxus, RA signaling initially acts via Hox1 to control anterior-posterior regionalization of the endoderm and subsequently assumes roles in pharyngeal morphogenesis that are independent of Hox1.

Tbx1/10 is involved in segmentation of the pharyngeal endoderm and patterning of the gill slit primordia in amphioxus

Since, in amphioxus, *Tbx1/10* is expressed in mesoderm and endoderm between adjacent gill slits [33], we knocked down *Tbx1/10* function to test whether it might be involved in RA-dependent segmentation of the pharynx. Larvae injected with a *Tbx1/10* MO exhibited a smaller pharynx and severely malformed and fused gill slit primordia (for embryos injected with the control MO, 10 of 10 were normal, while 22 of 30 embryos injected with the *Tbx1/10* MO had pharyngeal abnormalities) (Figure 6). The domain of *Pax1/9* expression was reduced, which was largely due to the shorter pharynx (Figure 6A, A'). The *Six1/2* signal was disorganized, in agreement with abnormal gill slit primordia that are elongated and possibly fused (Figure 6B, B'). This elongation and fusion is most apparent in larvae labeled with *Pax2/5/8*, as it marks where the gill slits will form (Figure 6C, C'). Expression of *Pax2/5/8* in the endostyle and mouth is not altered in *Tbx1/10* MO-injected larvae (Figure 6C, C'). *Pitx* expression remains detectable around the open mouth as well as in the club-shaped gland (Figure 6D, D'). In contrast, pharyngeal expression of *Cyp26-2* is disorganized and somewhat reduced in *Tbx1/10* MO-injected larvae (Figure 6E, E'), as is that of *TR2/4* (Figure 6F, F'), which is normally expressed in the pharyngeal endoderm, most conspicuously in mouth and gill slit primordia [15]. Importantly, after *Tbx1/10* knockdown, *TR2/4* expression remains detectable in the mouth as well as in the primordia of the abnormal gill slits (Figure 6F, F').

These data indicate that Tbx1/10 functions in separating gill slit primordia in the developing amphioxus pharynx. The phenotypes obtained from *Tbx1/10* knockdown are reminiscent of those obtained by RA treatments in a *Hox1* knockdown context, suggesting a mutual inhibition of RA signaling and Tbx1/10 function in amphioxus pharyngeal development, a feature previously proposed for vertebrate pharyngeal arch and cardiovascular development [10].

Discussion

RA signaling functions during distinct phases of amphioxus pharyngeal development

Our results indicate that, in amphioxus, RA signaling must be suppressed in the anterior portion of the gastrula for proper specification of the pharynx and that it must also be finely regulated during the neurula stage for proper partitioning of the pharynx into pharyngeal arches and pouches (Figure 7). RA signaling is controlled at several levels. The first is synthesis. Amphioxus has seven duplicates of *Aldh1/2* genes, at least some of which probably encode RA synthesizing enzymes. At the neurula stage, they are all expressed in the posterior half of the embryo [38]. However, amphioxus embryos are quite small, and RA readily diffuses through cell membranes. Therefore, RA is most likely able to diffuse anteriorly in the embryo. The second level of RA regulation is degradation by the three Cyp26 enzymes. While *Cyp26-2* is conspicuously expressed in extreme anterior and posterior tissues, *Cyp26-1* and *Cyp26-3* expression is chiefly limited to the anterior somites [18]. We found that inhibition of Cyp26 function leads to a loss of pharyngeal structures, most severely affecting the anteriormost pharynx. Finally, the third level of RA regulation is local inhibition. Our data suggest that Tbx1/10 and RA mutually inhibit each other. Thus, *Tbx1/10* turns on in mesoderm migrating into the pharyngeal arches and in adjacent pharyngeal endoderm, indicating that RA signaling must be locally suppressed for segmentation of the pharynx into pouches and arches. Later, *TR2/4*, a competitive inhibitor of RA signaling, turns on in gill slit primordia, indicating that further suppression of RA signaling in the gill slit primordia is required for gill slit penetration.

This work further allows us to propose a model describing the activity of RA signaling during amphioxus pharyngeal development (Figure 8). Following this model, RA signaling activity is high in the posterior half of the developing embryo, while the anterior portion is characterized by Cyp26 expression and hence by very low levels of RA. The endoderm is thus divided into anterior and posterior domains, with Cyp26 protecting the anterior domain from RA activity and with RA signaling patterning the posterior domain, including its regionalization, by defining the posterior limit of the pharynx. Expansion of the pharyngeal territory as development proceeds is accomplished through the elongation of the embryo, which concomitantly increases the distance between the posterior tip of the pharynx and the tissues producing RA and expressing RAR. This movement of RA signaling activity in the embryo hence creates a permissive, low RA environment at the posterior extremity of the pharynx allowing the patterning and formation of additional pharyngeal structures, such as gill slits. Importantly, a similar model for the activity of RA signaling has been proposed for the regulation of head mesoderm patterning in the chick embryo [39], strongly suggesting that regionalization and segmentation of the pharynx of the last common ancestor of all chordates was already dependent on RA signaling, possibly by employing an amphioxus-like molecular patterning mechanism.

It would be useful to know if our model for the first three gill slits is also applicable to the ones which subsequently form. However, it would be difficult at best to perform

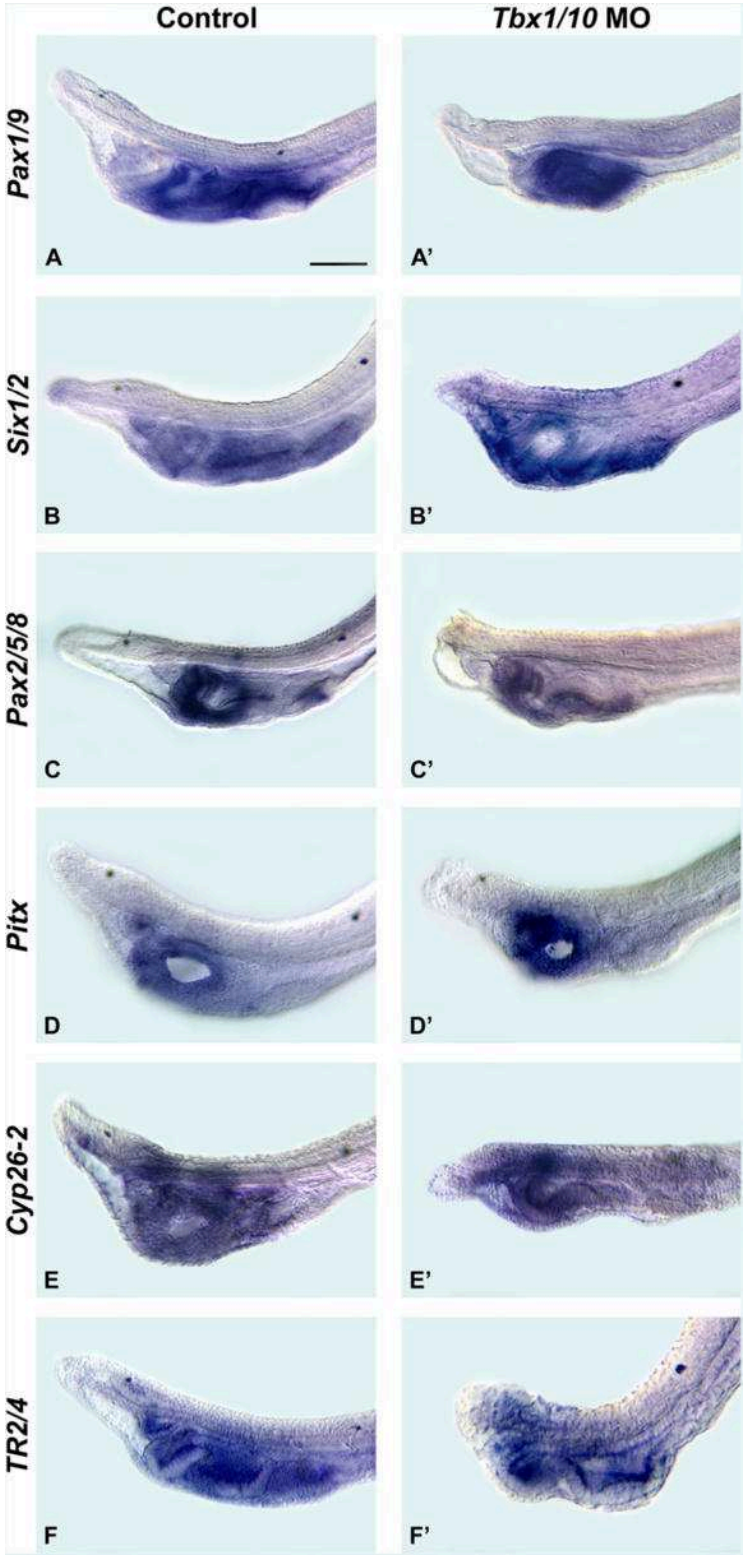


Figure 6 (See legend on next page.)

(See figure on previous page.)
Figure 6 Tbx1/10 function is required for gill slit formation in the amphioxus pharynx. The anterior portions of larvae at 36 hours post fertilization (hpf) are shown in lateral views (with anterior to the left). The 50-μm scale bar in **(A)** is applicable to the entire figure. Amphioxus larvae are shown that were injected before fertilization with either 1,000 μM control antisense morpholino oligonucleotide (MO) or with 1,000 μM *Tbx1/10* MO. The assayed genes are amphioxus *Pax1/9* **(A)**, **(A')**, *Six1/2* **(B)**, **(B')**, *Pax2/5/8* **(C)**, **(C')**, *Pitx* **(D)**, **(D')**, *Cyp26-2* **(E)**, **(E')** and *TR2/4* **(F)**, **(F')**.

comparable experiments to test this. The first three gill slits of amphioxus form during the first two days of development. By 30 hpf to 36 hpf, the young larvae begin feeding [31]. However, even if not fed, they will develop normally until at least 48 hpf. Therefore, even if MO-induced gene knockdowns or manipulation of signaling pathways cause abnormal development of the gill slits and/or mouth and prevent the larvae from feeding, they will keep developing for the duration of our experiments. Once the first three gill slits develop, there is typically a lag of 2 days before the larvae begin adding more gill slits posterior to the first three. During this lag, the animals grow considerably. Starting about 4 days hpf, gill slits are added sequentially for about 2 to 3 weeks until there is a total of nine to eleven gill slits. At that point metamorphosis ensues. During metamorphosis a second row of gill slits appears on the right side dorsal to the first row, and the first row of gill slits migrates to the left side at the same time as the mouth migrates anteriorly [28,29].

When we added RA after formation of the first three gill slits, additional gill slits did not form, but as the existing gill slits collapsed and the anus seemed to close, the animals stopped feeding (data not shown). The very first thing that happens, whenever amphioxus larvae are starved or poisoned even slightly is that all the gill slits collapse and the larvae cease feeding and cease growing. Consequently, we could not determine whether the failure of additional gill slits to form in the RA-treated larvae was due to direct effects of RA or due to starvation. These considerations severely limit the types of experiments that can be done in amphioxus to investigate whether the molecular mechanisms of gill slit formation are conserved between the first three gill slits and the ones that form later.

Pharyngeal patterning in deuterostomes is controlled by conserved genetic mechanisms

A segmental series of gill slits is present not only in chordates, but also in hemichordates among the deuterostomes

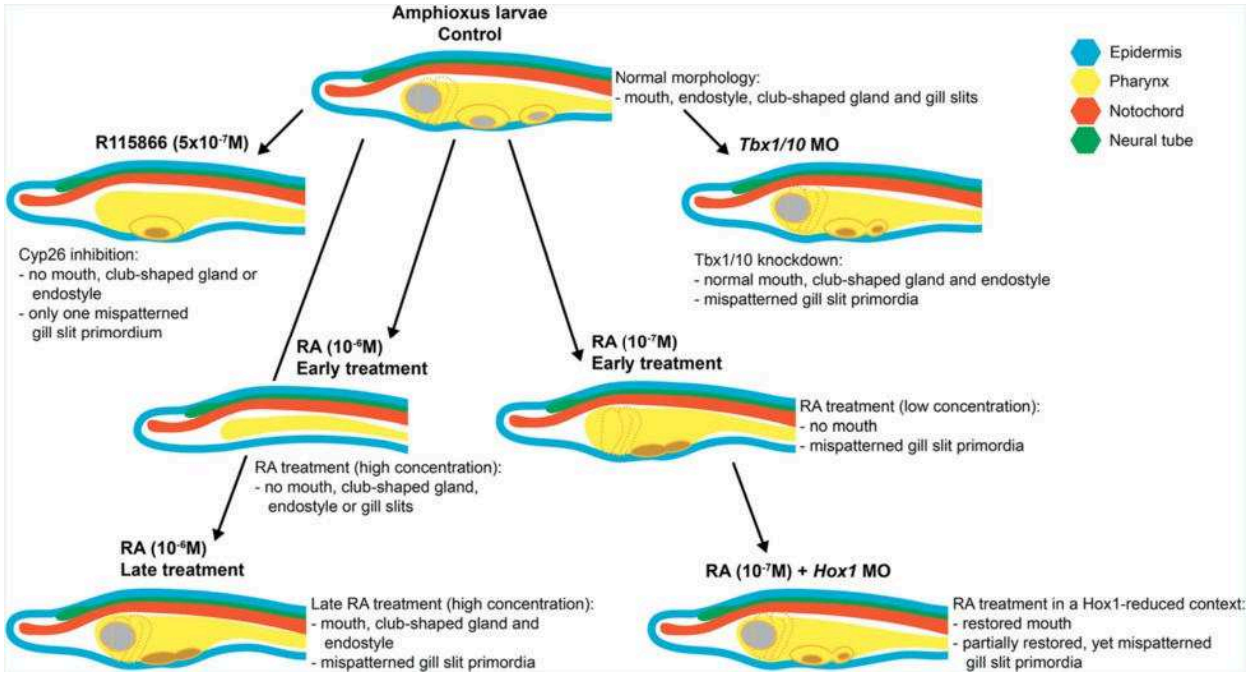


Figure 7 Diagrammatic summary of the morphological defects induced by pharmacological treatments and gene knockdown in the developing amphioxus pharynx. The effects are described in the pharynx of larvae at 36 hours post fertilization (hpf), which are characterized by a well-developed mouth (gray circle), an endostyle (anterior dotted lines), a club-shaped gland (posterior dotted lines) and two gill slit primordia (yellow ovals) with two gill slits (gray oval). Indicated are the results obtained from treatments with the Cyp26 inhibitor R115866, with retinoic acid (RA) at different developmental stages and at different concentrations, from injections of a *Tbx1/10*-specific antisense morpholino oligonucleotide (MO) and from the combination of RA treatments with the injection of a *Hox1*-specific MO. Mispatterned gill slit primordia with malformed gill slits are indicated as oval structures with brown centers. The results obtained with the retinoic acid receptor (RAR) antagonist BMS009 are not shown.

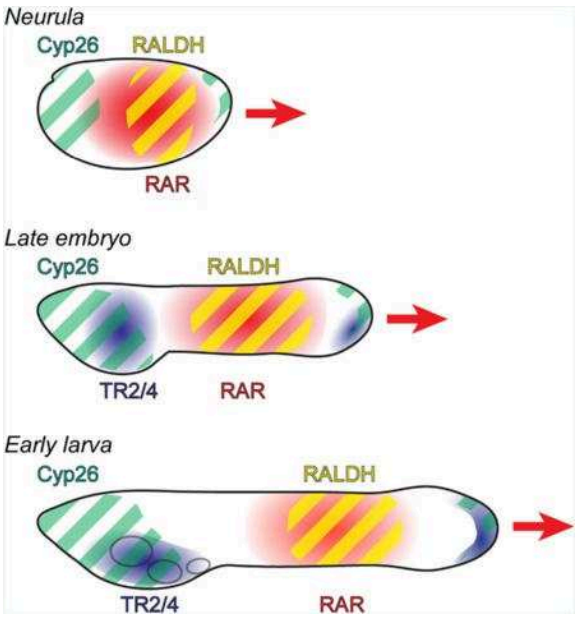


Figure 8 Stage-dependent functions of retinoic acid (RA) signaling during amphioxus pharyngeal development. Expression of the main mediators of RA signaling in the developing endoderm is indicated at three developmental stages: the neurula, the late embryo and the early larva. Shown are the domains of retinoic acid receptor (RAR), RA synthesizing enzymes (RALDH), RA degradation enzymes (Cyp26) and TR2/4, a competitive inhibitor of RA signaling. Posterior elongation of the embryo and larva in the course of development is indicated with a red arrow. As the embryo and larva elongate, the source of endogenous RA as well as the expression domain of RAR move posteriorly, hence creating a permissive environment for the patterning and development of additional pharyngeal structures, chiefly gill slits.

and has been described in some fossil echinoderms [40]. Comparisons of gene expression patterns indicate that at least some of the molecular mechanisms for pharyngeal specification and gill slit formation are conserved between hemichordates and chordates (Figure 9). For example, in both hemichordates and chordates, the pharyngeal endoderm expresses *Pax1/9*, *Six1/2* and *Eya* [13,41-44]. In the hemichordate *Saccoglossus kowalevskii*, as in amphioxus, *Hox1* is expressed just posterior to the third primary gill slit [43], suggesting the possibility that RA signaling might act via Hox1 to establish the posterior limit of the gill slits. Hemichordate genomes encode the retinoid receptors RAR and RXR as well as homologs of vertebrate RA synthesizing (RALDH) and degrading (Cyp26) enzymes [45], but their expression and functions are still unknown. The main difference between pharyngeal patterning in hemichordates and chordates is that *Tbx1/10* is not expressed in the developing gill bars of hemichordates [43].

Comparisons between amphioxus, ascidian tunicates and vertebrates indicate that amphioxus has retained the fundamental mechanism for specification of the chordate pharynx and for partitioning it into gill slits and pharyngeal arches. For example, in both amphioxus and vertebrates, high levels of RA signaling in the middle third of the embryo establish the posterior limit of the pharynx [14-16]. Interestingly, the pharynx of juvenile ascidian tunicates might be patterned by a similar mechanism, as excess RA applied during postlarval development leads to a graded loss of the juvenile pharynx by respecification of anterior endoderm to more posterior fates [46]. Furthermore, in amphioxus, *Hox1*, which is expressed in the pharyngeal endoderm just posterior to the gills, is directly regulated by RA signaling and its knockdown causes the posterior limit of the pharynx to expand posteriorly [16]. Although this regulation is likely not conserved with ascidian tunicates [45,47], *Hox1* expression has been described in the esophagus and posterior intestine of ascidian tunicate juveniles [47]. Taken together, the genetic mechanisms for pharyngeal specification and patterning were probably already present in the last common ancestor of all chordates and have been secondarily modified in the lineage leading to extant tunicates.

Low RA signaling levels are required for pharyngeal segmentation and gill slit formation

In amphioxus, we found that, while adding RA at progressively later stages during the neurula causes progressively less truncation of the pharynx, the gill slits are still abnormal when RA is added at the very late neurula (24 hpf) (Figure 7). Excess RA at the mid neurula stage reduced the number of gill slit primordia, as indicated by loss of the *Six1/2* domains, while, when added at the very late neurula, RA results in strong reduction or loss of *Tbx1/10* and *Pax2/5/8* expression in the gill bars and gill slit primordia, respectively. Furthermore, when Hox1 activity is reduced, exogenous RA specifically disrupts gene expression associated with the gill slit primordia. Altogether, these data demonstrate that RA signaling must be kept low for gill slits to form in amphioxus and suggest that this function of RA is independent of Hox1. Similarly, while adding RA to lamprey embryos at the gastrula stage causes truncation of the pharynx, delaying addition until the neurula stage restores the third through seventh pharyngeal pouches, but nonetheless results in an anterior-posterior compression of the pharynx, which is accompanied by occasional fusion of adjacent arches as well as a likely absence of the endostyle [23,24].

RA signaling interacts with Tbx1/10 and Pitx in both the amphioxus and vertebrate pharynx. In amphioxus, exogenous RA applied during the neurula stage disrupts pharyngeal segmentation and reduces *Tbx1/10* expression in the pharyngeal arches and *Pitx* expression around the mouth. Not surprisingly, inhibition of Cyp26 enzymes also downregulates *Tbx1/10* and causes the loss of a gill slit primordium. Similarly, in the lamprey, RA added during the

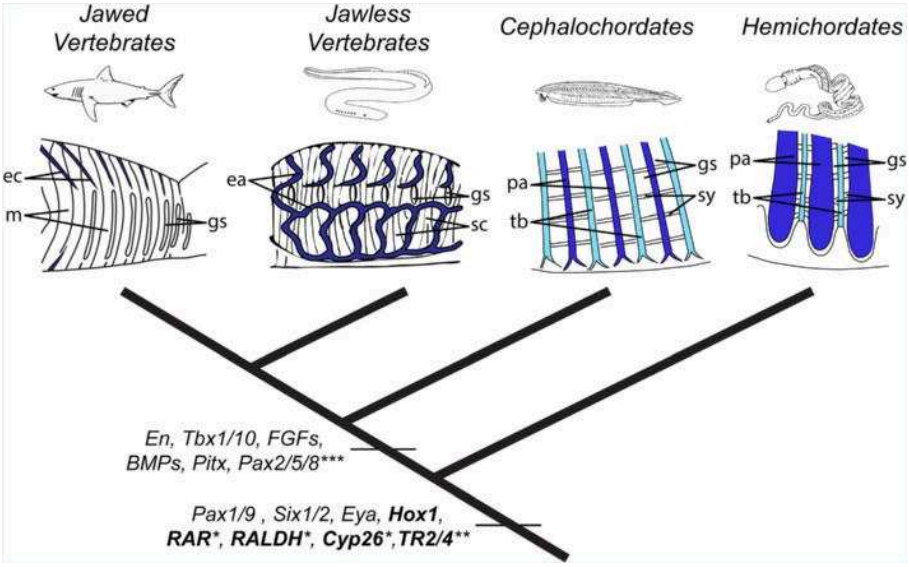


Figure 9 Evolution of pharyngeal patterning in deuterostomes. The last common ancestor of all deuterostomes possessed a segmented pharynx with alternating pharyngeal arches and tongue bars and with an extracellular collagenous skeleton secreted by endodermal cells. This ancestral pharynx was patterned by *Pax1/9*, *Six1/2* and *Eya* genes, with *Hox1* likely establishing the posterior boundary of the pharynx. The involvement of RA signaling in pharyngeal patterning in hemichordates is possible, but further work is required to validate this hypothesis. In the ancestral chordate, the pharyngeal apparatus was elaborated by *Tbx1/10*-expressing mesoderm originating from the anterior paraxial mesoderm. *Engrailed* (*En*) was co-expressed with *Tbx1/10* in the anterior somites and the gene network controlling chordate pharyngeal patterning further included *Pitx* and *Pax2/5/8* as well as fibroblast growth factor (FGF) and bone morphogenetic protein (BMP) signaling. In the vertebrate lineage, neural crest cells contribute extensively to the developing pharynx and segmentation of the head mesoderm was likely lost. Genes shown in bold are directly involved in mediating RA signaling in the pharynx. *Pharyngeal expression unknown in hemichordates; **Pharyngeal expression has not been assessed in hemichordates and vertebrates; ***Pharyngeal expression lost in terrestrial mammals. Abbreviations: ea, external branchial arches; ec, extrabranchial cartilage; gs, gill slits; mu, muscles; pa, pharyngeal arches; sc, superficial branchial constrictions; sy, synapticles; tb, tongue bar. For the sake of clarity, tunicates have not been included in the diagram.

neurula stage reduces the overall number of pharyngeal arches and induces fusion of adjacent arches [23,24]. Expression of lamprey *Tbx1/10* and *Pitx* genes was not determined in these embryos. The effects of RA on pharyngeal segmentation have not yet been assessed in shark embryos. For other gnathostomes, *Cyp26c1* is expressed in the anterior head mesoderm of chicken [39] and its inhibition downregulates *Tbx1*. Furthermore, the caudal pharyngeal arches are lost and the pharyngeal endoderm does not segment correctly [48]. Addition of RA has a similar effect, inhibiting *Tbx1* expression, indicating that low RA signaling levels are required for *Tbx1* expression.

Regulation of pharyngeal segmentation is conserved in amphioxus and vertebrates

The present study shows that RA signaling together with Tbx1/10 also has a role in partitioning of the amphioxus pharynx into pouches and arches (Figure 7). This is highly conserved with vertebrates. In the lamprey, as in amphioxus, *Tbx1/10* is expressed in the mesodermal core and endoderm in each of the pharyngeal arches, including the mandibular arch, as well as in the upper and lower lips, which develop from the mandibular arch [49]. Similarly, in the shark, *Tbx1* is expressed in a striped pattern in the pharyngeal mesoderm and endoderm, in the wall of the hyoid head cavity and in head mesoderm [7]. Interestingly, even though in the shark, as in all gnathostomes, the somites, which give rise to paraxial muscles, extend anteriorly only as far as the posterior hindbrain, the more anterior head mesoderm of the embryo is clearly segmented and extends sheets of mesoderm ventrally into the pharynx [9,50]. This is reminiscent of the *Tbx1/10*-expressing amphioxus somites extending to the anterior tip of the embryo and giving rise to sheets of *Tbx1/10*-expressing mesoderm that grow into the pharynx [9]. Some of the *Tbx1*-expressing head mesoderm in gnathostomes develops into the extrinsic eye muscles, which is in agreement with the theory that segmentation of the head mesoderm, as well as of the pharynx, is conserved in amphioxus and vertebrates, with the anterior somites of the ancestral chordate evolving into head muscles in vertebrates and not differentiating as paraxial muscle as they do in amphioxus.

Pitx genes are expressed anteriorly to *Tbx1/10* in amphioxus as well as in vertebrates. In amphioxus, *Pitx* is expressed on the left side of the body in somites as well as in mesoderm and endoderm, extending ventrally into the pharynx and becoming localized to endoderm and ectoderm around the mouth, which has been proposed to be a

modified gill slit [29]. Similarly, in the shark, *Pitx2* is expressed in the ectoderm near the mouth as well as in the walls of the hyoid and mandibular head cavities [7], while in the lamprey *PitxA* is expressed in the pre-mandibular mesoderm (head cavity), in ectoderm around the mouth, pharyngeal endoderm and in the ventral portion of the anteriormost somite [51]. In the chick, *Pitx2* is expressed in the ventral ectoderm of the head and in the first pharyngeal pouch as well as in precursors of the extra-ocular and mandibular arch muscles [39].

Tbx1/10 and *Pitx* genes appear to synergize in patterning the head in both amphioxus and vertebrates. Our results show that, in amphioxus, knockdown of *Tbx1/10* does not affect the expression pattern of *Pitx*, but may upregulate its expression. In mouse null mutants for *Tbx1, Pitx2* is downregulated [52]. In addition, mouse Tbx1 can physically interact with the C-terminus of Pitx2 and repress the ability of Pitx2 to activate promoters of several genes, including *Pitx2c* [53]. Conversely, knockdown of *Pitx2* in the zebrafish results in smaller and abnormal cartilages of the mandibular and hyoid arches and misshapen eyes [54], and deletion of *Pitx1* in mice results in downregulation of *Tbx1* [55]. In gnathostomes, Pitx2 is further known to specify head muscles and muscles derived from the first branchial arch and to regulate *Tbx1* expression [56]. Although the regulation of *Tbx1/10* expression by Pitx has not been assessed in amphioxus, it is tempting to speculate that functional interactions of Tbx1/10 and Pitx are required for mediating pharyngeal patterning in both amphioxus and vertebrates.

The vertebrate head mesoderm evolved from a segmented ancestor

The first sign of pharyngeal segmentation in amphioxus is the downregulation of *Pax1/9* in the pharyngeal endoderm and the simultaneous striped expression of *Tbx1/10* in the sheets of mesoderm that migrate in between the pharyngeal endoderm and ectoderm [9]. Thus, it is likely that the anterior somites and their ventral extensions, which constitute the head mesoderm in amphioxus, are pivotal in instructing pharyngeal segmentation. In gnathostomes, it has been proposed that pharyngeal segmentation is driven by expression of genes, such as *Nkx2.5*, *Wnts*, *FGFs* and *BMPs*, expressed in the head mesoderm prior to segmentation of the pharyngeal pouches [39,57]. These genes are also expressed in the amphioxus somites, and it will be interesting to see, if in addition to the permissive role of low RA signaling in the pharynx, segmental expression of these genes in the somites is also required for pharyngeal segmentation in amphioxus. If so, it would lend additional support to the hypothesis that the anterior somites of an ancestral chordate gave rise to the head mesoderm of vertebrates and that the highly disputed somitomeres seen in gnathostome embryos may be the evolutionary equivalent of these amphioxus somites [4]. This idea has been highly contested. One school of thought is that, although the head cavities in the lamprey evolved from the anterior somites of an amphioxus-like ancestor, those of the shark, which are formed by schizocoely, are the result of a gnathostome innovation [7]. To explain similar patterns of gene expression in head structures considered morphologically non-homologous [58], the concept of heterotopy has been invoked, that is, the idea that homologous structures can alter their position during evolution and, therefore, that structures in comparable locations in different organisms may not be homologous [59]. Thus, expression of homologous genes in similar places may not indicate homologous tissues or organs [58]. The alternative view is that not only does conserved expression of genes, including *Pax1/9*, *Six1/2* and *Eya*, in the pharyngeal endoderm of hemichordates, amphioxus, tunicates, lampreys and gnathostomes indicate that a segmented pharynx was present in the common ancestor of hemichordates and chordates, but that the anterior, enterocoelic somites of amphioxus, the head cavities of the lamprey and shark and the jaw and eye muscles of bony gnathostomes are homologous [9].

Our present results, revealing striking similarities in the regulation of pharyngeal patterning between amphioxus and vertebrates, add to the body of evidence that heterotopy probably does not explain similar gene expression in head mesoderm of amphioxus and gnathostomes (Figure 9). The conserved expression of *Tbx1* in head mesoderm of gnathostomes and somites of amphioxus and the similar effects of its knockdown in both groups argue for the common evolutionary ancestry of gnathostome head mesoderm and amphioxus somites. Comparisons of the roles of additional genes segmentally expressed in both the anterior somites and mesoderm migrating into the pharyngeal arches in amphioxus with those expressed in both the early head mesoderm and developing pharynx of gnathostomes could lend additional support to the hypothesis that both the pharynx and the anterior somites of an amphioxus-like ancestral chordate evolved into the pharynx and head mesoderm of vertebrates.

Conclusions

In this manuscript, we have used a combination of pharmacological treatments and morpholino-induced gene knockdown to study RA signaling functions during pharyngeal development of the cephalochordate amphioxus. The results allowed us to define distinct phases of RA activity in the amphioxus pharynx, mediating, for example, the anterior-posterior regionalization of the endoderm as well as the patterning and formation of pharyngeal structures. We were further able to show that Tbx1/10 is required for amphioxus gill slit development and that a

reduction of Tbx1/10 activity in the pharyngeal arches has a similar effect as late RA treatments. These data suggest that segmentation of the amphioxus pharynx requires mutual inhibition of RA signaling and Tbx1/10 function. Given that similar molecular mechanisms control the patterning and segmentation of the vertebrate head, the genetic mechanisms involving RA and Tbx1/10 for partitioning the pharynx into pharyngeal pouches were probably already present in the ancestral chordate and hence precede the evolutionary elaboration of neural crest. Finally, comparisons of our results from amphioxus with data from other deuterostomes indicate that at least some of the molecular components controlling pharyngeal patterning are conserved in hemichordates and chordates, which strongly suggests that the genetic mechanisms for pharyngeal segmentation predate the origin of chordates.

Additional files

Additional file 1: Figure S1. An antisense morpholino oligonucleotide (MO) targeting the amphioxus *Tbx1/10* sequence suppresses translation of the amphioxus *Tbx1/10* gene *in vitro*. Each lane contains 200 ng of amphioxus *Tbx1/10* expression plasmid. While 500 ng and 2,000 ng of the amphioxus *Tbx1/10* MO efficiently block the translation of *Tbx1/10* mRNA, the equivalent amounts of control MO do not affect the *in vitro* translation of *Tbx1/10* mRNA. The arrows indicate the Tbx1/10 protein band.

Additional file 2: Table S1. Stage-dependent effects of retinoic acid (RA) and RAR antagonist (BMS009) on the formation of pharyngeal structures in amphioxus. The proportion of larvae characterized by normal pharynx length and by the presence of mouth, endostyle, club-shaped gland, gill slit primordia and gill slits at 36 hours post fertilization (hpf) is indicated. Larvae were treated with 10^{-6} M RA or 2×10^{-6} M BMS009 from 16 hpf, 20 hpf or 24 hpf.

Abbreviations

CNS: central nervous system; CYP26: cytochrome p450 family 26; DIC: differential interference contrast; DMSO: dimethylsulfoxide; hpf: hours post fertilization; MO: morpholino oligonucleotide; RA: retinoic acid; RALDH: retinaldehyde dehydrogenase; RAR: retinoic acid receptor; RXR: retinoid X receptor.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DK performed most of the experiments and co-wrote the manuscript. JC, MT and JEC contributed preliminary results and supported data analyses. SA and ARdL synthesized and contributed the RAR antagonist. LZH and MS conceived the study and co-wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank John M Lawrence and Susan Bell at the University of South Florida in Tampa, USA, for providing laboratory space during the amphioxus spawning season. We are indebted to Janssen Research & Development, a division of Janssen Pharmaceutica NV, for providing the Cyp26 inhibitor and to Elisabeth Zieger for help with artwork. This work was supported by research grants from the *Agence Nationale de la Recherche* to MS (ANR-09-BLAN-0262-02 and ANR-11-JSV2-002-01). LZH and DK received support from NSF grant IOS 0743485. JEC is funded by a FCT doctoral fellowship (SFRH/BD/86878/2012).

Author details

[1]Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093-0202, USA. [2]Institut de Génomique Fonctionnelle de Lyon (CNRS UMR 5242, UCBL, ENS, INRA 1288), Ecole Normale Supérieure de Lyon, 69364 Lyon, Cedex 07, France. [3]Sorbonne Universités, UPMC Université Paris 06, CNRS, UMR 7009, Laboratoire de Biologie du Développement de Villefranche-sur-Mer, Observatoire Océanologique de Villefranche-sur-Mer, 06230 Villefranche-sur-Mer, France. [4]CNRS, UMR 7009, Laboratoire de Biologie du Développement de Villefranche-sur-Mer, Observatoire Océanologique de Villefranche-sur-Mer, 06230 Villefranche-sur-Mer, France. [5]Departamento de Química Orgánica, Universidade de Vigo, 33610 Vigo, Spain.

Received: 20 June 2014 Accepted: 27 August 2014
Published: 9 October 2014

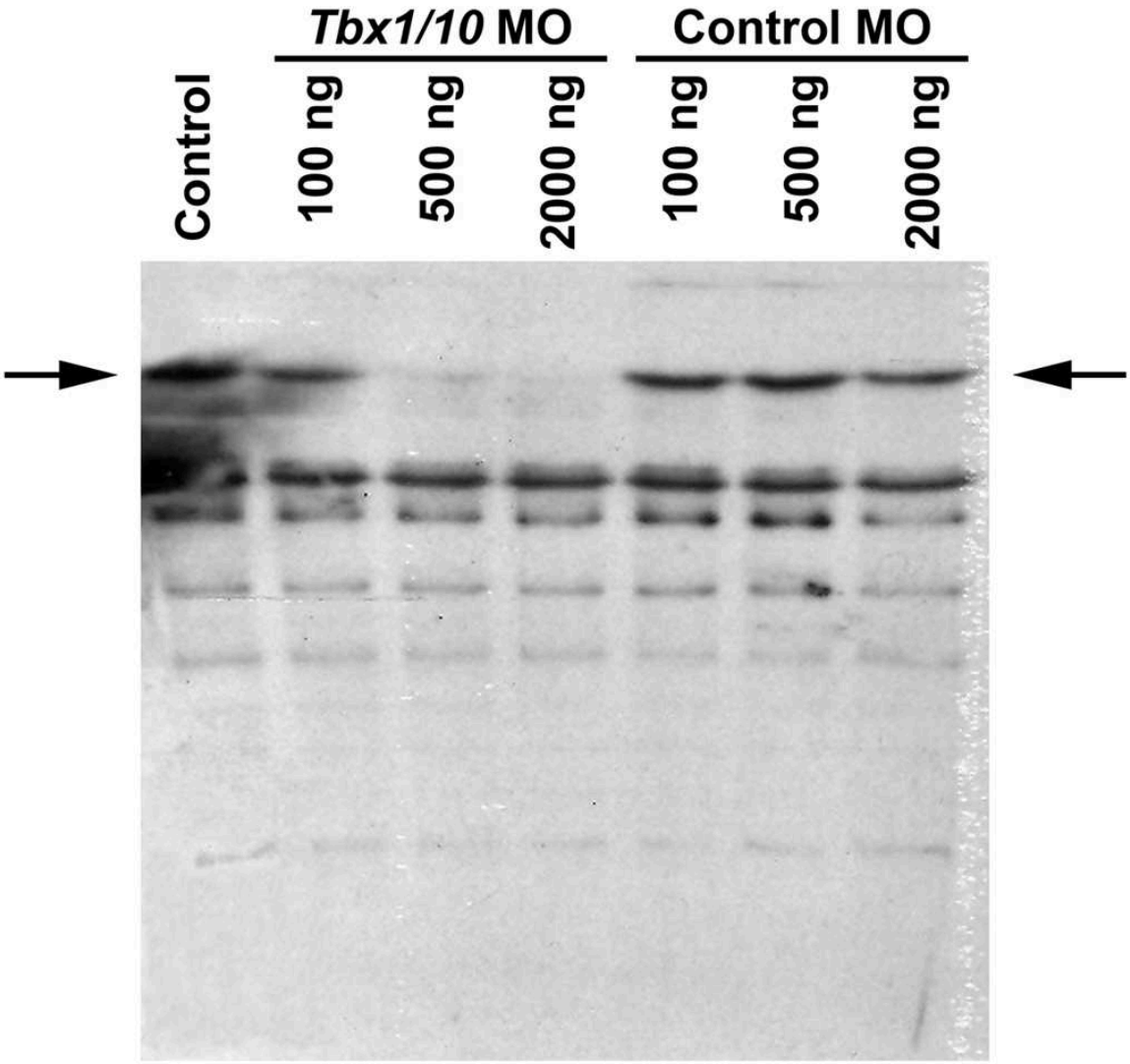
References

1. Chen JY, Huang DY, Li CW: **An early Cambrian craniate-like chordate.** *Nature* 1999, **402:**518–522.
2. Shu DG, Luo HL, Conway Morris S, Zhang XL, Hu SX, Chen L, Han J, Zhu M, Li Y, Chen LZ: **Lower Cambrian vertebrates from south China.** *Nature* 1999, **402:**42–46.
3. Holland ND, Chen JY: **Origin and early evolution of the vertebrates: new insights from advances in molecular biology, anatomy, and palaeontology.** *Bioessays* 2001, **23:**142–151.
4. Conway Morris S, Caron JB: **A primitive fish from the Cambrian of North America.** *Nature* 2014, **512:**419–422.
5. Holland LZ, Kene M, Williams NA, Holland ND: **Sequence and embryonic expression of the amphioxus *engrailed* gene (*AmphiEn*): the metameric pattern of transcription resembles that of its segment-polarity homolog in *Drosophila*.** *Development* 1997, **124:**1723–1732.
6. Matsuura M, Nishihara H, Onimaru K, Kokubo N, Kuraku S, Kusakabe R, Okada N, Kuratani S, Tanaka M: **Identification of four *Engrailed* genes in the Japanese lamprey, *Lethenteron japonicum*.** *Dev Dyn* 2008, **237:**1581–1589.
7. Adachi N, Takechi M, Hirai T, Kuratani S: **Development of the head and trunk mesoderm in the dogfish, *Scyliorhinus torazame: II*. Comparison of gene expression between the head mesoderm and somites with reference to the origin of the vertebrate head.** *Evol Dev* 2012, **14:**257–276.
8. Hatta K, Schilling TF, BreMiller RA, Kimmel CB: **Specification of jaw muscle identity in zebrafish: correlation with engrailed-homeoprotein expression.** *Science* 1990, **250:**802–805.
9. Holland LZ, Holland ND, Gilland E: **Amphioxus and the evolution of head segmentation.** *Integr Comp Biol* 2008, **48:**630–646.
10. Yutzey KE: **DiGeorge syndrome, Tbx1, and retinoic acid signaling come full circle.** *Circ Res* 2010, **106:**630–632.
11. Holland ND, Holland LZ, Kozmik Z: **An amphioxus *Pax* gene, *AmphiPax-1*, expressed in embryonic endoderm, but not in mesoderm: implications for the evolution of class I paired box genes.** *Mol Mar Biol Biotechnol* 1995, **4:**206–214.
12. Kozmik Z, Holland ND, Kalousova A, Paces J, Schubert M, Holland LZ: **Characterization of an amphioxus paired box gene, *AmphiPax2/5/8*: developmental expression patterns in optic support cells, nephridium, thyroid-like structures and pharyngeal gill slits, but not in the midbrain-hindbrain boundary region.** *Development* 1999, **126:**1295–1304.
13. Kozmik Z, Holland ND, Kreslova J, Oliveri D, Schubert M, Jonasova K, Holland LZ, Pestarino M, Benes V, Candiani S: ***Pax-Six-Eya-Dach* network during amphioxus development: conservation *in vitro* but context specificity *in vivo*.** *Dev Biol* 2007, **306:**143–159.
14. Holland LZ, Holland ND: **Expression of *AmphiHox-1* and *AmphiPax-1* in amphioxus embryos treated with retinoic acid: insights into evolution and patterning of the chordate nerve cord and pharynx.** *Development* 1996, **122:**1829–1838.
15. Escriva H, Holland ND, Gronemeyer H, Laudet V, Holland LZ: **The retinoic acid signaling pathway regulates anterior/posterior patterning in the nerve cord and pharynx of amphioxus, a chordate lacking neural crest.** *Development* 2002, **129:**2905–2916.
16. Schubert M, Yu JK, Holland ND, Escriva H, Laudet V, Holland LZ: **Retinoic acid signaling acts via Hox1 to establish the posterior limit of the pharynx in the chordate amphioxus.** *Development* 2005, **132:**61–73.
17. Koop D, Holland ND, Sémon M, Alvarez S, de Lera AR, Laudet V, Holland LZ, Schubert M: **Retinoic acid signaling targets *Hox* genes during the**

amphioxus gastrula stage: insights into early anterior-posterior patterning of the chordate body plan. *Dev Biol* 2010, **338:**98–106.
18. Carvalho JE: *The spatiotemporal activity of retinoic acid signaling in the amphioxus embryo: developmental functions and evolutionary implications, Master thesis.* Portugal: University of Lisbon; 2012.
19. Dupé V, Ghyselinck NB, Wendling O, Chambon P, Mark M: **Key roles of retinoic acid receptors alpha and beta in the patterning of the caudal hindbrain, pharyngeal arches and otocyst in the mouse.** *Development* 1999, **126:**5051–5059.
20. Veitch E, Begbie J, Schilling TF, Smith MM, Graham A: **Pharyngeal arch patterning in the absence of neural crest.** *Curr Biol* 1999, **9:**1481–1484.
21. Quinlan R, Gale E, Maden M, Graham A: **Deficits in the posterior pharyngeal endoderm in the absence of retinoids.** *Dev Dyn* 2002, **225:**54–60.
22. Matt N, Ghyselinck NB, Wendling O, Chambon P, Mark M: **Retinoic acid-induced developmental defects are mediated by RARβ/RXR heterodimers in the pharyngeal endoderm.** *Development* 2003, **130:**2083–2093.
23. Kuratani S, Ueki T, Hirano S, Aizawa S: **Rostral truncation of a cyclostome, *Lampetra japonica,* induced by all-*trans* retinoic acid defines the head/trunk interface of the vertebrate body.** *Dev Dyn* 1998, **211:**35–51.
24. Jandzik D, Hawkins MB, Cattell MV, Cerny R, Square TA, Medeiros DM: **Roles for FGF in lamprey pharyngeal pouch formation and skeletogenesis highlight ancestral functions in the vertebrate head.** *Development* 2014, **141:**629–638.
25. Vandersea MW, McCarthy RA, Fleming P, Smith D: **Exogenous retinoic acid during gastrulation induces cartilaginous and other craniofacial defects in *Fundulus heteroclitus*.** *Biol Bull* 1998, **194:**281–296.
26. Graham A, Okabe M, Quinlan R: **The role of the endoderm in the development and evolution of the pharyngeal arches.** *J Anat* 2005, **207:**479–487.
27. Bayha E, Jørgensen MC, Serup P, Grapin-Botton A: **Retinoic acid signaling organizes endodermal organ specification along the entire antero-posterior axis.** *PLoS One* 2009, **4:**e5845.
28. Holland ND, Paris M, Koop D: **The club-shaped gland of amphioxus: export of secretion to pharynx in pre-metamorphic larvae and apoptosis during metamorphosis.** *Acta Zool (Stockholm)* 2009, **90:**372–379.
29. Yasui K, Kaji T: **The lancelet and ammocoete mouths.** *Zoolog Sci* 2008, **25:**1012–1019.
30. Schubert M, Holland ND, Laudet V, Holland LZ: **A retinoic acid-*Hox* hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus.** *Dev Biol* 2006, **296:**190–202.
31. Holland LZ, Yu JK: **Cephalochordate (amphioxus) embryos: procurement, culture, and basic methods.** *Methods Cell Biol* 2004, **74:**195–215.
32. Holland LZ, Holland PWH, Holland ND: **Revealing homologies between body parts of distantly related animals by *in situ* hybridization to developmental genes: amphioxus versus vertebrates**. In *Molecular Approaches to Zoology and Evolution.* Edited by Palumbi S, Ferraris JD. New York, USA: John Wiley; 1996:267–282. 473–483.
33. Mahadevan NR, Horton AC, Gibson-Brown JJ: **Developmental expression of the amphioxus *Tbx1/10* gene illuminates the evolution of vertebrate branchial arches and sclerotome.** *Dev Genes Evol* 2004, **214:**559–566.
34. Boorman CJ, Shimeld SM: **Pitx homeobox genes in *Ciona* and amphioxus show left-right asymmetry is a conserved chordate character and define the ascidian adenohypophysis.** *Evol Dev* 2002, **4:**354–365.
35. Theodosiou M, Laudet V, Schubert M: **From carrot to clinic: an overview of the retinoic acid signaling pathway.** *Cell Mol Life Sci* 2010, **67:**1423–1445.
36. Albalat R, Brunet F, Laudet V, Schubert M: **Evolution of retinoid and steroid signaling: vertebrate diversification from an amphioxus perspective.** *Genome Biol Evol* 2011, **3:**985–1005.
37. Guidato S, Barrett C, Guthrie S: **Patterning of motor neurons by retinoic acid in the chick embryo hindbrain *in vitro*.** *Mol Cell Neurosci* 2003, **23:**81–95.
38. Sobreira TJ, Marlétaz F, Simões-Costa M, Schechtman D, Pereira AC, Brunet F, Sweeney S, Pani A, Aronowicz J, Lowe CJ, Davidson B, Laudet V, Bronner M, de Oliveira PS, Schubert M, Xavier-Neto J: **Structural shifts of aldehyde dehydrogenase enzymes were instrumental for the early evolution of retinoid-dependent axial patterning in metazoans.** *Proc Natl Acad Sci USA* 2011, **108:**226–231.
39. Bothe I, Tenin G, Oseni A, Dietrich S: **Dynamic control of head mesoderm patterning.** *Development* 2011, **138:**2807–2821.
40. Clausen S, Smith AB: **Palaeoanatomy and biological affinities of a Cambrian deuterostome (Stylophora).** *Nature* 2005, **438:**351–354.
41. Zou D, Silvius D, Davenport J, Grifone R, Maire P, Xu PX: **Patterning of the third pharyngeal pouch into thymus/parathyroid by *Six* and *Eya1*.** *Dev Biol* 2006, **293:**499–512.
42. Zajac JD, Danks JA: **The development of the parathyroid gland: from fish to human.** *Curr Opin Nephrol Hypertens* 2008, **17:**353–356.
43. Gillis JA, Fritzenwanker JH, Lowe CJ: **A stem-deuterostome origin of the vertebrate pharyngeal transcriptional network.** *Proc R Soc B* 2012, **279:**237–246.
44. Ogasawara M, Wada H, Peters H, Satoh N: **Developmental expression of *Pax1/9* genes in urochordate and hemichordate gills: insight into function and evolution of the pharyngeal epithelium.** *Development* 1999, **126:**2539–2550.
45. Campo-Paysaa F, Marlétaz F, Laudet V, Schubert M: **Retinoic acid signaling in development: tissue-specific functions and evolutionary origins.** *Genesis* 2008, **46:**640–656.
46. Hinman VF, Degnan BM: **Retinoic acid disrupts anterior ectodermal and endodermal development in ascidian larvae and postlarvae.** *Dev Genes Evol* 1998, **208:**336–345.
47. Sasakura Y, Kanda M, Ikeda T, Horie T, Kawai N, Ogura Y, Yoshida R, Hozumi A, Satoh N, Fujiwara S: **Retinoic acid-driven *Hox1* is required in the epidermis for forming the otic/atrial placodes during ascidian metamorphosis.** *Development* 2012, **139:**2156–2160.
48. Roberts C, Ivins S, Cook AC, Baldini A, Scambler PJ: ***Cyp26* genes *a1*, *b1* and *c1* are down-regulated in *Tbx1* null mice and inhibition of Cyp26 enzyme function produces a phenocopy of DiGeorge Syndrome in the chick.** *Hum Mol Genet* 2006, **15:**3394–3410.
49. Tiecke E, Matsuura M, Kokubo N, Kuraku S, Kusakabe R, Kuratani S, Tanaka M: **Identification and developmental expression of two *Tbx1/10*-related genes in the agnathan *Lethenteron japonicum*.** *Dev Genes Evol* 2007, **217:**691–697.
50. Gilland E, Baker R: **Conservation of neuroepithelial and mesodermal segments in the embryonic vertebrate head.** *Acta Anat* 1993, **148:**110–123.
51. Boorman CJ, Shimeld SM: **Cloning and expression of a Pitx homeobox gene from the lamprey, a jawless vertebrate.** *Dev Genes Evol* 2002, **212:**349–353.
52. Nowotschin S, Liao J, Gage PJ, Epstein JA, Campione M, Morrow BE: **Tbx1 affects asymmetric cardiac morphogenesis by regulating *Pitx2* in the secondary heart field.** *Development* 2006, **133:**1565–1573.
53. Cao H, Florez S, Amen M, Huynh T, Skobe Z, Baldini A, Amendt BA: **Tbx1 regulates progenitor cell proliferation in the dental epithelium by modulating Pitx2 activation of p21.** *Dev Biol* 2010, **347:**289–300.
54. Liu Y, Semina EV: ***pitx2* deficiency results in abnormal ocular and craniofacial development in zebrafish.** *PLoS One* 2012, **7:**e30896.
55. Mitsiadis TA, Drouin J: **Deletion of the Pitx1 genomic locus affects mandibular tooth morphogenesis and expression of the *Barx1* and *Tbx1* genes.** *Dev Biol* 2008, **313:**887–896.
56. Shih HP, Gross MK, Kioussi C: **Muscle development: forming the head and trunk muscles.** *Acta Histochem* 2008, **110:**97–108.
57. Choe CP, Collazo A, Trinh LA, Pan L, Moens CB, Crump JG: **Wnt-dependent epithelial transitions drive pharyngeal pouch formation.** *Dev Cell* 2013, **24:**296–309.
58. Kuratani S, Murakami Y, Nobusada Y, Kusakabe R, Hirano S: **Developmental fate of the mandibular mesoderm in the lamprey, *Lethenteron japonicum*: comparative morphology and development of the gnathostome jaw with special reference to the nature of the trabecula cranii.** *J Exp Zool B* 2004, **302:**458–468.
59. Haeckel E: **The gastraea-theory, the phylogenetic classification of the animal kindgom and the homology of the germ-lamellae.** *Q J Microsc Sci* 1874, **14:**142–165. 223–247.

doi:10.1186/2041-9139-5-36
Cite this article as: Koop *et al.*: Roles of retinoic acid and Tbx1/10 in pharyngeal segmentation: amphioxus and the ancestral chordate condition. *EvoDevo* 2014 **5**:36.

Additional file 1: Figure S1. An antisense morpholino oligonucleotide (MO) targeting the amphioxus *Tbx1/10* sequence suppresses translation of the amphioxus *Tbx1/10* gene *in vitro*. Each lane contains 200 ng of amphioxus *Tbx1/10* expression plasmid. While 500 ng and 2000 ng of the amphioxus *Tbx1/10* MO efficiently block the translation of *Tbx1/10* mRNA, the equivalent amounts of control MO do not affect the *in vitro* translation of *Tbx1/10* mRNA. The arrows indicate the Tbx1/10 protein band.



Additional file 2: Table S1. Stage-dependent effects of retinoic acid (RA) and RAR antagonist (BMS009) on the formation of pharyngeal structures in amphioxus. The proportion of larvae characterized by normal pharynx length and by the presence of mouth, endostyle, club-shaped gland, gill slit primordia and gill slits at 36 hours post fertilization (hpf) is indicated. Larvae were treated with 10^{-6} M RA or $2x10^{-6}$ M BMS009 from 16 hpf, 20 hpf or 24 hpf.

Treatment		Pharynx length	Mouth	Endostyle	Club-shaped gland	Gill slit primordia	Gill slits
Control		16/17	17/17	17/17	17/17	17/17	17/17
RA	16h-36h	0/20	0/20	1/20	0/20	4/20	0/20
	20h-36h	0/19	0/19	18/19	10/19	19/19	0/19
	24h-36h	12/17	15/17	17/17	17/17	17/17	9/17
BMS009	16h-36h	9/20	20/20	17/20	17/20	19/20	10/20
	20h-36h	12/19	18/19	18/19	18/19	19/19	10/19
	24h-36h	19/20	19/20	19/20	19/20	20/20	16/20

Appendix 7

Published article

Expression of fluorescent proteins in *Branchiostoma lanceolatum* by mRNA injection into unfertilized oocytes.

Hirsinger E, Carvalho JE, Chevalier C, Lutfalla G, Nicolas J-F, Peyriéras N, Schubert M.

Journal of Visualized Experiments. 2015;(95).

(15 pages)

Video Article

Expression of Fluorescent Proteins in *Branchiostoma lanceolatum* by mRNA Injection into Unfertilized Oocytes

Estelle Hirsinger[1], João Emanuel Carvalho[2], Christine Chevalier[1,3], Georges Lutfalla[4], Jean-François Nicolas[1], Nadine Peyriéras[5], Michael Schubert[2]

[1]Département de Biologie du Développement et Cellules Souches, Institut Pasteur

[2]Laboratoire de Biologie du Développement de Villefranche-sur-Mer (UMR7009 CNRS/UPMC Univ Paris 06), Sorbonne Universités

[3]Equipe Epigenetic Control of Normal and Pathological Hematopoiesis, Centre de Recherche en Cancérologie de Marseille

[4]Unité de Dynamique des Interactions Membranaires Normales et Pathologiques, CNRS UMR5235/DAA/cc107/Université Montpellier II

[5]Plateforme BioEmergences IBiSA FBI, CNRS-NED, Institut de Neurobiologie Alfred Fessard

Correspondence to: Nadine Peyriéras at nadine.peyrieras@inaf.cnrs-gif.fr

URL: http://www.jove.com/video/52042
DOI: doi:10.3791/52042

Keywords: Developmental Biology, Issue 95, Amphioxus, cephalochordate, gene expression vectors, *in vivo* imaging, microinjection protocol, model organism

Date Published: 1/12/2015

Citation: Hirsinger, E., Carvalho, J.E., Chevalier, C., Lutfalla, G., Nicolas, J.F., Peyriéras, N., Schubert, M. Expression of Fluorescent Proteins in *Branchiostoma lanceolatum* by mRNA Injection into Unfertilized Oocytes. *J. Vis. Exp.* (95), e52042, doi:10.3791/52042 (2015).

Abstract

We report here a robust and efficient protocol for the expression of fluorescent proteins after mRNA injection into unfertilized oocytes of the cephalochordate amphioxus, *Branchiostoma lanceolatum*. We use constructs for membrane and nuclear targeted mCherry and eGFP that have been modified to accommodate amphioxus codon usage and Kozak consensus sequences. We describe the type of injection needles to be used, the immobilization protocol for the unfertilized oocytes, and the overall injection set-up. This technique generates fluorescently labeled embryos, in which the dynamics of cell behaviors during early development can be analyzed using the latest *in vivo* imaging strategies. The development of a microinjection technique in this amphioxus species will allow live imaging analyses of cell behaviors in the embryo as well as gene-specific manipulations, including gene overexpression and knockdown. Altogether, this protocol will further consolidate the basal chordate amphioxus as an animal model for addressing questions related to the mechanisms of embryonic development and, more importantly, to their evolution.

Video Link

The video component of this article can be found at http://www.jove.com/video/52042/

Introduction

During development, a single cell gives rise to an entire organism in a highly complex process that involves both cell divisions and movements. To better understand the biological principles underlying the dynamics of cell behavior, developmental biologists have started to use fluorescence-based *in vivo* imaging techniques. Specific compartments of cells, such as cell membranes, can either be labeled by treatments with fluorescent dyes, an approach hampered by a lack of specificity and of tissue penetration[1], or by the specific introduction into the embryo of exogenous mRNAs encoding fluorescent proteins[2]. Different techniques can be used for the efficient delivery of exogenous compounds, such as mRNAs. These include, but are not limited to, microinjection, electroporation, bombardment with microparticles, lipofection and transduction[3,4]. Although all of these approaches can be used to introduce exogenous compounds into a developing embryo, only microinjection allows the application of predefined and precise quantities into each cell[3]. Microinjection techniques have been described for all major developmental model systems[4] (*e.g.*, fruit flies, nematode worms, zebrafish, frogs, mice) as well as for some alternative models[4], including those used for comparative studies aimed at understanding the evolution of developmental mechanisms (*e.g.*, sea anemones, annelid worms, sea urchins, ascidian tunicates, the cephalochordate amphioxus).

Cephalochordates, which together with tunicates and vertebrates establish the chordate phylum, are particularly well-suited models to study the evolution of chordates and the diversification of vertebrates from an invertebrate ancestor[5-8]. The cephalochordate lineage diverged very early during chordate evolution; and extant cephalochordates, which are subdivided into three genera (*Branchiostoma*, *Asymmetron* and *Epigonichthys*), resemble vertebrates both in terms of overall anatomy and genome architecture[5-8]. Of the about 30 species of cephalochordates that have been described so far, five are available for embryological and developmental studies[6,9]: *Asymmetron lucayanum* (the Bahama lancelet), *Branchiostoma floridae* (the Florida amphioxus), *Branchiostoma lanceolatum* (the European amphioxus), *Branchiostoma belcheri* (the Chinese amphioxus) and *Branchiostoma japonicum* (the Japanese amphioxus). Ripe adults of three of these species (*B. lanceolatum*, *B. belcheri* and *B. japonicum*) can be induced to spawn on-demand during the breeding season[10,11]. In addition, at least for *B. lanceolatum*, efficient spawning can also be induced in artificial sea water[12], thereby making this particular cephalochordate species accessible for laboratories that do not have access to natural seawater. The combination, in *B. lanceolatum*, of a convenient and reliable access to embryos with an efficient

delivery method, such as microinjection, so far the only delivery technique developed in amphioxus (in both *B. floridae* and *B. belcheri*)[13-15], will enable the development of a novel suite of manipulative techniques, including lineage tracing- and dynamic cell behavior-based approaches.

A protocol for the efficient microinjection of mRNAs to express fluorescent proteins in the *B. lanceolatum* embryo was hence developed. Furthermore, to provide a basic toolkit for live imaging of *B. lanceolatum* embryos, vector systems were developed that allow membrane-associated and nuclear expression of fluorescent proteins. For membrane targeting, enhanced green fluorescent protein (eGFP) was fused to the human HRAS CAAX box and nuclear localization of mCherry and eGFP was obtained by fusion to the zebrafish histone 2B (H2B) exon (**Figure 1**, Supplementary File 1). Furthermore, with the goal to optimize protein translation, the Kozak sequences and codons of the constructs have been modified and adapted to usage in *B. lanceolatum*. Taken together, the injection method and expression vectors presented here will serve as a basis for the generation of new experimental approaches for cephalochordates, notably analyses using the latest fluorescence-based *in vivo* imaging techniques.

Protocol

1. Preparation of Instruments and Reagents

1. Transfer Pasteur pipettes
    1. Generate a series of transfer Pasteur pipettes with different tip diameters by pulling 230 mm long Pasteur pipettes above a flame at different speeds. Ensure that the taper is as long as possible for smooth and fine control of aspiration.
    2. With a diamond scribe, scratch the pipette along a line perpendicular to the length of the pipette. With both hands, pull the pipette parallel to its length to generate a blunt cut. Swiftly flame-polish the pipette without sealing the tip.
    3. Vary the diameter of the tip with the stage of the oocytes/embryos to be pipetted: 300-400 µm for unfertilized oocytes (around 150 µm in diameter), 600 µm for fertilized eggs (around 500 µm in diameter), 200-400 µm for hatched neurulae. Use pipettes with a mouth-monitored aspiration tube to transfer oocytes/embryos from one dish to another.

2. Injection needles
    1. Manipulate the capillaries with gloves to ensure RNase-free conditions. Use borosilicate glass with filament capillaries with dimensions of OD 1.20 mm, ID 0.94 mm, length 10 mm.
    2. If capillaries are pulled on the type of heating-filament needle puller described in the Materials List, use the following settings: Heat 600, Pull 50, Velocity 80, Time 60, Pressure 200 or 300. Otherwise, ensure that the shape, which is crucial for successful injections, is as shown in **Figure 2** with the following properties: 4-8 µm outer tip diameter, 2 cm taper length.
    NOTE: Needles can be pulled before the spawning season and used throughout the season.

3. 5% Phenol Red stock solution (4x)
    1. Prepare fresh before each spawning season.
    2. In a 0.22 µm-filtration tube, weigh out 25 mg of Phenol Red powder.
    3. Add 0.5 ml of DNase- and RNase-free water to the tube containing the powder.
    4. Spin for 3-5 min at 18,000 x g at room temperature to filter-sterilize the solution and remove crystals that could clog the injection needle.
    5. Store at 4 °C or store 250 µl aliquots at -20 °C.

4. 0.25 mg/ml poly-lysine solution
    1. Prepare fresh before each spawning season.
    2. Dissolve 5 mg of poly-L-lysine in 20 ml distilled water. Store 5 ml aliquots at -20 °C.
    3. Use a defrosted aliquot immediately and only once to ensure reproducible and robust adhesion of the oocytes to the poly-lysine-coated dish.

5. mRNA synthesis
    1. Prepare fresh before each spawning season.
    2. Linearize 5 µg of DNA of interest with the adequate enzyme (usually for 2 hr, at 37 °C). To check the completeness of the digestion, run 2% in volume of the digestion mix on a 1% agarose-TBE gel in TBE buffer at 150 W for 20 min.
    3. Extract the linearized DNA with 25:24:1 phenol (pH 8.0):chloroform:isoamyl alcohol. Vortex for 20 sec, centrifuge 10 min at 18,000 x g and collect the aqueous (upper) phase.
    4. Extract the aqueous phase again with 24:1 chloroform:isoamyl alcohol. Vortex for 20 sec, centrifuge 10 min at 18,000 x g and collect the aqueous phase.
    5. Precipitate the linearized DNA with 100:10:300 linearized DNA:3 M sodium acetate (pH 5.2):100% ethanol overnight at -20 °C.
    6. Centrifuge 20 min at 18,000 x g at 4 °C. Rinse in 70% ethanol.
    7. Centrifuge 10 min at 18,000 x g at 4 °C. Let dry and resuspend in RNase-free water at a final concentration of 0.5 µg/µl.
    8. Transcribe 1 µg of linearized DNA using an mRNA synthesis kit with the appropriate polymerase, according to manufacturer's instructions.
    9. Extract the mRNA with 5:1 phenol (pH 4.7):chloroform and ammonium acetate stop solution provided with the kit.
    10. Vortex for 20 sec, centrifuge 10 min at 18,000 x g and collect the aqueous phase.
    11. Extract the aqueous phase again with 24:1 chloroform:isoamyl alcohol. Vortex for 20 sec, centrifuge 10 min at 18,000 x g and collect the aqueous phase.
    12. Precipitate the mRNA with 100% isopropanol overnight at -20 °C.
    13. Centrifuge 20 min at 18,000 x g at 4 °C. Rinse in 80% ethanol.
    14. Centrifuge 10 min at 18,000 x g at 4 °C. Let the pellet dry at room temperature for no longer than 5-10 min as it will then be difficult to resuspend. Resuspend in DNase- and RNase-free water to a final concentration of at least 2 µg/µl to ensure a decent final mRNA concentration in the injection mix.

15. To check the quality and size of the transcription product, run 0.5 µl of the mRNA on a RNase-free 1% agarose-TBE gel in TBE buffer at 150 W for 20 min. Store 2 µl aliquots at -80 °C.

6. Poly-lysine-coated dishes
NOTE: Poly-lysine coated dishes are used to immobilize the oocytes during injection.
    1. For each 35 mm cell-culture Petri dish (5 in total), cover the bottom of the Petri dish with 1 ml of the thawed 0.25 mg/ml poly-lysine solution. Incubate at room temperature for 5 min.
    2. For each 35 mm cell-culture Petri dish (5 in total), transfer the 0.25 mg/ml poly-lysine solution into another 35 mm cell-culture Petri dish. Incubate at room temperature for 5 min.
    3. Discard the 0.25 mg/ml poly-lysine solution.
    4. Let the Petri dishes dry, upside-down at room temperature for 2 hr.
    5. Store the poly-lysine-coated dishes wrapped in a plastic wrap at 4 °C to avoid contamination for one week maximum.

7. Agarose-coated dishes
NOTE: They are used to culture injected embryos. The agarose provides a cushion for the injected embryos and prevents them from sticking to the bottom of the dish.
    1. Make artificial seawater (ASW) using 37-38 g/L commercial salts + 0.25 mM $NaHCO_3$ in reverse osmosis water.
    2. Dissolve agarose to a 1% concentration in 0.22 µm-filtered ASW by heating the solution in a microwave.
    3. Swiftly pour the warm agarose solution from one 35 mm Petri dish into another one in order to ensure a very thin agarose coating of the dish.
    4. Store the agarose-coated dishes wrapped in Saran wrap at 4 °C to avoid contamination for one week maximum.

8. Injection mix and loading of the injection needles
    1. About 2 hr before starting the injections, make a 2 µl injection mix in RNase- and DNase-free water with final concentrations of 1-1.8 µg/µl of mRNA, 15% glycerol, 1.25% Phenol Red.
    NOTE: The Phenol Red colors the solution, which allows monitoring of the injection efficiency and the identification of successfully injected embryos. Glycerol favors mRNA diffusion within the oocyte.
    2. Centrifuge 4 min at 18,000 x g to pellet crystals. Keep on ice until use.
    3. With a 10 µl pipette, collect 0.5 µl of the injection mix, avoiding the bottom of the tube, where the crystals have been pelleted.
    4. Backfill at least two injection needles (in case one breaks during the injection) by pipetting the 0.5 µl drop of injection mix at the large opening of the needle.
    5. Install the needles in a storage jar with liquid at the bottom at 4 °C to prevent evaporation of the injection mix. Let the injection mix slowly travel to the tip of the injection needle for at least 1 hr to minimize the creation of bubbles.
    6. Store additional injection mix at -80 °C for maximum of three additional uses, after which the mRNA quality deteriorates (data not shown).

2. Collection of Biological Material, Microinjection and Embryo Culture

1. Oocyte and sperm collection
NOTE: See Theodosiou *et al.*[12] for a detailed protocol for inducing spawning and for gamete collection.
    1. Shock males and females in ASW at 23 °C for 24 hr.
    2. One to two hours before sunset, transfer the adults into individual cups in ASW at 19 °C because most adults will spawn 1-2 hr after sunset.
    3. Rinse 35 mm Petri dishes in filtered ASW and let them dry upside-down to prevent the oocytes from sticking to the bottom of the dish.
    4. Upon spawning, immediately collect sperm and oocytes with a 1,000 µl pipette.
        1. Keep sperm and oocytes separate from adult amphioxus because the contact is detrimental for gamete health (data not shown).
        2. Furthermore, avoid startling the adults, which leads to movements that dissipate, and hence dilute, both sperm and oocytes. To keep the sperm active as long as possible and to optimize fertilization rate, collect the sperm as concentrated as possible.
    5. Keep sperm on ice in a 1.5 ml tube.
    6. Transfer oocytes in filtered ASW into the pre-rinsed 35 mm Petri dishes.
    7. Transfer 100-500 oocytes with a previously pulled 300-400 µm transfer Pasteur pipette to another 35 mm Petri dish to perform the injections.
    8. Fertilize the remainder of the clutch as a control for sperm and oocyte quality or for other experiments.

2. Oocyte injection
    1. Install the injection needle on the micromanipulator at a 50° angle relative to the horizontal plane.
    NOTE: Angles of less than 50° will push the oocytes around on the dish, while angles of more than 50° will not allow an appropriate monitoring of the needle position relative to the oocyte.
    2. Under a fluorescent dissecting scope with 25X oculars, transfer 30 oocytes with the 300-400 µm transfer Pasteur pipette on a poly-lysine-coated dish containing filtered ASW.
    3. Deposit the oocytes along a line to carry out injections in an ordered way and to distinguish injected from non-injected oocytes. Inject small numbers (30 oocytes) to minimize the exposure time of oocytes to poly-lysine, which tends to deform developing embryos (data not shown).
    4. Use the dark field illumination to render the oocytes as translucent as possible.
    5. With the coarse movement knob of the micromanipulator, bring the injection needle close to an oocyte.
    6. With fine forceps, cut open the needle at the level where the tip starts to be curved. By pulsing with the injector, verify that red injection mix is actually flowing out of the needle.

7. At 200X magnification and with the fine movement knob of the micromanipulator, gently move the injection needle inside the core of the oocyte.
   NOTE: If inserted too superficially, the injected solution will not remain inside the oocyte. If inserted too far, the oocyte will be destroyed.
8. Inject with 1-3 pulses of 120 msec duration and 1-10 psi pressure. If the needle is fine enough, inject with continuous flow at constant pressure. Ensure that the injection volume corresponds to 1/5 to 1/3 of the volume of a single oocyte.
9. Following injection, pull the needle out swiftly to avoid leakage of the oocyte.
10. Verify that the injected solution remains within the oocyte and that after a few sec, the injected solution spreads throughout the oocyte.
11. Move on to the next oocyte in line.
12. Keep some uninjected embryos of each series as negative control to estimate the background fluorescence when scanning for injected embryos.

3. Fertilization, selection of injected embryos and embryo culture
   1. Fertilize the oocytes as soon as a series has been injected. As oocyte quality declines with time, inject and fertilize oocytes within 1 hr after spawning[12].
   2. Depending on the sperm concentration, add 1-5 drops of sperm to the oocytes and swirl the dish.
      NOTE: The fertilization envelope should become apparent on the embryos after about 1 min.
   3. Allow the embryos to detach from the poly-lysine-coated dish, while injecting another series of oocytes.
   4. Transfer the embryos with the 600 µm transfer Pasteur pipette into an agarose-coated Petri dish. Remove the embryos from the poly-lysine-coated dish as soon as possible, if at all possible before the 2-cell stage.
      NOTE: In case of prolonged exposure to poly-lysine, the embryos tend to become densely-packed blastulae, flattened on the side touching the bottom of the dish.
   5. At the 2-cell to 4-cell stage, select with a fluorescent dissecting scope with DSR filter the successfully-injected embryos, *i.e.,* those with a normal morphology that exhibit a Phenol Red-derived red fluorescent signal.
   6. Keep the embryos in culture in filtered ASW in agarose-coated Petri dishes at 19 °C until the desired stage for *in vivo* imaging.

Representative Results

The protocol detailed above provides the basis for the microinjection of *B. lanceolatum* oocytes and hence for the introduction into developing *B. lanceolatum* embryos of mRNA encoding fluorescent proteins for *in vivo* imaging. Although the technique is certainly robust and reliable, the rate of successful injections using this protocol remains variable (**Table 1**). The very likely explanation for this intriguing fact is the extreme variability of oocyte clutches: different egg batches do indeed behave very differently, when subjected to the injection pressure. Some oocytes are rather elastic and tend to flatten at the bottom of the dish, while others are quite stiff and remain round upon injection. Furthermore, some oocyte batches tend to inflate their chorion membranes upon injection, while others simply lyse (data not shown). No obvious correlation could be established between oocyte behavior upon injection and embryonic development after fertilization. It is thus difficult to anticipate which category of oocytes is best suited for microinjection. However after fertilization, embryos undergoing normal development can be identified as early as cleavage stages: 2-cell to 4-cell stage embryos can be considered normal when the contact surface between blastomeres is small, while a compacted cleavage-stage embryo, where individual cells are difficult to discern, is definitely abnormal.

In our hands, about half of the oocytes do not survive the trauma caused by the injection and about half of the embryos that do survive the injection exhibit a specific fluorescent label. Thus, we estimate that with this injection protocol, between 50 and 120 embryos can be successfully injected on a given spawning day yielding between 15 and 60 labeled embryos.

The red fluorescence of the Phenol Red in 2-cell to 4-cell stage embryos very reliably marks embryos that have been properly injected, as 100% of embryos selected in this manner subsequently produce the fluorescent proteins encoded by the injected mRNA. Additionally, there is a very clear positive correlation between the intensity of Phenol Red fluorescence at the 2-cell to 4-cell stage and the time of onset of the fluorescence produced by the proteins encoded by the injected mRNA. This correlation thus allows the identification of those embryos that have received the highest quantity of mRNA during the injection process.

If the signal of the fluorescent protein encoded by the injected mRNA is not detected after selection of Phenol Red-positive embryos, we recommend to run the injection mix on an RNAse-free gel in order to check for mRNA degradation or trapping (**Figure 3**). In lane 1, an intact mRNA band is detected. Injection of this mix led to strong fluorescent signal in the embryo. On the contrary, in lane 2 which corresponds to another experiment where Phenol Red was replaced with Dextran dye in the mix (see discussion for further details), the mRNA seems to have been trapped by the Dextran dye and injection of this mix never led to fluorescent signal in the embryo.

Using this microinjection technique and our constructs (**Tables 2 and 3**, Supplementary File 1) (see discussion), we can thus reproducibly produce homogenous fluorescent labeling throughout the embryo with mCherry or eGFP in the nucleus and with eGFP at the membrane (**Figure 4**). The imaging protocol used to generate **Figure 4** will be described elsewhere (Faure *et al.,* currently in preparation). Depending on the amount of mRNA injected, fluorescent protein expression typically becomes detectable between 16-cell (**Figure 4A**) and 64-cell stages and stays detectable at least up to the late neurula stage (data not shown). Later stages have not been tested. Nuclear signal typically appears earlier than membrane signal, potentially because of the intrinsically diffuse nature of the membrane compared to the compact nature of the nucleus (data not shown). Although it has not been monitored systematically, embryos injected with both a nuclear and a membrane label appear to be less healthy than embryos injected exclusively with a nuclear eGFP label. Intriguingly, this effect seems to be independent of the total amount of mRNA injected as the total amount of injected mRNA is the same, whether one or two mRNA species are included to the mix (data not shown).

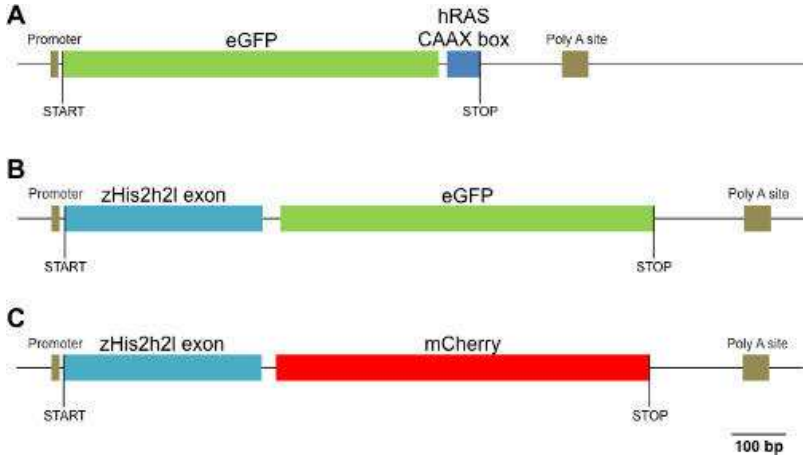


Figure 1: Construct details. Sketches of **(A)** the membrane-targeted eGFP (eGFP:CAAX box), **(B)** the nuclear-targeted eGFP (H2B:eGFP) and **(C)** the nuclear-targeted mCherry (H2B:mCherry) constructs are shown.

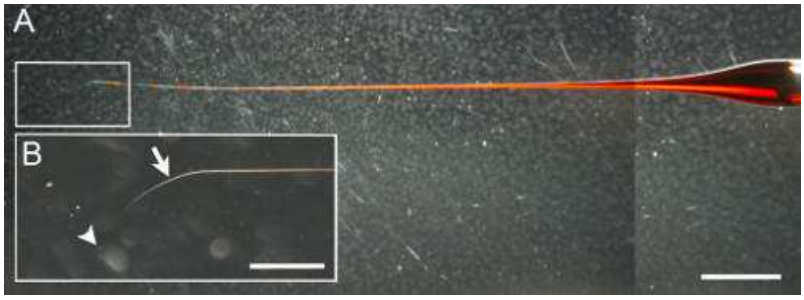


Figure 2: Shape of a representative injection needle. (A) The taper is about 2 cm in length and the needle curves at the tip. Scale bar = 2 mm. **(B)** Magnification of the boxed area in A. The position for clipping the needle (arrow) is in the curve of needle. The tip of the needle is indicated by the arrowhead. Scale bar = 500 µm. Please click here to view a larger version of this figure.

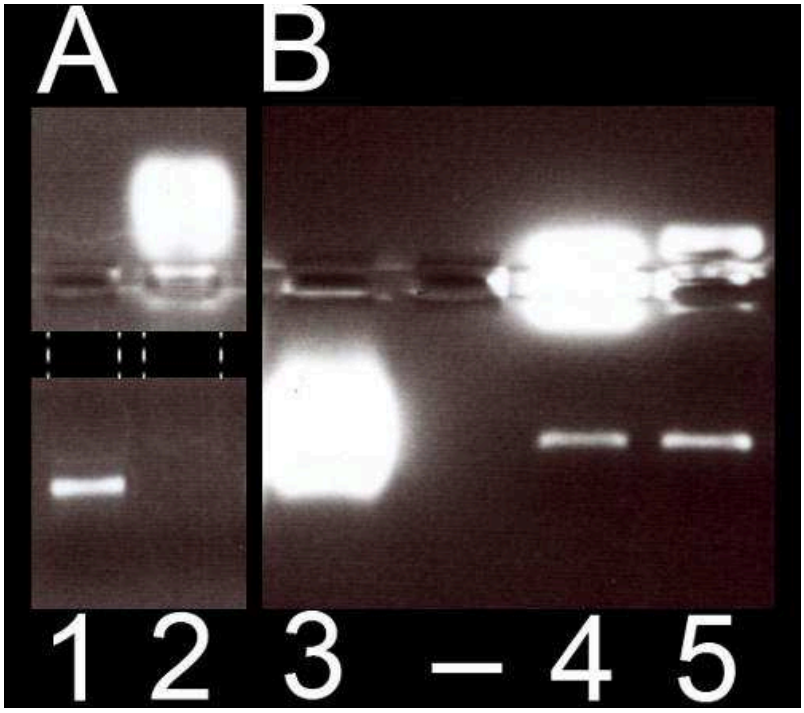


Figure 3: Interactions of mRNA and selected fluorescent dyes. **(A)** Result of the migration of a mix containing mRNA and Phenol Red in lane 1 and of mRNA and Texas Red dextran (at a final concentration of 3.3 mg/ml) in lane 2. **(B)** Result of the migration of a mix containing mRNA and Oregon Green dextran (at a final concentration of 3.3 mg/ml) in lane 3, of mRNA and FITC dextran (at a final concentration of 3.3 mg/ml) in lane 4 and of mRNA and Rhodamine dextran (at a final concentration of 3.3 mg/ml) in lane 5. Gel migration time in A and B was different. For the sake of clarity, the black rectangle in A masks a light reflection and the dashed lines delineate the migration lanes. Please click here to view a larger version of this figure.

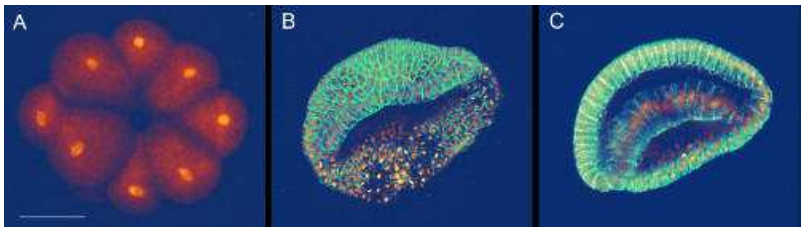


Figure 4: 3D rendering (Amira software) of amphioxus embryos injected with mRNAs coding for fluorescent proteins. Images are extracted from 3D time-lapses acquired on an optimized Leica SP5 2-photon laser-scanning microscope. **(A)** 16-cell stage embryo expressing nuclear H2B:eGFP. **(B)** Gastrula stage embryo expressing nuclear H2B:mCherry and the membrane-targeted eGFP:CAAX box. **(C)** Gastrula stage embryo expressing nuclear H2B:mCherry and the membrane-targeted eGFP:CAAX box. The optical section shows internal cells with nuclear mCherry and membrane-associated eGFP signals. Scale bar = 50 µm.

Experiment number	Injected construct mRNA	Number of injected embryos	Number of labeled embryos
1	H2B:eGFP	53	23
2	H2B:eGFP	50	31
3	H2B:eGFP	48	20
4	H2B:eGFP	54	24
5	H2B:eGFP	114	47
6	H2B:eGFP	80	46
7	H2B:eGFP or	62	24
	H2B:mCherry + eGFP:CAAX box		
8	H2B:eGFP	87	60
9	H2B:eGFP	77	45
10	eGFP:CAAX box or	110	51
	H2B:mCherry + eGFP:CAAX box		
11	H2B:mCherry + eGFP:CAAX box	45	26
12	H2B:mCherry + eGFP:CAAX box	52	16
13	eGFP:CAAX box or	74	35
	H2B:mCherry + eGFP:CAAX box		
	Total	**906**	**448**
		Success rate	**49%**

Table 1: Injection success rates. The injected constructs, the number of injected embryos that survived the injection and the number of labeled (or successfully injected) embryos are indicated, as is the overall percentage of successful injections.

Construct	Position in the pCS2+ vector	Original sequence	Mutated sequence
pCS2+ eGFP:CAAX box	79..87	GGA TCC ACC	**ACC GT**C A**A**C
pCS2+ H2B:eGFP	82..90	TCC GAC ACG	**A**CC GT**C** A**AC**
pCS2+ H2B:mCherry	82..90	TCC GAC ACG	**A**CC GT**C** A**AC**

Table 2: Tentative optimization of Kozak sequences. Original and mutated Kozak sequences are indicated for each construct. The introduced mutations recover the Kozak sequence of the *B. lanceolatum* actin gene, which very closely resembles that of the inferred theoretical preferred Kozak sequence of amphioxus (A/CGA/C G/TTC A/GAC atg TG/CT).

pCS2+ eGFP:CAAX box			
Original codon	**Mutated codon**	**Position in the pCS2+ vector**	
(usage level)	**(usage level)**		
GTA	GT**G**	154..156	eGFP sequence
(0.08)	(0.43)		
AGA	AG**G**	814..816	Linker
(0.09)	(0.27)		
pCS2+ H2B:eGFP			
Original codon	**Mutated codon**	**Position in the pCS2+ vector**	
(usage level)	**(usage level)**		
GTA	GT**G**	223..225	H2B sequence
(0.08)	(0.43)		
		289..291	H2B sequence
		469..471	Linker
		568..570	eGFP sequence
CTA	CT**G**	226..228	H2B sequence
(0.06)	(0.51)		
pCS2+ H2B:mCherry			
Original codon	**Mutated codon**	**Position in the pCS2+ vector**	
(usage level)	**(usage level)**		
GTA	GT**G**	223..225	H2B sequence
(0.08)	(0.43)		
		289..291	H2B sequence
		469..471	Linker
		913..915	mCherry sequence
CTA	CT**G**	226..228	H2B sequence
(0.06)	(0.51)		

Table 3: Tentative optimization of codon usage. Original and mutated codons are indicated for each construct. While this has not been experimentally tested, the introduced mutations are meant to optimize mRNA translation in amphioxus.

Supplementary File 1: Construct sequences. The nucleotide sequences of (A) the membrane-targeted eGFP (eGFP:CAAX box), (B) the nuclear-targeted eGFP (H2B:eGFP) and (C) the nuclear-targeted mCherry (H2B:mCherry) constructs are given in the context of the pCS2+ vector backbone. The main domains encoded by each of the constructs are indicated in color: eGFP (green), mCherry (red), membrane localization signal from human HRAS (CAAX box) (cyan blue), zebrafish histone 2B exon (H2B) (yellow).

Discussion

In this article, we present, for the first time, a detailed and reproducible protocol for the injection of *B. lanceolatum* oocytes, which, after *B. floridae*[13,14] and *B. belcheri*[15], is thus the third amphioxus species, for which such a technique has been described. Importantly, the protocol described here also includes the description of vector systems suited for the production of fluorescent proteins in *B. lanceolatum* from injected mRNA produced *in vitro* (described below). Together, these new tools allow *in vivo* imaging of the early development of amphioxus and the analysis of dynamic cell behaviors that underlie the earliest morphogenetic events in the embryo, such as cleavage and gastrulation.

In general, the microinjection technique has the benefit of allowing the delivery, into a target cell, of a pre-defined volume of a specific compound. With this being said, one major limitation of this technique is the limited number of cells that can be injected during a given experiment. For example, the protocol described here for *B. lanceolatum* allows the injection of 100 to 120 eggs per injection session, of which about half will be labeled (**Table 3**). For *B. floridae*, the numbers are very similar with 500 or more eggs that can be injected per day, of which more than 50% will survive the injection[13,14]. While the protocols for injecting *B. lanceolatum* and *B. floridae* eggs closely resemble each other, the *B. belcheri* technique is slightly different: the eggs are placed on poly-lysine-coated coverslips and injections are performed under an inverted microscope using a microinjector of a different brand. This approach apparently allows the injection of between 200 and 300 *B. belcheri* eggs per injection session with a survival rate, after hatching at the neurula stage, of 89.39% to 95.83%[15]. Considering the differences in the protocols, it is difficult to know whether the differences in success rates are due to species-related or protocol-related differences. One would have to test the different species in parallel with the different protocols to answer this question. Nonetheless, to further increase the number of oocytes targeted for the delivery of exogenous compounds, other protocols will have to be developed for amphioxus. Candidate techniques include electroporation,

bombardment with microparticles, lipofection, and transduction[4]. Of note, electroporation has already been established as a standard technique for the introduction of exogenous material into fertilized ascidian tunicate eggs, another invertebrate chordate model used for studying the evolution of developmental mechanisms[8].

For successful microinjection and subsequent *in vivo* imaging of fluorescent proteins, suitable expression vectors are a mandatory prerequisite. Towards this end, three plasmids have been developed and experimentally validated (**Figure 1**, Supplementary File 1): for membrane targeting, the eGFP gene was fused at its 3' end to the human HRAS CAAX box (resulting in a eGFP:CAAX box construct)[16] and, for nuclear targeting, both the eGFP and the mCherry genes were fused at their 5' ends to the zebrafish histone 2B (H2B) exon (resulting, respectively, in a H2B:eGFP and a H2B:mCherry construct)[17]. Each one of these constructs has been cloned into the pCS2+ plasmid containing a SV40 late polyadenylation site and a SP6 promoter allowing *in vitro* capped mRNA synthesis[18,19]. Furthermore, with the goal to optimize translation of the constructs in *B. lanceolatum*, both the Kozak sequences and the codons have been adapted using single nucleotide mutagenesis (**Tables 2 and 3**, Supplementary File 1). Based on the approach by Nakagawa[20], a small pool of *B. lanceolatum* genes was analyzed to identify a preferred Kozak sequence in this species. The closest known naturally-occurring sequence to this theoretical preferred sequence is that of the *B. lanceolatum actin* gene. The Kozak sequences of the three constructs were hence modified to match this *actin* gene sequence. Furthermore, the codon usage of *B. lanceolatum* was analyzed using the codon usage database[21] and codons in the constructs with a low usage probability in *B. lanceolatum* (<10%) were replaced, wherever possible, by equivalent codons with a higher usage probability.

The results of the injections show that the level of protein expression from these vectors is suitable for *in vivo* imaging analyses. Given that the expression output of the modified vectors has not been compared with that of unmodified constructs, we cannot conclude on the efficiency of the modifications that were introduced in the Kozak consensus and the coding sequences. It would certainly be interesting to test whether these modifications indeed have an impact on protein expression levels. Along these lines, a recent analysis based on microinjections into *B. belcheri* suggests that other, unmodified vectors can also be used for the synthesis of mRNA for fluorescent protein production in amphioxus[15].

Amphioxus contains endogenous green fluorescent proteins[22]. The embryos therefore exhibit low levels of green as well as red fluorescence (our unpublished observations). However, these endogenous levels are negligible with respect to the fluorescence levels resulting from the injection of our mRNA constructs. Therefore, the endogenous fluorescence of amphioxus embryos does not disturb experimental observations using fluorescent binoculars or multi-laser scanning microscopes. In addition, the injection experiments reveal that the intensity of the exogenous fluorescence generated by mRNA injection depends on the actual amount of injected material, which is directly correlated with the final concentration of mRNA used in the injection mix. *B. lanceolatum* embryos tend to tolerate a concentration of up to 1.8 µg/µl mRNA, although, independent of the actual mRNA concentration, some mRNA combinations seem to be less tolerated by *B. lanceolatum* embryos than others. Thus, the combination of nuclear mCherry and membrane eGFP mRNAs appeared to be more toxic than the injection of nuclear eGFP alone.

Texas Red dextran has previously been used as a tracer for successful injections into amphioxus oocytes[13-15]. Intriguingly, a fluorescent signal derived from the injected mRNA was never obtained after injection of solutions containing Texas Red dextran and this in both amphioxus oocytes (n = 4 experiments) and zebrafish embryos (n = 3 experiments). When an injection mix containing *in vitro* transcribed mRNA and Texas Red dextran is loaded on a RNase-free agarose gel (**Figure 3**, lane 2), the band corresponding to the mRNA is absent. In contrast, *in vitro* transcribed mRNA is detectable on an RNase-free agarose gel in the presence of Phenol Red (**Figure 3**, lane 1). Given that there is no evidence for mRNA degradation on the gel, these results suggest that Texas Red dextran tends to trap mRNA. When used in an injection solution, Texas Red dextran might thus prevent the translation of the injected mRNA. Following this observation, the interactions of other dyes (FITC dextran, Rhodamine dextran and Oregon Green dextran) with mRNA were tested, both on RNase-free agarose gels and by injection into *B. lanceolatum* or zebrafish (**Figure 3**). The analyses indicate that FITC dextran does not trap mRNA on gel (**Figure 3**, lane 4) and leads to fluorescent protein expression in zebrafish (n = 1 experiment), but not in embryos of *B. lanceolatum* (n=1 experiment) (data not shown). Rhodamine dextran does not trap mRNA on gel (**Figure 3**, lane 5) and leads to fluorescent protein expression in zebrafish embryos (n = 1 experiment) (data not shown), but its activity could unfortunately not be tested in *B. lanceolatum* embryos. Finally, as the Oregon Green dextran migrates at the same level as the mRNA, mRNA trapping could not be assessed by agarose gel (**Figure 3**, lane 3). This latter dye prevented fluorescent protein expression in amphioxus (n = 1 experiment), but not zebrafish embryos (n=1 experiment) (data not shown). In sum, these data suggest that some dextrans can efficiently inhibit the translation of injected mRNA. Given that dose-response experiments were not carried out, these data are qualitative. Intriguingly, this inhibitory effect seems to be dependent on the animal species (zebrafish versus amphioxus), which highlights the need for further studies to understand the mechanisms underlying this effect. Furthermore, these results call for preparatory analyses, if dextrans are to be used as color tracers for injections.

The development of the microinjection technique for a given model opens the door for gene-specific manipulations. In amphioxus, for example, the first description of microinjections (in *B. floridae*) was followed by the successful knockdown of a gene, *hox1*, using morpholino oligonucleotide injections[23,24]. The number of *B. lanceolatum* embryos that can successfully be injected every day using the protocol presented here is certainly compatible with this type of study. Furthermore, with these methods and the novel constructs at hand, one can now also design and perform overexpression and knockdown studies by injection of mRNAs of interest (native or dominant-negative forms) or use other strategies for disrupting gene expression (for example microRNA- or shRNA-based strategies). So far, amphioxus injections can only be performed at the oocyte stage, simply because this is the only stage that can efficiently be immobilized. After injection at the oocyte stage, the molecule of interest is ubiquitously expressed in the embryo. Thus, if gene expression needs to be altered or monitored only in a subset of cells, DNA constructs need to be injected, as these are expressed mosaically in the developing amphioxus embryo[15,25-27]. DNA construct injections can also be used to test the activity of regulatory regions during amphioxus development in transient transgenic assays, with the aforementioned drawback of mosaic expression[15,25-27].

Ultimately, the amphioxus microinjection technique also opens the door for stable transgenesis, including the targeted knock-out or knock-in of specific genetic loci. Systems for stable insertion of exogenous DNA already exist, for example with the fish transposon-based Tol2 system that seems to function in both insects and amniote vertebrates[28]. Furthermore, approaches based on modified zinc-finger nucleases (ZFNs) or transcription activator-like effector nucleases (TALENs) or RNA-guided genome editing tools (CRISPR/Cas system) could be used to generate point mutations and knock-in modifications in specific regions of the genome[29]. These efforts will require the year-round breeding of amphioxus in captivity, which has already been set up for *B. belcheri*[11,30] and *B. japonicum*[30,31] and is currently being developed for *A. lucayanum*, *B. floridae* and the amphioxus species featured here, *B. lanceolatum*[32].

Disclosures

The authors have nothing to disclose.

Acknowledgements

The authors would like to acknowledge the support by the “Animalerie Centrale de Gif-sur-Yvette” for animal husbandry. This work was supported by funds from ANR (ANR-09-BLAN-0262-02 and ANR-11-JSV2-002-01) to Michael Schubert, by the European Union FP6 grant “Embryomics” and by the ANR grant “ANR-10-BLAN-121801 Dev-Process” to Jean-François Nicolas and Nadine Peyriéras. João Emanuel Carvalho is financed by a FCT doctoral fellowship (SFRH/BD/86878/2012).

Requests for the vectors described here can be addressed directly to the authors.

References

1. Weber, T., Köster, R. Genetic tools for multicolor imaging in zebrafish larvae. *Methods.* **62**, 279-291, doi: 10.1016/j.ymeth.2013.07.028 (2013).
2. Weil, T.T., Parton, R.M., Davis, I. Making the message clear: visualizing mRNA localization. *Trends. Cell. Biol.* **20**, 380-390, doi: 10.1016/j.tcb.2010.03.006 (2010).
3. Zhang, Y., Yu, L.C. Microinjection as a tool of mechanical delivery. *Curr. Opin. Biotechnol.* **19**, 506-510, doi: 10.1016/j.copbio.2008.07.005 (2008).
4. Stepicheva, N.A., Song, J.L. High throughput microinjections of sea urchin zygotes. *J. Vis. Exp.* **83**, e50841, doi: 10.3791/50841 (2014).
5. Schubert, M., Escriva, H., Xavier-Neto, J., Laudet, V. Amphioxus and tunicates as evolutionary model systems. *Trends Ecol. Evol.* **21**, 269-277, doi: 10.1016/j.tree.2006.01.009 (2006).
6. Bertrand, S., Escriva, H. Evolutionary crossroads in developmental biology: amphioxus. *Development.* **138**, 4819-4830, doi: 10.1242/dev.066720 (2011).
7. Holland, L.Z. Evolution of new characters after whole genome duplications: insights from amphioxus. *Semin. Cell Dev. Biol.* **24**, 101-109, doi: 10.1016/j.semcdb.2012.12.007 (2013).
8. Lemaire, P. Evolutionary crossroads in developmental biology: the tunicates. *Development.* **138**, 2143-2152, doi: 10.1242/dev.048975 (2011).
9. Yu, J.K., Holland, L.Z. Cephalochordates (amphioxus or lancelets): a model for understanding the evolution of chordate characters. *Cold Spring Harb. Protoc.* **2009**, pdb.emo130, doi: 10.1101/pdb.emo130 (2009).
10. Fuentes, M., *et al.* Insights into spawning behavior and development of the European amphioxus (*Branchiostoma lanceolatum*). *J. Exp. Zool. B.* **308**, 484-493, doi: 10.1002/jez.b.21179 (2007).
11. Li, G., Shu, Z., Wang, Y. Year-round reproduction and induced spawning of Chinese amphioxus, *Branchiostoma belcheri*, in laboratory. *PLoS One.* **8**, e75461, doi: 10.1371/journal.pone.0075461 (2013).
12. Theodosiou, M., *et al*. Amphioxus spawning behavior in an artificial seawater facility. *J. Exp. Zool. B.* **316**, 263-275, doi: 10.1002/jez.b.21397 (2011).
13. Holland, L.Z., Yu, J.K. Cephalochordate (amphioxus) embryos: procurement, culture, and basic methods. *Methods Cell Biol.* **74**, 195-215, doi: 10.1016/S0091-679X(04)74009-1 (2004).
14. Holland, L.Z., Onai T. Analyses of gene function in amphioxus embryos by microinjection of mRNAs and morpholino oligonucleotides. *Methods Mol. Biol.* **770**, 423-438, doi: 10.1007/978-1-61779-210-6_16 (2011).
15. Liu, X., Li, G., Feng, J., Yang, X., Wang, Y.Q. An efficient microinjection method for unfertilized eggs of Asian amphioxus *Branchiostoma belcheri*. *Dev. Genes Evol.* **223**, 269-278, doi: 10.1007/s00427-013-0441-0 (2013).
16. Harvey, K.J., Lukovic, D., Ucker, D.S. Membrane-targeted green fluorescent protein reliably and uniquely marks cells through apoptotic death. *Cytometry.* **43**, 273-278, doi: 10.1002/1097-0320(20010401)43:4<273::AID-CYTO1059>3.0.CO;2-3 (2001).
17. Maruyama, J., Nakajima, H., Kitamoto, K. Visualization of nuclei in *Aspergillus oryzae* with EGFP and analysis of the number of nuclei in each conidium by FACS. *Biosci. Biotechnol. Biochem.* **65**, 1504-1510, doi: 10.1271/bbb.65.1504 (2001).
18. Rupp, R.A., Snider, L., Weintraub, H. *Xenopus* embryos regulate the nuclear localization of XMyoD. *Genes Dev.* **8**, 1311-1323, doi: 10.1101/gad.8.11.1311 (1994).
19. Turner, D.L., Weintraub, H. Expression of achaete-scute homolog 3 in *Xenopus* embryos converts ectodermal cells to a neural fate. *Genes Dev.* **8**, 1434-1447, doi: 10.1101/gad.8.12.1434 (1994).
20. Nakagawa, S., Niimura, Y., Gojobori, T., Tanaka, H., Miura, K. Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. *Nucleic Acids Res.* **36**, 861-871, doi: 10.1093/nar/gkm1102 (2008).
21. Nakamura, Y., Gojobori, T., Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**, 292, doi: 10.1093/nar/28.1.292 (2000).
22. Deheyn, D.D., *et al.* Endogenous green fluorescent protein (GFP) in amphioxus. *Biol. Bull.* **213**, 95-100, doi: 10.2307/25066625 (2007).
23. Schubert, M., *et al*. Retinoic acid signaling acts via *Hox1* to establish the posterior limit of the pharynx in the chordate amphioxus. *Development.* **132**, 61-73, doi: 10.1242/dev.01554 (2005).
24. Schubert, M., Holland, N.D., Laudet, V., Holland, L.Z. A retinoic acid-*Hox* hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. *Dev. Biol.* **296**, 190-202, doi: 10.1016/j.ydbio.2006.04.457 (2006).
25. Yu, J.K., Holland, N.D., Holland, L.Z. Tissue-specific expression of *FoxD* reporter constructs in amphioxus embryos. *Dev. Biol.* **274**, 452-461, doi: 10.1016/j.ydbio.2004.07.010 (2004).
26. Beaster-Jones, L., Schubert, M., Holland, L.Z. *Cis*-regulation of the amphioxus *engrailed* gene: insights into evolution of a muscle-specific enhancer. *Mech. Dev.* **124**, 532-542, doi: 10.1016/j.mod.2007.06.002 (2007).
27. Holland, L.Z., *et al.* The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res.* **18**, 1100-1111, doi: 10.1101/gr.073676.107 (2008).

28. Urasaki, A., Mito, T., Noji, S., Ueda, R., Kawakami, K. Transposition of the vertebrate *Tol2* transposable element in *Drosophila melanogaster*. *Gene.* **425**, 64-68, doi: 10.1016/j.gene.2008.08.008 (2008).
29. Li, G., *et al.* Mutagenesis at specific genomic loci of amphioxus *Branchiostoma belcheri* using TALEN method. *J. Genet. Genomics.* **41**, 215-219, doi: 10.1016/j.jgg.2014.02.003 (2014).
30. Zhang, Q.J., *et al.* Continuous culture of two lancelets and production of the second filial generations in the laboratory. *J. Exp. Zool. B.* **308**, 464-472, doi: 10.1002/jez.b.21172 (2007).
31. Yasui, K., Igawa, T., Kaji, T., Henmi, Y. Stable aquaculture of the Japanese lancelet *Branchiostoma japonicum* for 7 years. *J. Exp. Zool. B.* **320**, 538-547, doi: 10.1002/jez.b.22540 (2013).
32. Benito-Gutiérrez, E., Weber, H., Bryant, D.V., Arendt, D. Methods for generating year-round access to amphioxus in the laboratory. *PLoS One.* **8**, e71599, doi: 10.1371/journal.pone.0071599 (2013).

Materials List for:

Expression of Fluorescent Proteins in *Branchiostoma lanceolatum* by mRNA Injection into Unfertilized Oocytes

Estelle Hirsinger[1], João Emanuel Carvalho[2], Christine Chevalier[1,3], Georges Lutfalla[4], Jean-François Nicolas[1], Nadine Peyriéras[5], Michael Schubert[2]

[1]Département de Biologie du Développement et Cellules Souches, Institut Pasteur

[2]Laboratoire de Biologie du Développement de Villefranche-sur-Mer (UMR7009 CNRS/UPMC Univ Paris 06), Sorbonne Universités

[3]Equipe Epigenetic Control of Normal and Pathological Hematopoiesis, Centre de Recherche en Cancérologie de Marseille

[4]Unité de Dynamique des Interactions Membranaires Normales et Pathologiques, CNRS UMR5235/DAA/cc107/Université Montpellier II

[5]Plateforme BioEmergences IBiSA FBI, CNRS-NED, Institut de Neurobiologie Alfred Fessard

Correspondence to: Nadine Peyriéras at nadine.peyrieras@inaf.cnrs-gif.fr

URL: http://www.jove.com/video/52042
DOI: doi:10.3791/52042

Materials

Name	Company	Catalog Number	Comments
Consumables			
35 mm Petri dishes	Falcon	353001	Culture-treated
Filtration unit (Stericup 1 L)	Fisher	W21719	0.22 µm
Spin-X tubes	Costar	8160	0.22 µm
Needle storage jar for 1.2 mm diameter capillaries	WPI	E212	
Pasteur pipettes			230 mm long
Aspiration tube	Dutscher	75056	
Capillaries for injection needles	Sutter	BF 120-94-10	Borosilicate glass with filament, OD 1.20 mm, ID 0.94 mm, length 10 mm
Reagents			
Low-melting agarose	Sigma	A9414	
Phenol Red	Sigma	114537	
Glycerol	Sigma	G2025	
Poly-L-lysine hydrobromide	Sigma	P9155	
H_2O, DNase/RNase-free	Gibco	10977-035	
mMessage mMachine SP6 Transcription kit	Ambion	AM1340	mRNA synthesis kit
Phenol pH 8	Sigma	P4557	
24:1 chloroform:isoamylic alcohol	Sigma	C0549	
5:1 phenol pH 4.7:chloroform	Sigma	P1944	
Reef Crystal salts (200 kg)	Europrix		Commercial salts
$NaHCO_3$	Sigma	S6014	
Equipment			
Fluorescent dissecting scope with 200X magnification	Leica	MZ16F	25X oculars, DSR and GFP2 filters
Micromanipulator	Marzhauzer	M-33	
Injector	Picospritzer	model II or III	
Needle puller	Sutter	P97	Heating-filament needle puller
Fine forceps	Fine Science Tools GmbH	11252-30	Dumont #5

A

eGFP:CAAX box
eGFP: from pEGFP-1 vector (GenBank U55761)
CAAX box: membrane localization signal from human HRAS (GenBank BT019421)
Mutated sites are indicated in bold and underlined.

```
1     CGCCATTCTGCCTGGGGACGTCGGAGCAAGCTTGATTTAGGTGACACTAT
51    AGAATACAAGCTACTTGTTCTTTTTGCAACCGTCAACATGGTGAGCAAGG
101   GCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC
151   GACGTGAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGC
201   CACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC
251   CCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGC
301   TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGC
351   CATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACG
401   GCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTG
451   AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCT
501   GGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGG
551   CCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAAC
601   ATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCC
651   CATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCC
701   AGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTG
751   CTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTA
801   CAAGGGAGGAGGAAGGTCTAAGCTGAACCCTCCTGATGAGAGTGGCCCCG
851   GCTGCATGAGCTGCAAGTGTGTGCTCTCCTGATCTAGAACTATAGTGAGT
901   CGTATTACGTAGATCCAGACATGATAAGATACATTGATGAGTTTGGACAA
951   ACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGA
1001  TGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACA
1051  ACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAG
1101  GTGTGGGAGGTTTTTTAATTCGCGGCCGCGGCGCCAATGCATTGGGCCCG
```

B

H2B:eGFP
H2B: zebrafish histone 2B (hist2h2l) exon (Ensembl ENSDARG00000068996.2)
eGFP: from pEGFP-1 vector (GenBank U55761)
Mutated sites are indicated in bold and underlined.

```
1     CGCCATTCTGCCTGGGGACGTCGGAGCAAGCTTGATTTAGGTGACACTAT
51    AGAATACAAGCTACTTGTTCTTTTTGCAGGAACCGTCAACATGCCCGAAC
101   CTGCGAAGTCAGCGCCCGCTCCCAAAAAAGGCTCTAAAAAAGCTGTCGCC
151   AAGACCCAGAAGAAGGGGGATAAGAAAAGGCGTAAGACCAGGAAAGAGAG
201   TTACGCCATTTACGTGTACAAAGTGCTGAAACAAGTCCACCCGGACACTG
251   GCATCTCCTCAAAGGCGATGGGCATTATGAACTCATTTGTGAACGACATC
301   TTCGAGCGCATCGCCGGAGAAGCGTCGCGCCTGGCGCATTACAACAAGCG
351   CTCCACTATCACATCCCGGGAGATCCAGACGGCCGTGCGCCTGCTCCTGC
401   CCGGAGAACTGGCCAAACACGCTGTGTCTGAGGGCACAAAGGCCGTGACC
451   AAGTACACCAGCTCCAAGGTGCCGCGGGCCCGGGATCCACCGGTCGCCAC
501   CATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
551   TCGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGTCCGGCGAG
601   GGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGG
651   CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCG
701   TGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTC
751   AAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
801   GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACA
851   CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGC
901   AACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTA
951   TATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCC
1001  GCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAG
1051  AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCT
1101  GAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACA
1151  TGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
1201  GAGCTGTACAAGTAAAGCGGCCGCGACTCTAGAACTATAGTGAGTCGTAT
1251  TACGTAGATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCAC
1301  AACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTA
1351  TTGCTTTATTTGTAACCATTTATAAGCTGCAATAAACAAGTTAACAACAA
1401  CAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTT
1451  TTTAATTCGCGGCCGCGGCGCCAATGCATTGGGCCCGGTACCCAGCTTTT
```

C

H2B:mCherry
H2B: zebrafish histone 2B (hist2h2l) exon (Ensembl ENSDARG00000068996.2)
mCherry: from pRSET-B-mCherry plasmid (the Forchheimer Center plasmid collection)
Mutated sites are indicated in bold and underlined.

```
1     CGCCATTCTGCCTGGGGACGTCGGAGCAAGCTTGATTTAGGTGACACTAT
51    AGAATACAAGCTACTTGTTCTTTTTGCAGGAACCGTCAACATGCCCGAAC
101   CTGCGAAGTCAGCGCCCGCTCCCAAAAAAGGCTCTAAAAAAGCTGTCGCC
151   AAGACCCAGAAGAAGGGGGGATAAGAAAAGGCGTAAGACCAGGAAAGAGAG
201   TTACGCCATTTACGTGTACAAAGTGCTGAAACAAGTCCACCCGGACACTG
251   GCATCTCCTCAAAGGCGATGGGCATTATGAACTCATTTGTGAACGACATC
301   TTCGAGCGCATCGCCGGAGAAGCGTCGCGCCTGGCGCATTACAACAAGCG
351   CTCCACTATCACATCCCGGGAGATCCAGACGGCCGTGCGCCTGCTCCTGC
401   CCGGAGAACTGGCCAAACACGCTGTGTCTGAGGGCACAAAGGCCGTGACC
451   AAGTACACCAGCTCCAAGGTGCCGCGGGCCCGGGATCCCGCCACCATGGT
501   GAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCT
551   TCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAG
601   GGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAA
651   GGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTC
701   AGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC
751   GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGAT
801   GAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGC
851   AGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCC
901   TCCGACGGCCCCGTGATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTC
951   CGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGA
1001  GGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACC
1051  TACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACAT
1101  CAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGT
1151  ACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTAC
1201  AAGTAAGAATTCAAGGCCTCTCGAGCCTCTAGAACTATAGTGAGTCGTAT
1251  TACGTAGATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCAC
1301  AACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTA
1351  TTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAAC
1401  AATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTT
1451  TTAATTCGCGGCCGCGGCGCCAATGCATTGGGCCCGGTACCCAGCTTTTG
```

Appendix 8

Published article

The human Mixed Lineage Leukemia 5 (MLL5), a sequentially and structurally divergent SET domain-containing protein with no intrinsic catalytic activity.

Mas-y-Masa S, Barbona M, Teyssierd C, Déménéa H, Carvalho JE, Lebedevf A, Fattorig J, Schubert M, Dumasa C, Bourgueta W, le Maire A.

PLoS One. 2016;11(11): e0165139.

(20 pages)

PLOS ONE

RESEARCH ARTICLE

The Human Mixed Lineage Leukemia 5 (MLL5), a Sequentially and Structurally Divergent SET Domain-Containing Protein with No Intrinsic Catalytic Activity

Sarah Mas-y-Mas[1,2,3☯], Marta Barbon[1,2,3☯¤], Catherine Teyssier[3,4], Hélène Déméné[1,2,3], João E. Carvalho[5], Louise E. Bird[6], Andrey Lebedev[7], Juliana Fattori[8], Michael Schubert[5], Christian Dumas[1,2,3], William Bourguet[1,2,3], Albane le Maire[1,2,3,8]*

1 Inserm U1054, Centre de Biochimie Structurale, Montpellier, France, **2** CNRS UMR5048, Centre de Biochimie Structurale, Montpellier, France, **3** Université de Montpellier, Montpellier, France, **4** IRCM, Institut de Recherche en Cancérologie de Montpellier, Montpellier, France, **5** Sorbonne Universités, UPMC Université Paris 06, CNRS, Laboratoire de Biologie du Développement de Villefranche-sur-Mer, Observatoire Océanologique de Villefranche-sur-Mer, Villefranche-sur-Mer, France, **6** OPPF-UK, Research Complex at Harwell, Rutherford Appleton Laboratory, Oxfordshire, OX11 0FA, United Kingdom, **7** CCP4, Research Complex at Harwell, Rutherford Appleton Laboratory, Oxfordshire, OX11 0FA, United Kingdom, **8** Centro Nacional de Pesquisa em Energia e Materiais, Laboratório Nacional de Biociências, Campinas, SP, Brazil

☯ These authors contributed equally to this work.
¤ Current address: DNA Replication Group, Institute of Clinical Science, Imperial College, London, United Kingdom
* albane.lemaire@cbs.cnrs.fr

Citation: Mas-y-Mas S, Barbon M, Teyssier C, Déméné H, Carvalho JE, Bird LE, et al. (2016) The Human Mixed Lineage Leukemia 5 (MLL5), a Sequentially and Structurally Divergent SET Domain-Containing Protein with No Intrinsic Catalytic Activity. PLoS ONE 11(11): e0165139. doi:10.1371/journal.pone.0165139

Editor: Albert Jeltsch, Universität Stuttgart, GERMANY

Received: September 12, 2016

Accepted: October 6, 2016

Published: November 3, 2016

Data Availability Statement: The structure is available at the PDB (ID: 5HT6).

Funding: This work was supported by a research grant from the Association pour la Recherche sur le Cancer (ARC) (to A.l.M). J.E.C. is a FCT doctoral fellow (SFRH/BD/86878/2012).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Mixed Lineage Leukemia 5 (MLL5) plays a key role in hematopoiesis, spermatogenesis and cell cycle progression. Chromatin binding is ensured by its plant homeodomain (PHD) through a direct interaction with the N-terminus of histone H3 (H3). In addition, MLL5 contains a Su(var)3-9, Enhancer of zeste, Trithorax (SET) domain, a protein module that usually displays histone lysine methyltransferase activity. We report here the crystal structure of the unliganded SET domain of human MLL5 at 2.1 Å resolution. Although it shows most of the canonical features of other SET domains, both the lack of key residues and the presence in the SET-I subdomain of an unusually large loop preclude the interaction of MLL5 SET with its cofactor and substrate. Accordingly, we show that MLL5 is devoid of any *in vitro* methyltransferase activity on full-length histones and histone H3 peptides. Hence, the three dimensional structure of MLL5 SET domain unveils the structural basis for its lack of methyltransferase activity and suggests a new regulatory mechanism.

Introduction

MLL5 was initially identified as a candidate tumor suppressor gene located in the commonly deleted 2.5 Mb segment of human chromosome band 7q22 in myeloid malignancies [1].

Knock-out studies in mice showed that the murine MLL5 protein is required in adult hematopoiesis [2–4] and normal spermatogenesis [5]. Overexpression and knockdown of MLL5 both induce cell cycle arrest at various phases, suggesting a versatile function of MLL5 throughout the cell cycle [6]. A recent study showed that MLL5 associates with chromatin regions downstream of transcriptional start sites of active genes and provided insights into the regulation of this association during mitosis and development [7].

Full-length MLL5 is a 205 kDa protein (1858 residues), although a smaller isoform (around 75 kDa) including a single Plant HomeoDomain (PHD) and a Su(var)3-9, Enhancer of zeste, Trithorax (SET) domain has been described [1]. The C-terminal region of MLL5, truncated in the shorter isoform, displays no apparent homology to any known structural domain and is predicted to be disordered in solution. The MLL5 protein is well conserved between species, with the human sequences for example, showing 80% and 38% identity with the PHD and SET domains of its *Drosophila* counterpart UpSET, respectively. Interestingly, both the *Drosophila* UpSET and the Yeast ortholog SET3, which are found in histone deacetylase complexes, were shown to contain inactive SET domains [8,9]. Even if a hallmark of SET domain-containing proteins is the histone methyltransferase (HMT) activity, it remains controversial whether the human and mouse MLL5 SET domains possess a catalytic HMT activity [2,10,11]. To better understand the molecular mechanism of human MLL5 function, we solved the crystal structure of its SET domain at 2.1 Å resolution. Although some common characteristics were observed, our structure reveals significant differences with canonical SET domains endowed with catalytic activity. Moreover, a correlative analysis of 3D structures and sequences allowed us to identify MLL5 specific features unfavorable for methyltransferase activity. Accordingly, biochemical and biophysical experiments showed that the SET domain of MLL5 is devoid of any binding activity towards histone peptides or the cofactor S-Adenosyl Methionine (SAM) *in vitro* and does not exhibit HMT activity on full-length histones. Having provided compelling evidence that MLL5 is not a functioning enzyme, the biological role of this protein remains to be established.

Results

The MLL5 SET domain lacks most of the residues important for HMT activity

The MLL5 protein contains a SET domain (residues 323 to 433) followed by a POSTSET domain (residues 434 to 473). The prominent feature of the POSTSET domain is a zinc-binding cage formed from three cysteine residues of the C-terminal region with a fourth cysteine provided by the SET-C subdomain (Fig 1A). A common feature of SET domains is a channel through the protein linking the cofactor (SAM) binding surface on one side with the substrate binding surface on the other [12]. This channel, made up of residues from the SET and POST-SET regions, encloses the lysine residue of the substrate and holds it in an appropriate chemical environment and position for methyl transfer to take place (S1 Fig).

Sequence analysis of the MLL5 SET domain suggests that it does not contain all the conserved sequence elements required for methyltransferase activity. Contrary to other SET-domain containing methyltransferases, MLL5 does not have the residues usually involved in cofactor binding (Fig 1A). Instead of the highly conserved XGXG, Y and NH motifs, MLL5 displays NKKI (Asn 339-Ile 342), F (Phe 381) and RR (Arg 408—Arg 409) motifs. Key residues involved in histone H3 recognition and in the active center are also poorly conserved in MLL5 (Fig 1A). In particular, of the three conserved tyrosine residues in the active site of SET domains, two are replaced by other residues (Ile 444 and Phe 446 in MLL5). These substitutions are very likely to either significantly reduce the catalytic activity or to affect the mono, di-

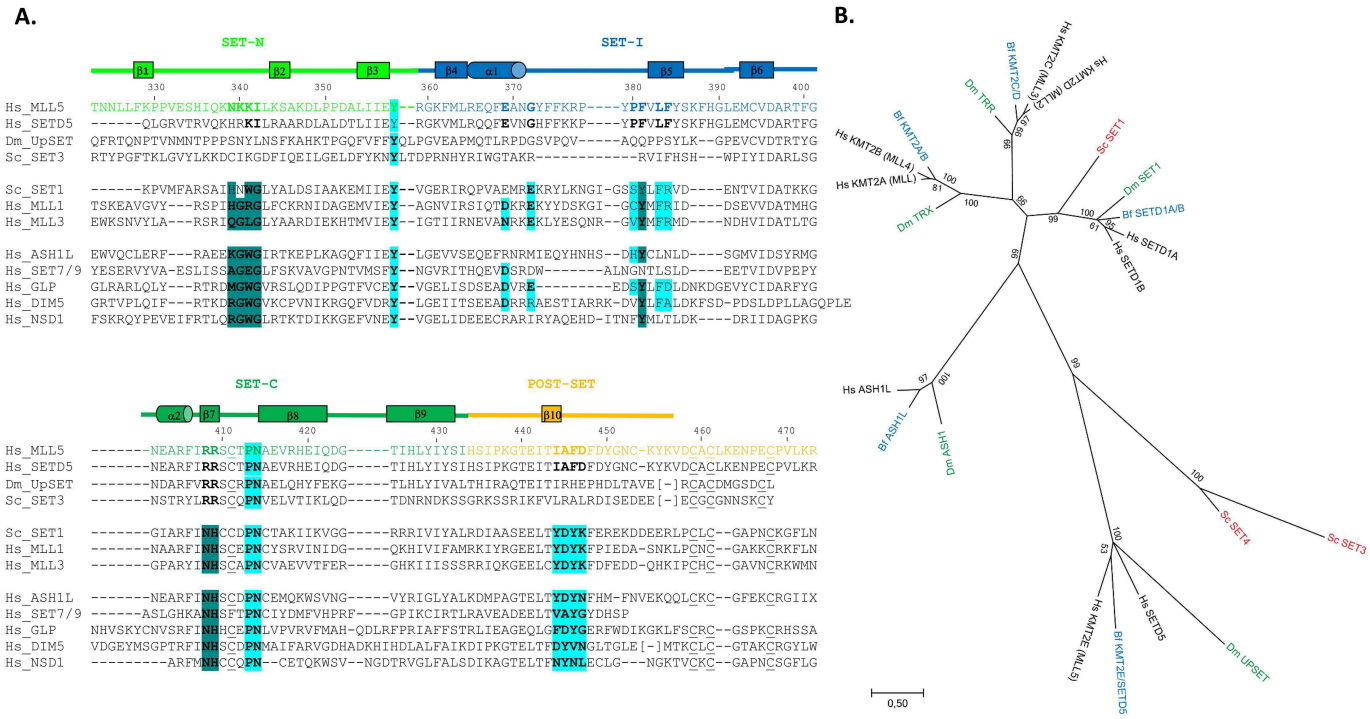


Fig 1. MLL5 SET domain lacks important residues for methyltransferase activity. (A). Sequence alignment of the human MLL5 SET domain with SET domains from other methyltransferases (human SETD5, *Drosophila* UPSET, yeast SET3, yeast SET1, human MLL1, human MLL3, human ASH1L, human SET7/9, human GLP, human DIM5 and human NSD1). Structural alignment was calculated using the Dali server [56] when structure of the SET domain were available. Protein subdomains, indicated above the sequences, are based on the human MLL1 SET domain [14]. Secondary structure elements, derived from the structure of the MLL5 SET domain, are also shown above the sequences. Residues usually important for cofactor binding (dark blue) and for substrate binding and maintenance in the active center (cyan) are highlighted. Conserved cysteines from the zinc-binding cage are underlined. **(B)**. Phylogenetic tree of methyltransferases related to MLL5 using sequences from humans (*Homo sapiens*) (Hs, black), the cephalochordate amphioxus (*Branchiostoma floridae*) (Bf, blue), the fruit fly (*Drosophila melanogaster*) (Dm, green) and the yeast (*Saccharomyces cerevisiae*) (Sc, red). The tree resulting from the Maximum Likelihood (ML) analysis is shown and branch lengths are representative of sequence substitution rates. Branch support is indicated as Bootstrap percentages (ranging from 0 to 100) and as posterior probabilities (ranging from 0 to 1) obtained from the Bayesian Inference (BI) analysis (provided in parentheses at each node). "(—)" indicates that the branching patterns of the ML and BI analyses diverged at this node and that a posterior probability score is thus unavailable.

doi:10.1371/journal.pone.0165139.g001

or tri-methylation specificity of the MLL5 SET domain [13]. Altogether, these important sequence differences strongly suggest that the MLL5 SET domain displays altered functional properties, when compared to SET domains of other proteins.

The MLL5 protein is definitively not a member of the MLL family

In mammals, the Mixed Lineage Leukemia (MLL) family of SET domain-containing methyltransferases includes at least six members (MLL1, MLL2, MLL3, MLL4, SETD1A and SETD1B), all of which are involved in gene activation [14]. All members of this family share a distinct SET domain along with various numbers of PHD domains and catalyze the addition of methyl groups to the ε-amino moiety of lysine. Importantly, MLL family proteins are highly specific for lysine 4 of histone H3, which is mediated by their SET domain [15]. MLL5 was originally included in this MLL family due to its sequence conservation with the SET domain of MLL1 [1], although it was always considered as a distantly related member. Our phylogenetic analysis of human, *Drosophila*, yeast and amphioxus sequences shows that MLL5 and SETD5 form a family that is clearly distinct from the MLL1-4, SETD1A, B group (S2 Fig, Fig 1B). Moreover, while yeast SET1 is very closely related to human SETD1A, B and thus to

human MLL1-4, yeast SET3 and SET4 are orthologs of human MLL5 and SETD5. Similarly, *Drosophila* UpSET belongs to the MLL5/SETD5 family. The presence of MLL5/SETD5 proteins in yeast, flies and humans indicates that this family is evolutionarily very ancient and was already present in the last common ancestor of fungi and animals (Fig 1B).

Like MLL5, yeast SET3 and SET4 as well as *Drosophila* UpSET also lack the XGXG, Y and NH catalytic sequence signatures present in MLL1-4 (Fig 1A) [15], suggesting that all members of the MLL5/SETD5 clade are characterized by a selective loss of catalytic activity. This notion is supported by functional analyses of yeast SET3 and *Drosophila* UpSET, which demonstrated that both proteins lack methyltransferase activity [8,9]. Although the catalytic capacities of yeast SET4 and human SETD5 remain to be determined, it is very likely that the loss of catalytic activity is characteristic of all MLL5 and SETD5 sequences and thus took place soon after the evolutionary origin of this family.

The isolated MLL5 SET domain is not able to bind the S-adenosyl-methionine and the H3K4 histone peptide

Although the MLL5 SET domain lacks some essential residues for methyltransferase activity, we investigated whether it was capable of interacting with the substrate or the cofactor SAM, two binding events necessary for catalysis. For these functional studies, we prepared a MLL5 fragment including the SET domain and the POSTSET region at the C-terminus (SET-POSTSET, residues 323–473) as well as a longer fragment containing an additional N-terminal helix (Nh-SET-POSTSET, residues 262–473), which was previously shown to be important for the structural integrity of the SET domain [13,16]. Isothermal titration calorimetry (ITC) measurements indicated that both fragments failed to bind the cofactor SAM (Fig 2A), and thermal denaturation shift assays revealed no stabilization of both SET fragments even in the presence of 20-fold molar excess of SAM (data not shown), thus confirming the ITC results.

Similarly, we ran fluorescence anisotropy experiments to measure the affinity of the presumed substrate H3K4 peptide (ARTKQTARKST), with un-methylated or mono-methylated lysine 4 [7], for the SET-POSTSET fragment. These measurements clearly showed that the interaction between the MLL5 SET domain and both histone H3 peptides is extremely weak (Fig 2B).

To ensure that the MLL5/H3K4 peptide interaction was not inhibited by the fluorescein moiety attached to the N-terminus of the peptides, we conducted NMR titration and saturation transfer difference (STD) experiments to monitor the interaction between the SET-POSTSET fragment and the unlabeled non-, mono- and di-methylated H3K4 peptides. Titration and STD analyses were performed by increasing the protein/peptide ratio from 1/20 to 1/5 in a stepwise manner. As shown in Fig 2C (upper panel), the line widths and the position of proton resonances of the un-methylated peptide remained unchanged, even at the highest protein concentration, although in the case of a low affinity complex, any interaction with the protein would cause a small line width increase and/or a shift of some proton resonances in the peptide spectrum [17]. STD experiments (Fig 2C lower panel), which are very sensitive to the presence of low affinity complexes (Kd in the 10^{-4}–10^{-7} M range) did not reveal any resonances of atoms from the peptide in interaction with the protein. Identical results were obtained with all forms of the H3K4 peptide.

As we did not observe any interaction of the MLL5 SET domain with non-, mono- or di-methylated H3K4 peptides, we sought to identify other histone substrates for the MLL5 SET domain using GST-pull down and a mixture of histones from calf thymus. As shown in S3 Fig, no binding could be observed between the GST-fused SET-POSTSET fragment and any component of the histone mixtures.

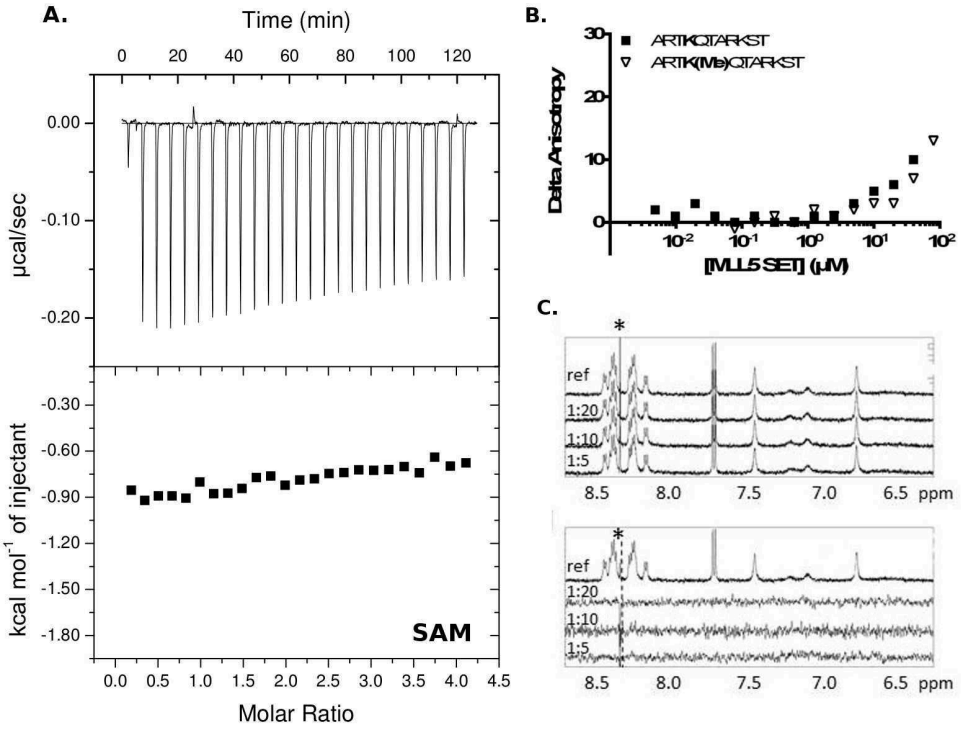


Fig 2. Absence of cofactor and substrate (H3K4 peptide) binding to the MLL5 SET domain. (A). Isothermal titration calorimetry shows that the MLL5 SET domain in combination with the N-terminal flanking helix and the POSTSET does not bind the cofactor S-adenosyl methionine (SAM). Results were the same with the fragment SET-POSTSET. **(B)**. Fluorescence anisotropy shows that the MLL5 SET domain (SET-POSTSET construct) does not bind either the non-methylated or mono-methylated H3K4 peptide. **(C)**. ^{1}H NMR spectra of SET-POSTSET construct and un-methylated H3K4 peptide. Upper panel: Expansion of ^{1}H NMR spectra of un-methylated H3K4 peptide (top) at 400 µM, and in presence of different ratios protein/peptide (1:20, 1:10, 1:5). Lower panel: Expansion of ^{1}H (top) and saturation transfer difference (STD) spectra of un-methylated H3K4 peptide at 400 µM in presence of different ratios of protein. In all spectra, the fine peak labeled with an asterisk (*), if present, corresponds to an impurity, whose resonance is imperfectly subtracted in STD on- and off-resonance experiments.

doi:10.1371/journal.pone.0165139.g002

Taken together, our data indicate that the isolated MLL5 SET domain in combination with the POSTSET region fails to bind either the cofactor SAM or any histone substrate. In addition, the presence of the N-terminal flanking region is not helpful for the SAM binding.

The MLL5 SET domain does not exhibit any methyltranferase activity

Next, we performed an *in vitro* histone methyltransferase assay (HMT) to address whether the MLL5 SET domain is enzymatically functional. We followed the same protocol as the one used in the study of UpSET, the Drosophila ortholog of MLL5, showing that this SET domain does not exhibit histone methyltransferase activity [9]. We found that the bacterially expressed and purified Nh-SET-POSTSET fragment does not have HMT activity (Fig 3), either alone or in presence of nuclear extracts to allow the presence of additional proteins that could be necessary for a potential methyltransferase activity of MLL5. In fact, whereas the SET7 protein shows a strong methyltransferase activity as expected [18], the MLL5 fragment, alone or after incubation with mammalian nuclear extracts, exhibits a level of activity comparable to the negative control.

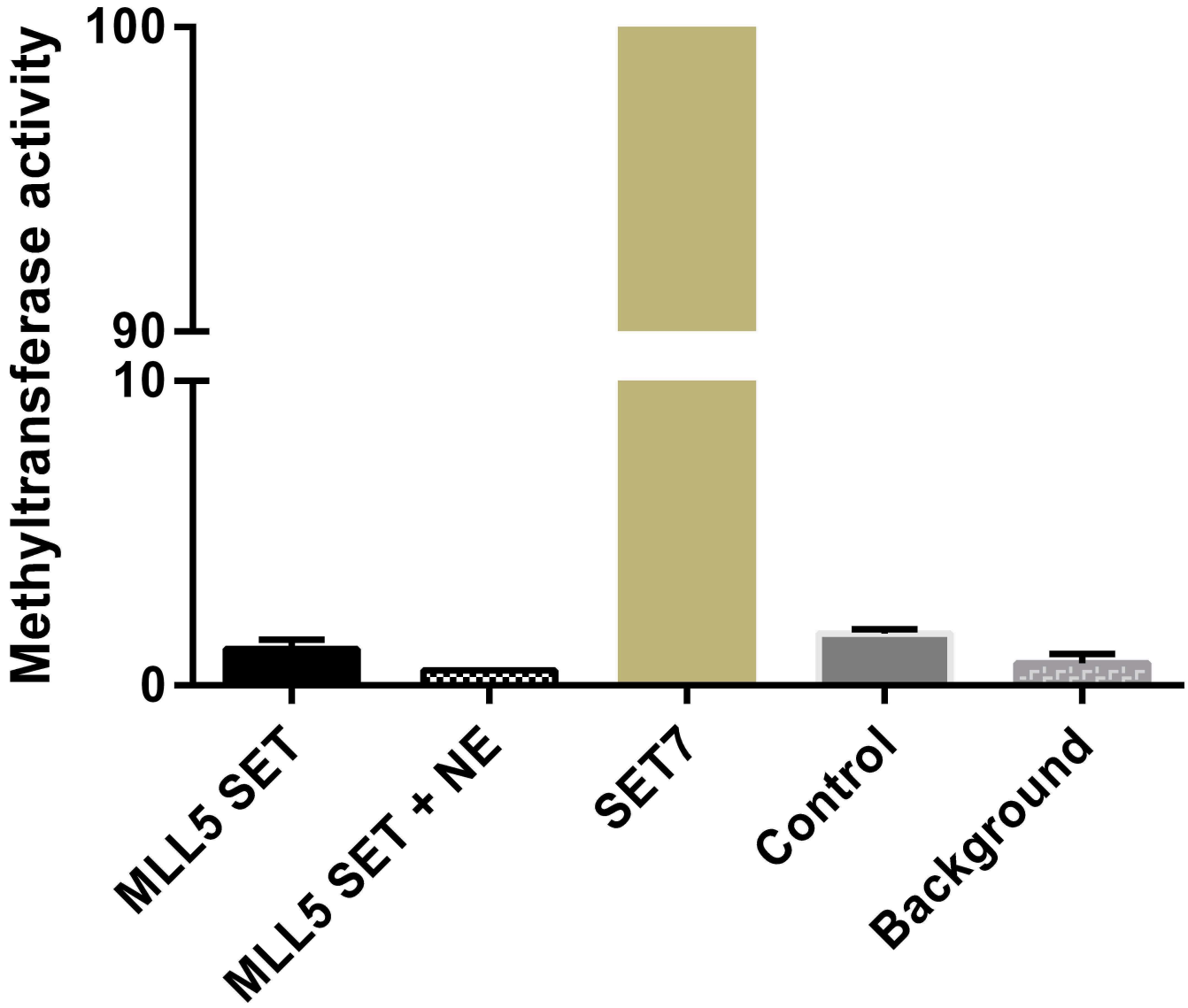


Fig 3. The MLL5 SET domain does not possess histone methyltransferase (HMT) activity. Radioactive HMT assays were performed with the bacterially purified Nh-SET-POSTSET fragment and histones from calf thymus as substrate. Recombinant proteins were incubated with HeLa nuclear extracts (+ NE) or not. SET7 was used as positive control (set as 100% activity) and the same condition in absence of any enzyme was used as negative control (Control).

doi:10.1371/journal.pone.0165139.g003

The overall MLL5 SET domain structure

In order to better understand the molecular basis for the absence of HMT activity in the MLL5 SET domain, we solved its crystal structure. Attempts to crystallize the Nh-SET-POSTSET and SET-POSTSET fragments were unsuccessful. NMR studies (S4 and S5 Figs, S1 File) showed that both the N-terminal helix flanking region and the POSTSET region display substantial dynamics, but did not induce major changes in the structure of the SET domain. Such a flexibility, already observed in other SET-POSTSET proteins (like EZH2, [19]), would explain the difficulty to crystallize the longer fragments of the MLL5 SET-POSTSET domain.

The SET domain alone (residues 323 to 458) crystallized in the space group $P3_121$, and diffracted up to 2.1 Å resolution (Table 1, Fig 4A). To allow crystallization, we mutated an exposed and non-conserved cysteine (Cys 453) into alanine since this residue caused heterogeneity of the protein due to dimerization by formation of an intermolecular disulfide bridge. The crystal structure was determined by molecular replacement with many difficulties. First, most of the datasets collected could be integrated in the $P3_121$ space group with two different unit cell parameters (a = b = 66.2 Å, c = 56.7 Å, or a' = b' = 66.1 Å, c' = 113.3Å), due to the presence of a pseudo-translation along the c axis. A pool of 16 homologous SET models with less than 30% sequence identity was selected for initial molecular replacement trials. The best template models were then edited based on the local fit to the density maps to produce an ensemble of viable search models. Finally, by combining molecular replacement with this ensemble model in the large cell using Phaser-MR [20] and cyclic iterations of manual rebuilding with Coot [21] followed by auto-rebuilding with Buccaneer [22] and Parrot [23], we managed to generate an hybrid model resulting in a good molecular replacement solution (S6 Fig). The initial R / Rfree factors calculated after rigid body and restrained refinement steps with Refmac [24] were equal to 27% / 34%. The final crystal structure of the SET domain of MLL5 in its free form was refined at 2.1 Å resolution with an R-factor equal to 20.9% and R-free equal to 24.2% with good stereochemistry (Table 1).

The asymmetric unit contained two unliganded MLL5 molecules which superposed with a root mean square deviation (RMSD) of 0.64 Å (110 Cα atoms). Molecule A, which includes most of the SET polypeptide chain (residues 327–331, 340–457) was used for the structural

Table 1. Data collection and structure refinement statistics.

	PDB entry 5HT6
Data collection	
Space Group	$P3_1$ 2 1
Cell dimensions	
a, b, c (Å)	65.90, 65.90, 112.90
α, β, γ (˚)	90.00, 90.00, 120.00
Resolution (Å)	50.93–2.09 (2.21–2.09)*
Rsym	0.06 (0.39)*
I/σI	5.87 (2.0)*
Completeness (%)	99.5 (99.7)*
Refinement	
Resolution (Å)	50.93–2.10
No. reflections	17250
Rfree test set	2432 (7.55%)*
Rwork/Rfree	0.209/0.242
No. atoms	
Protein	1931
Water	80
B-factors $Å^2$	
Protein chains A/B	50.45/69.43
Water	46.79
R.m.s deviations	
Bond lengths (Å)	0.004
Bond angles (˚)	0.841

* Values in parentheses are for highest-resolution shell.

doi:10.1371/journal.pone.0165139.t001

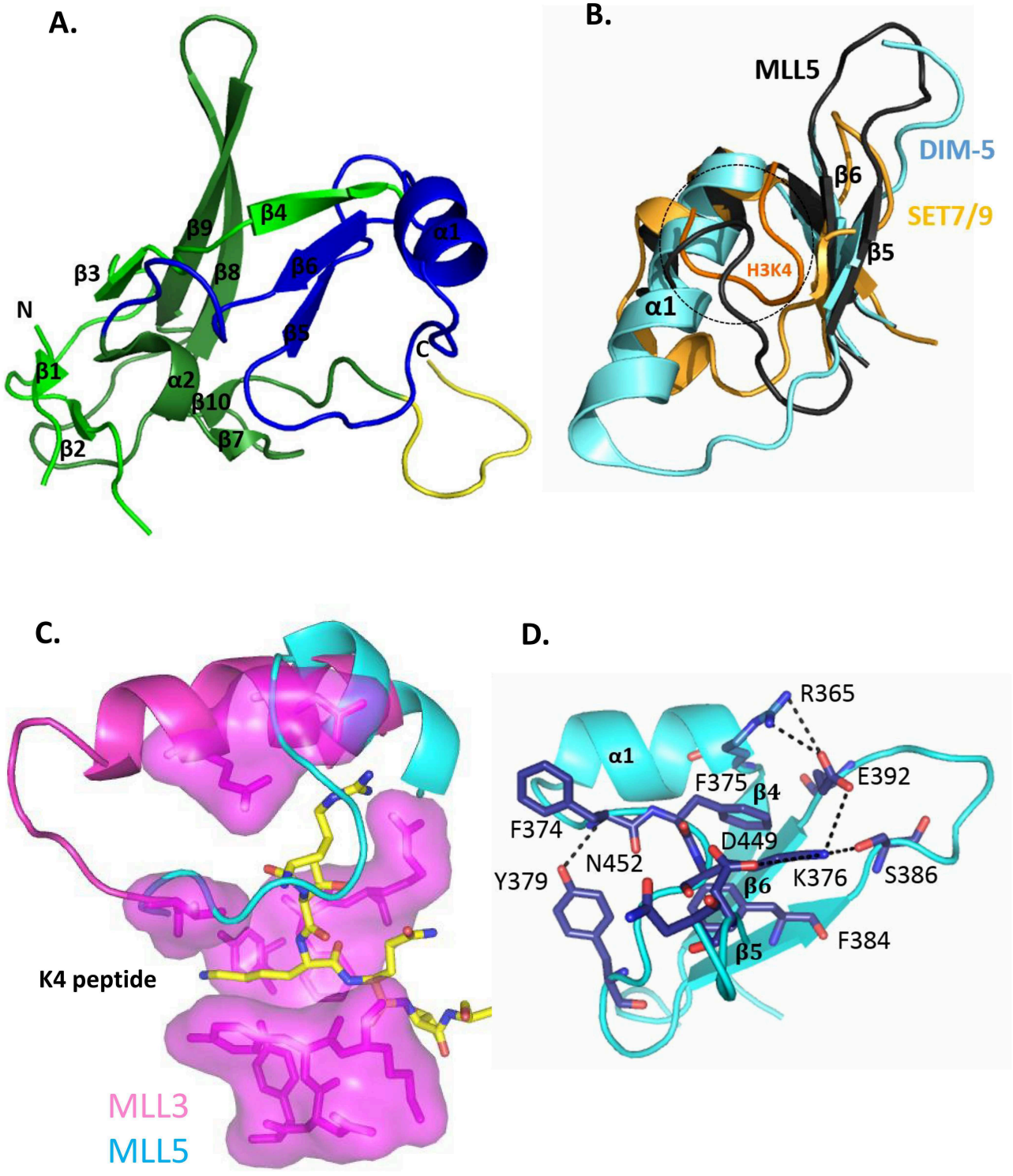


Fig 4. Crystal structure of the MLL5 SET domain highlights unexpected feature. (A). The structure is represented as a cartoon model and secondary structure elements are labeled. The subdomains are: SET-N in pale green, SET-I in blue, SET-C in bright green and the beginning of the POSTSET domain in yellow. **(B)**. Part of the MLL5-SET (dark grey), DIM-5-SET (cyan, PDB entry 1PEG) and SET7/9-SET (light orange, PDB entry 1O9S) domains are superimposed and represented by cartoons. The histone peptide in complex with the SET7/9-SET-domain is colored in orange. The loop of the MLL5-SET domain occupies the region of the substrate binding groove (highlighted in a circle). **(C)**. Groove of

K4 peptide binding of MLL3 (pink surface, PDB entry 5F6K) superimposed with the MLL5 structure (cyan) to highlight the occupation of the substrate groove by the MLL5 loop between α1 and β5. **(D)**. Zoom-out view of the interface (salt-bridge and hydrogen bonds as well as hydrophobic stacking between aromatic residues shown as dashed lines) that maintains the loop between α1 and β5 in the peptide groove.

doi:10.1371/journal.pone.0165139.g004

analysis. Molecule B had well defined electron density for SET core, however residues 327 to 339 at the N-terminus and 449 to the end at the C-terminus were disordered and could not be modeled from X-ray data. However, the MLL5 POSTSET zinc binding site could be easily modeled using the MODELLER v9.2 program [25], and the homologous domain of the SETD2 methyltransferase (pdb entry 4H12) as a template [26] (S4 Fig, S1 File). This model suggests that, together with Cys 411 from the SET domain, the three conserved cysteine residues of the POSTSET region (Cys 459, 461 and 468) form a canonical zinc-binding site.

The overall tertiary structure of the MLL5 SET domain is similar to that of previously determined SET domain structures (Fig 4A). The RMSD calculated between backbone atoms of the core residues included in the SET-N (start to Leu 364), SET-I (Arg 365 to Gly 401) and SET-C (Asn 402 to Iso 433) sub-domains of the MLL5 and MLL1 SET domains is 1.3 Å (on 112 Cα atoms). The core of the SET domain is formed by a three-stranded (β3, β8, β9) anti-parallel β-sheet pressed diagonally across a second smaller one (β4, β5, β6). The opposite face of this smaller β-sheet is decorated by an α-helix (α1). A single-turn 3_{10} helix (α2) is packed against the core on the opposite side by one two-strand anti-parallel β-sheet (β1, β2) and one parallel β-sheet (β7, β10). Finally, a typical pseudo-knot is formed by the C-terminal segment of the SET domain passing through a loop formed by a preceding sequence stretch.

The isolated MLL5 SET domain adopts a conformation that precludes SAM and substrate binding

A careful analysis of the crystal structure of the SET domain reveals several features that most likely account for the lack of catalytic activity of MLL5. A unique structural feature of the SET domain of MLL5 is a very short helix α1 (two turns, from Arg 365 to Asn 371) where other SET domains usually display a four-turn α-helix (e.g. DIM-5 and MLL3) (Fig 4B and 4C). None of the 3D prediction programs have anticipated this short α-helix, even if the presence of a helix-destabilizing residue (Gly 372) could explain the premature discontinuation of the helix.

Accordingly, the shortening of this helix results in the formation of a long loop (from Gly 372 to Phe 381) that is very uncommon in SET domains. This large loop between α1 and β5 occupies a groove which, in other SET domains, serves as a substrate docking site (Fig 4B and 4C). This groove normally mediates sequence-specific recognition and positions the substrate lysine side chain into a conserved channel that penetrates deep into the core of the SET domain (Fig 4C) and meets the cofactor at the active site (S1 Fig). In these SET domains, the long α1 helix forces the corresponding loop to pass outward of the SET domain and to delineate the substrate channel (Fig 4B and 4C). To our knowledge, only two other SET domains, from the methyltransferases SET7/9 and GLP, display such a short helix. However, contrary to MLL5, the elongated loops of these methyltransferases sit outside of the domain (Fig 4B). It is noteworthy that this MLL5-specific loop conformation is observed in both molecules of the asymmetric unit and does not rely on crystallographic constraints. In addition, residues from helix α1 and from the following loop create salt-bridge and hydrogen bonds with residues from the substrate channel (Fig 4D), maintaining it in this position. A measure of stability is the persistence of these hydrogen bonds and salt bridges throughout a molecular dynamics (MD) simulation. A 96.5% persistence over all frames of the trajectories is observed for the salt-bridge between Arg 365 and Glu 392 (S7 Fig), suggesting a limited flexibility of this region. In

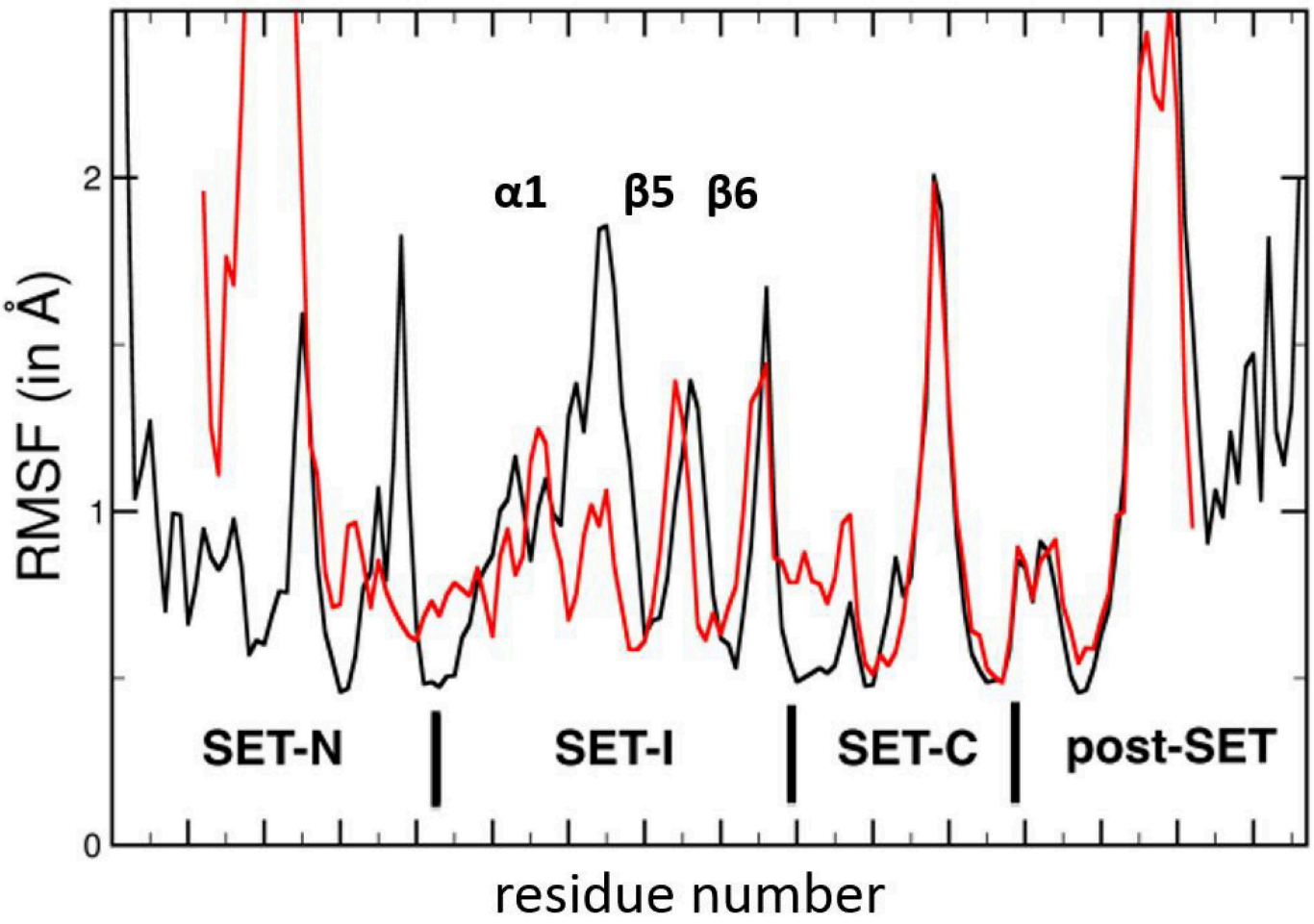


Fig 5. Dynamic properties of the unliganded MLL5 SET domain. The root mean square fluctuations (RMSF) of the SET domain residues during 100 ns of MD simulation are shown in red. The equivalent plot for the MLL3 SET domain isolated, taken from [16] and whose data where kindly sent by Y. Zhang, is shown in black.

doi:10.1371/journal.pone.0165139.g005

addition, 67% occupancy is detected for the salt bridge between Arg 376 and Asp 449, and an occupancy of 51.6% for the hydrogen bond between Lys 376 with Ser 386 and Asn 452. Our MD simulations also support the existence of a stable hydrogen bond between the phenolic hydroxyl group of Tyr 379 and the main chain nitrogen of residue 374, present for 83% of the simulation time and providing additional stability to the conformation of this loop segment. The calculation of root mean square fluctuations (RMSF) shows that the α1 helix and the following α1-β5 loop are much more constrained in MLL5 (Fig 5) than in apo-MLL1 and apo-MLL3 [16]. This suggests that the loop may not open up to accommodate the histone peptide. Moreover, normal mode analysis of apo-MLL5 did not reveal large low-frequency collective motion of the SET-I subdomain (data not shown), thus excluding local structural rearrangements of this loop. In conclusion, the elongation of the α1-β5 loop in MLL5 and its positioning in the peptide groove is very likely to preclude the binding of any substrate.

Of note, the NH motif that is usually highly conserved in SET domains and necessary for the binding of the SAM cofactor is replaced by a RR (Arg 408-Arg 409) motif in MLL5 (Figs 1 and 6). The second arginine (Arg 409) is probably able to re-orient upon cofactor binding as

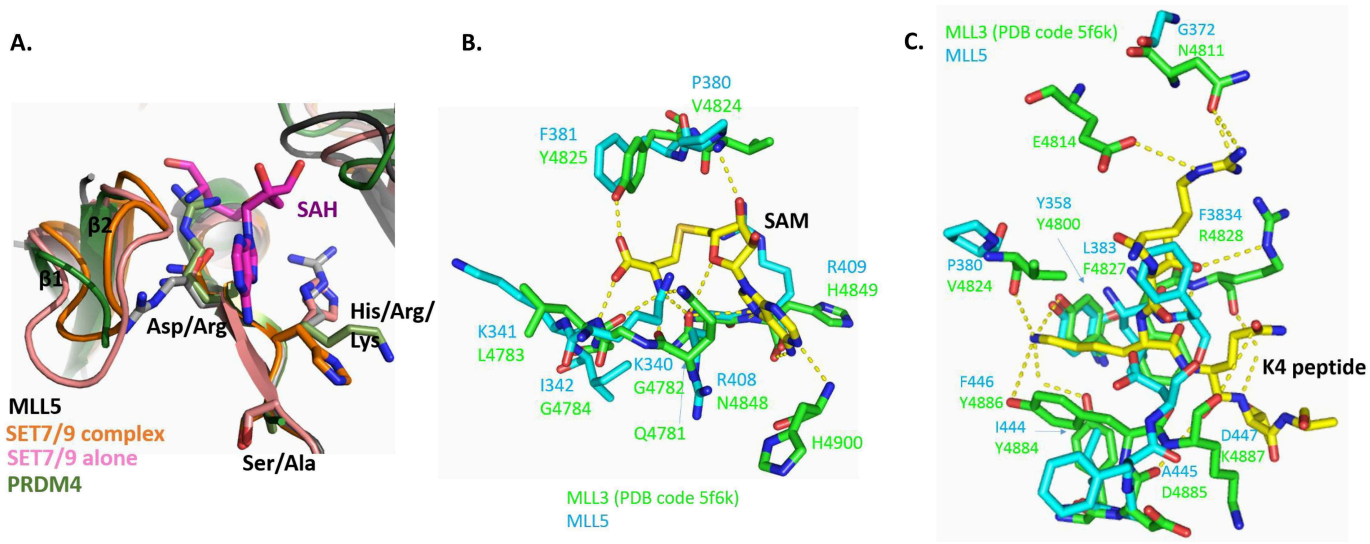


Fig 6. The S-adenosylmethionine (SAM) binding motif is not present and the active site is incomplete in the MLL5 SET domain. (A). Superposition of the motif RRS of MLL5 SET domain (black and gray sticks), the motif NHS of SET7/9 SET domain in its apo (pink, PDB entry 1H3I) form and in ternary complex (orange, PDB entry 1O9S) and of the motif RKA of PRDM4 SET domain (green, PDB entry 3DB5). **(B)**. Residues of the MLL3 SET domain (PDB entry 5F6K, green sticks) forming hydrogen bonds with the cofactor SAM (yellow sticks). Corresponding residues of the MLL5 SET domain are superimposed (cyan sticks). **(C)**. Superimposition of MLL5 residues (cyan sticks) and the MLL3 active site (PDB entry 5F6K, green sticks). Hydrogen bonds stabilizing the histone peptide (yellow sticks) are shown as dashed lines.

doi:10.1371/journal.pone.0165139.g006

illustrated by the reorientation of the corresponding residue (His 297) in SET7/9 (PDB entry 1H3I) upon substrate binding and in the presence of the cofactor SAH (PDB entry 1O9S) (Fig 6A). In the ternary complex, the histidine points towards the outside of the SET domain, a position that is also permitted for MLL5 Arg 409. On the other hand, the first residue of the motif (Asn 4848) of the MLL3 SET domain makes hydrogen bonds with the hydroxyl group of its side chain with three atoms of the cofactor SAM (Fig 6B). The replacement of this residue by an arginine (Arg 408) in the MLL5 SET domain precludes the formation of these hydrogen bonds (Fig 6B). In addition, this first arginine of the motif (Arg 408) in the MLL5 SET domain appears to prevent the proper positioning of the loop between the two first strands of the domain (β1 and β2). Indeed, the β1-β2 loop is disordered in the crystal structure, suggesting a high flexibility of this region. This loop which is well defined in both structures of SET7/9 in its free form and in the ternary complex [12] is necessary for the formation of a functional cofactor binding site. In PRDM4, for which no methyltransferase activity has been reported, the corresponding arginine residue (Arg 491) can also block the SAM binding site (PDB entry 3DB5) (Fig 6A). These results strongly suggest that the substitution, in the NH motif, of an asparagine by an arginine led to the loss of SAM binding by MLL5 proteins.

Regarding the catalytic center itself, the conserved Tyr 358 which favors the methyl transfer from the cofactor SAM to the histone peptide in other HMTs such as the MLL1-4 proteins [13,16] appears to be favorably positioned in MLL5 (Fig 6C). In contrast, all others residues important for peptide binding through hydrogen bonds in MLL3 and other MLL proteins are not conserved in MLL5 (Fig 6C). In particular, two tyrosine residues (Tyr 4884 and Tyr 4886 in MLL3), whose hydroxyl group form hydrogen bonds with the side chain of the lysine 4 of the histone peptide in the MLL3 structure complex, are replaced, respectively, by Ile 444 and Phe 446 in MLL5, both of which do not allow the formation of such hydrogen bonds (Fig 6C). In conclusion, the MLL5 SET domain lacks essential residues for the correct binding and positioning of both cofactor and substrate.

Discussion

Human MLL5 has no intrinsic HMT catalytic activity

The data presented here indicate that the isolated and bacterially expressed SET domain of MLL5, either alone or in combination with a complete POSTSET sub-domain, is not capable of binding either the cofactor SAM or the histone substrate and thus completely lacks histone methyltransferase activity. Even if we could not exclude a contribution of post-translational modifications in eukaryotic organisms, this absence of catalytic activity can be explained by several factors. First, the MLL5 protein lacks essential residues for cofactor and substrate binding. Interestingly, comparison of the human MLL5 SET sequence with those of orthologs from various mammalian species indicates that the complete MLL5 SET domain is strictly conserved (not shown). Therefore, the selective pressure on residues within the putative SAM and histone binding sites is comparable to that on structural residues, which suggests that the results obtained here on the human MLL5 protein can likely be applied to all mammalian MLL5 sequences. Second, structural analysis and molecular dynamics simulations of the MLL5 SET domain reveal the presence of a large loop between α1 and β5 that stably folds back to the protein core and prevents histone peptide binding. Previous studies described the reduction or lack of catalytic activity for isolated SET domains due to the incorrect positioning of structural elements around the substrate binding groove [13, 26, 27]. For example, in the isolated MLL1 SET domain, the peptide binding cleft is too wide, resulting in the suboptimal positioning of the substrate and a subsequent loss of catalytic activity [13]. In contrast, the EZH2 SET domain structure reveals a closed conformation that prevents peptide binding [26,27]. The recent crystal structure of the human PRC2 complex [28] allow to show that a re-orientation and stronger stabilization of the EZH2 SET-I in the complex permit the methyltransferase activity of EZH2. To our knowledge, MLL5 represents the first SET domain structure with such an occupation of the peptide binding cleft by a part of the protein. In conclusion, the MLL5 SET domain, in addition to lacking catalytic activity, is further characterized by an absence of H3K4 docking capability. A recent study suggests that, in human MLL5, this latter function is achieved by the PHD domain [29].

Human MLL5 is not a member of the MLL family

Phylogenetic analyses of human, *Drosophila*, yeast and amphioxus sequences demonstrated that MLL5 and SETD5 belong to the same family, and that this family is clearly distinct from the MLL1-4 and SETD1A, B group. Contrary to MLL5, the SET domains of the MLL1-4 proteins have all the required catalytic residues and the methyltransferase activity of these proteins is well described [4,13,16]. In the case of MLL1, the relative positioning of the SET subdomains is not optimal and its association of with the WDR5 (WD-repeat protein-5) complex is required for full activation. Likewise, the association of MLL3 and MLL4 with the RBBP5 (Retinoblastoma-binding protein-5)—ASH2L (Absent Small Homeotic disks-2-like) heterodimer reduces the inherent flexibility of their SET domains, favoring the formation of a catalytically competent conformation [16]. Interaction of these MLL proteins with a tri-subunit complex composed of WDR5-RBBP5-ASH2L involves a WIN (WDR5 interacting) motif in close proximity of the SET domain [30]. While conserved in MLL1-4 as well as in SETD1A, B, MLL5 lacks a WIN interaction motif [31]. In addition, Zhou and colleagues have recently shown that the WDR5, RBBP5 and ASH2L proteins are absent from the MLL5 complex [32]. Comparison of our MLL5 SET structure with that of the MLL3-RBBP5-ASH2L complex confirms that MLL5 does not have the residues required for interacting with this complex [16]. Instead of Val 4809 and Thr 4803 found in MLL3, MLL5 has bulky residues (Gln 367 and Lys

361, respectively) that are unable to accommodate the shallow pocket formed by RBBP5 and the SET domain. Based on our phylogenetic, structural, and functional analyses, it is thus very clear that MLL5 is not a member of the MLL family.

What is the biological role for the human MLL5 SET region?

Although MLL5 lacks HMT activity, the presence of members of the MLL5/SETD5 family in various species (including, humans, cephalochordates, *Drosophila* and yeast) nonetheless suggests that they have important functions in the organism that are independent of the catalytic site. In addition, given the high level of sequence identity between SETD5 and MLL5, it is conceivable that they are functionally redundant, which might explain the relatively mild phenotypes observed in MLL5 knockout mice [5].

Instead of catalyzing histone methylation, recent studies point to a role of MLL5 in the regulation of the expression of other histones-modifying enzymes, like LSD1 or SET7/9. For example, Zhou and colleagues recently showed that MLL5 knockdown cells are characterized by a reduced H3K4 trimethylation rate at E2F1 target promoters, suggesting that MLL5 associates with other proteins (like HCF-1) to indirectly stimulate methylation [32]. Furthermore, human MLL5 as well as its *Drosophila* and yeast orthologs, UpSET and SET3, contain a PHD domain that strongly and specifically binds to histone H3K4me3 through a non-canonical mechanism [7,29,33]. Like human MLL5, *Drosophila* UpSET and yeast SET3 also have no HMT activity, but are found in multi-protein complexes exhibiting deacetylase activity [8,9]. Interestingly, the yeast SET3 complex (termed Set3C) shows similarities to the mammalian HDAC3 complex, involving the corepressors SMRT and N-CoR [8].

In conclusion, we hypothesize that MLL5 functions as a mediator of gene regulation in humans, not through a HMT activity, but as a structural component ensuring multi-protein complex integrity. Thus, MLL5 could have been retained in the genomes of a wide variety of organisms because of its capacity to engage in protein-protein interactions.

Materials and Methods

Phylogenetic analyses

Amino acid sequences included in the analysis are listed in S1 Table. Amino acid alignments were based on conserved protein domains identified using a NCBI CD-Search [34]. All alignments were performed using MUSCLE as implemented in MEGA7 [35], followed by manual refinement. All positions containing gaps and missing data were hence eliminated. In total, two separate phylogenetic analyses were performed, one using only human HMT proteins and one including HMT sequences from humans (*Homo sapiens*), the cephalochordate amphioxus (*Branchiostoma floridae*), the fruit fly (*Drosophila melanogaster*) as well as the yeast (*Saccharomyces cerevisiae*). The first tree was calculated with 24 sequences and a total of 105 amino acid positions, while the second one was obtained based on 22 sequences and a total of 206 amino acid positions. The phylogenetic relationships were assessed using both the Maximum Likelihood (ML) and Bayesian inference (BI) methods. For the ML analyses, two different phylogenetic models were used: the human HMT tree was calculated with the Whelan And Goldman + Freq. model [36] and a discrete Gamma distribution to model evolutionary rate differences among sites (5 categories) and the tree containing several species was computed with the Le and Gascuel model [37] and a discrete Gamma distribution to model evolutionary rate differences among sites (5 categories) with the rate variation model allowing for some sites to be evolutionarily invariable (+I). The robustness of each node of the resulting trees was assessed by Bootstrap analyses (with 1000 pseudoreplicates) [38]. All ML analyses were conducted using MEGA7 [35]. To validate the obtained ML results, trees were also computed using the Bayesian Inference

(BI) method, as implemented in MrBayes version 3.1.2 [39]. BI phylogenies were obtained using the WAG model with a discrete Gamma distribution for amino acid substitutions [36] in two runs of 1,000,000 generations with sampling of trees every 100 generations and a burn-in period of 25%. All trees are drawn to scale, with branch lengths indicating the number of substitutions per site.

Cloning, expression and purification

The human MLL5 cDNA was a gift from Shigeaki Kato (University of Tokyo, Japan). A fragment from this cDNA encoding the SET domain (residues 323–458) was sub-cloned into the pOPINJ vector from Ray Owens's lab [40], allowing the expression as a His_6-GST fusion protein (named SET). In order to limit dimerization of the protein, a mutation of cysteine 453 into an alanine was introduced using the QuikChange site-directed mutagenesis lightning kit (Stratagene) (named SETC453A). This mutant was used only for crystallization purpose. Two longer constructs including the POSTSET domain were prepared by sub-cloning a fragment encoding residues 323–473 (named SET-POSTSET) and residues 262–473 (named Nh-SET-POSTSET). These two latest constructs do not contain the C453A mutation. All the proteins were expressed in Rosetta (DE3) *E. coli* cells. Cultures were grown overnight at 20°C after induction with 0.5mM IPTG. The cell lysate was first applied to a nickel affinity column (HiTrap HP, GE Healthcare) washed in 50mM Tris, HCl pH 7.5, 150mM NaCl and 1mM DTT. After cleavage of the GST-tag with PreScission protease on the column, a final purification step was added consisting in a gel filtration on a Superdex 75 column (GE Healthcare) equilibrated in the same buffer. 20μM of $ZnSO_4$ was added in all purification buffers for the constructs that include the POSTSET domain. Fractions containing the protein of interest were pooled and concentrated to 8 mg/ml. The intact mass of the purified protein was confirmed by ESI mass spectrometry.

Crystallization, data collection and structure determination

The protein SETC453A was crystallized by sitting drop vapor diffusion using a 1:1 drop ratio against a mother liquor containing 0.1 M MES pH 6.5 and 25% (w/v) PEG 3000. Crystals around 100 μm in size, were cryo-cooled by immersion in liquid nitrogen using 25% glycerol as cryoprotectant. Datasets were collected at the ID23-1 beam-line at the European Synchrotron Radiation Facility (Grenoble, France) and processed and scaled with XDS and XSCALE [41]. For molecular replacement, model rebuilding and refinement, a number of programs in CCP4 [24] and PHENIX [42] were used in combination (Phaser [20], Buccaner [22], Parrot [23], COOT [21], Refmac [43] and phenix.refine [44]).

Nuclear Magnetic Resonance (NMR) experiments

The following protein buffer was used for NMR experiments: 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT (addition of 20 μM ZnSO4 for the SET-POSTSET protein). All spectra were recorded on a 500 MHz Bruker Advance spectrometer equipped with a cryogenic H/C/D/N probe with a Z-axis gradient at 300 K. For 1D titration experiments, series of 1D Saturation Transfer Difference (STD) NMR spectra [45] were recorded on different mixtures of isolated MLL5 SET-POSTSET protein and several forms of H3K4 peptide (un-, mono- or di-methylated). The peptide concentration was typically 400 μM and the peptide:protein ratio was systematically varied from 20:1 to 5:1. STD spectra were acquired with 2048 scans and selective saturation of MLL5 SET-POSTSET resonances at -0.5 ppm (-30 ppm for reference spectra). Saturation transfer of 2 s was achieved with equally spaced 50-ms Gaussian shaped pulses (separated by a 1-ms delay). The most shielded resonance of peptide is at 1.11 ppm, but an

irradiation test was also performed on a free peptide sample to verify that only MLL5 SET-POSTSET resonances were irradiated. Subtraction of free induction decay values with on- and off-resonance protein saturation was achieved by phase cycling. A relaxation delay of 2.5 s (Aq + D_1) and 128 dummy scans were employed to reduce subtraction artifacts. All NMR Spectra were processed and analyzed using Gifa [46] and Topspin 2.6 (Bruker Biopsin, Germany).

Isothermal Titration Calorimetry (ITC)

The ITC experiment was carried out at 20°C on a MicroCal VP microcalorimeter. The Nh-SET-POSTSET protein sample was buffer exchanged by overnight dialysis into the ITC buffer (50 mM Na phosphate pH 7.5, 150 mM NaCl). Titration of MLL5 Nh-SET-POSTSET domain (30 μM) in the cell (2 mL) was performed by sequential addition of SAM (600μM, 25 injections of 10 μL). Integrated raw ITC data were fitted to a one binding site model using nonlinear least squares regression analysis with the MicroCal Origin plugin after the control experiments (titration of the SAM from the syringe into the buffer) were subtracted.

Fluorescence-based thermal shift assays (Thermofluor®)

Solutions of 15 μL containing 40 μM of MLL5 SET-POSTSET protein, 2X Sypro® Orange (Sigma) and different concentrations (ranging from 2 to 20 molar excess of protein) of SAM, S-adenosyl homocysteine (SAH) or histone peptide in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT were added to the wells of a 96-well PCR plate. The plates were sealed with an optical sealing tape (Bio-Rad) and heated in an Mx3005P Q-PCR system (Stratagene) from 25°C to 95°C at 1°C intervals. Fluorescence changes in the wells were monitored with a photomultiplier tube. The wavelengths for excitation and emission were 545 nm and 568 nm, respectively.

Fluorescence anisotropy measurements

The binding affinities of fluorescein-labeled H3 peptides (ARTKQTARKST with K4 un- or mono-methylated, purchased at EZBiolab) for MLL5 SET-POSTSET protein in the presence of SAM (at a concentration of 10 molar excess of the higher protein concentration) were measured using a Safire microplate reader (TECAN) with the excitation wavelength set at 470 nm and emission measured at 530 nm at room temperature. The following buffer solution was used for this experiment: 50 mM Tris HCl pH 7.5, 150 mM NaCl, 20 μM $ZnSO_4$, 2 mM β-mercaptoethanol, 10 mM $MgCl_2$, 10% glycerol.

Histone methyltransferase assay

GST-MLL5 Nh-SET-POSTSET (1 μg) protein was incubated with or without HeLa nuclear extracts (200 μg at $5.0X10^9$, Cilbiotech) at 4°C in buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.5% NP-40, 12.5 mM $MgCl_2$, 1 mM β-Mercaptoethanol, 10 μM $ZnSO_4$). 24 hours later, 5 μg of bulk histones from calf thymus (Roche) and 1 μl ^{3}H-SAM (15 Ci/mM, Perkin Elmer) were added for 90 min at 30°C. Methylation assay with SET7 was used as a positive control by mixing 5 μg of histones with purified Set 7 (1 μg) in 20 mM Tris-HCl pH 8.0, 4 mM EDTA, 1 mM PMSF, 0.5 mM DTT, and 1 μl ^{3}H-SAM (15 Ci/mM, Perkin Elmer) for 90 min at 30°C. Methylation was measured by liquid scintillation assay as described [47]. After the incubation period, 10μl of the reaction mixture were spread onto Glass Microfibre filters GF/A (Whatman, GE Healthcare, Life Sciences). Filters were washed with 50 mM $NaHCO_3$ (pH 9.0) and air dried for 2 hours. The amounts of methylated products were quantified by liquid scintillation on a Liquid Scintillation Analyzer Tri-Carb 2900 TR (Beckman Coulter).

Molecular Dynamics (MD) simulations

The initial model used for all-atom MD simulations corresponds to the MLL5 molecule A in the asymmetric unit. The disordered missing loop connecting residues 331 and 341 in the N-terminal segment was modeled using MODELLER [48]. The protein was solvated in TIP3P water molecules [49] and neutralized in 150 mM NaCl salt. The MD simulations of the solvated proteins were performed using NAMD2 [50] with the CHARMM27 force field [51]. The protein was initially minimized and then heated to 310 K with harmonic restraints added during the first 250-ps equilibration steps. A 2-ns unrestrained simulation was sufficient to achieve equilibration. All simulations were performed with periodic boundary conditions using the NPT micro-canonical ensemble with pressure set to 1 atmosphere and temperature set to 310 K. The pressure and temperature were maintained using the Nose-Hoover Langevin piston and the Langevin thermostat respectively. The van der Waals interactions were calculated using a switching distance of 10 Å and a cutoff of 12 Å. The particle mesh Ewald (PME) method [52] was used to calculate long range electrostatic interactions, with a grid spacing of 1 Å. The total number of atoms was approximately 23,300. All the production runs were performed with 2-fs time steps in conjunction with the SHAKE algorithm to constrain covalent bond lengths between heavy and hydrogen atoms [53]. Two independent 100-ns simulations were performed on an Intel Xeon cluster. Trajectory data from the simulations were recorded and analyzed in VMD [54]. Normal modes were calculated with the WEBnm@ program [55] using the default parameters and Cα atoms of the MLL5 protein. WEBnm@ uses the elastic network model and calculates the six lowest-frequency vibrational modes (modes 7–12).

Accession code

The atomic coordinates and structure factors have been deposited in the Protein Data Bank under the accession code 5HT6.

Supporting Information

S1 Fig. Comparison of the active center of MLL1 (PDB code 2W5Z) and MLL3 (PDB code 5F6K) complex structures. MLL3 residues (PDB entry 5F6K, green sticks) that form hydrogen bonds with the cofactor SAM (yellow sticks) and the K4 histone peptide (yellow sticks). Corresponding residues of the MLL1 SET domain are superimposed (PDB entry 2W5Z, pink sticks).
(TIF)

S2 Fig. Phylogenetic analysis of human histone methyltransferase (HMT) proteins. (A). Maximum Likelihood (ML) tree of human HMT proteins, with Bootstrap percentages (ranging from 0 to 100) as branch support indicated at each node. Branch lengths are representative of sequence substitution rates. **(B).** Bayesian Inference (BI) tree of human HMT proteins, with posterior probabilities (ranging from 0 to 1) as branch support indicated at each node. Branch lengths are representative of sequence substitution rates.
(TIF)

S3 Fig. No histone binding by the MLL5 SET domain. GST pull-down shows that MLL5 SET domain does not bind either histone H3 or a mixture of histones from calf thymus.
(TIF)

S4 Fig. Influence of the POSTSET of MLL5 SET on its structure. (A). Model of the MLL5 SET-POSTSET fragment. The MLL5 POSTSET zinc binding site was modeled using the MODELLER v9.14 program (32), and the homologous POSTSET domain of SETD2 methyl transferase (pdb entry 4H12) (50) as template. **(B).** Superposition of the ^{1}H-^{15}N HSQC spectra of

MLL5 SET domain (blue) and MLL5 SET-POSTSET fragment (green) recorded at 500 MHz. The 2D heteronuclear experiments were recorded at 300K, with the two proteins at a concentration of 400 and 100 μM concentration respectively, in 50 mM Tris pH7.5, 150 mM NaCl, 1mM DTT. Peaks labeled with a red asterisk are present only in the spectrum of the SET domain, whereas peaks framed in a box are present only in the spectrum of SET-POSTSET fragment.
(TIF)

S5 Fig. Influence of the extending N-terminal flanking region of MLL5 SET on its structure. Superposition of the ^{1}H-^{15}N HSQC spectra of MLL5 SET-POSTSET domain (green) and Nh-SET-POSTSET (purple) recorded at 500 MHz. The 2D heteronuclear experiments were recorded at 300K, with the two proteins at a concentration of 100 mM concentration, in 50 mM Tris pH7.5, 150 mM NaCl, 1 mM DTT.
(PDF)

S6 Fig. Models from the different steps of rebuilding. Superposition of the final MLL5 SET domain structure (same colors as in Fig 4A) with one of the best model used for molecular replacement (PDB entry 3BO5, colored in brown) and with the intermediate hybrid model (colored in red). The two regions whose modeling was crucial are circled.
(TIF)

S7 Fig. Distance calculations suggest a persistent salt-bridge Arg 365-Glu 392 during MLL5 MD simulation. Plot of the distance between Arg 365 and Glu 392 (in red) and, as comparison, of the distance between Lys 376 and Asp 449. Asp 449 is localized in the POSTSET flexible segment. These values correspond to the distance between the center of mass of the oxygens in the acidic side chain and center of mass of the nitrogens in the basic side chain. The cutoff distance between the oxygen atoms of acidic residues and the nitrogen atoms of basic residues is 3.2A.
(TIF)

S1 File. Comments on S4 and S5 Figs and supplementary methods.
(PDF)

S1 Table. Amino acid sequences included in the phylogenetic analysis.
(PDF)

Acknowledgments

We acknowledge the experimental assistance during data collection by the staff of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). We acknowledge the Spectroscopy and Calorimetry Facility (LEC) from the LNBIO/CNPEM (Campinas, Brasil) for ITC measurements. The authors would further like to thank Nicolas Robert for technical help with the phylogenetic analyses.

Author Contributions

Conceptualization: AlM WB.

Funding acquisition: AlM.

Investigation: SM-y-M AL LB MB HD JF CT CD JEC MS AlM.

Supervision: AlM.

Writing – original draft: AlM.

Writing – review & editing: SM-y-M AL LB MB HD JF CT CD JEC MS WB AlM.

References

1. Emerling BM, Bonifas J, Kratz CP, Donovan S, Taylor BR, Green ED, et al. MLL5, a homolog of Drosophila trithorax located within a segment of chromosome band 7q22 implicated in myeloid leukemia. Oncogene. 2002; 21, 4849–4854. doi: 10.1038/sj.onc.1205615 PMID: 12101424
2. Madan V, Madan B, Brykczynska U, Zilbermann F, Hogeveen K, Döhner K, et al. Impaired function of primitive hematopoietic cells in mice lacking the Mixed-Lineage-Leukemia homolog MLL5. Blood. 2009; 113, 1444–1454. doi: 10.1182/blood-2008-02-142638 PMID: 18952892
3. Heuser M, Yap DB, Leung M, de Algara TR, Tafech A, McKinney S, et al. Loss of MII5 results in pleiotropic hematopoietic defects, reduced neutrophil immune function, and extreme sensitivity to DNA demethylation. Blood. 2009; 113, 1432–1443. doi: 10.1182/blood-2008-06-162263 PMID: 18854576
4. Zhang Y, Wong J, Klinger M, Tran MT, Shannon KM, Killeen N. MII5 contributes to hematopoietic stem cell fitness and homeostasis. Blood. 2009; 113, 1455–1463. doi: 10.1182/blood-2008-05-159905 PMID: 18818388
5. Yap DB, Walker DC, Prentice LM, McKinney S, Turashvili G, Mooslehner-Allen K, et al. Mll5 is required for normal spermatogenesis. PLoS One. 2011; 6(11):e27127. doi: 10.1371/journal.pone.0027127 PMID: 22069496
6. Deng LW, Chiu I, Strominger JL. MLL 5 protein forms intranuclear foci, and overexpression inhibits cell cycle progression. Proc. Natl. Acad. Sci. U. S. A. 2004; 101, 757–762. doi: 10.1073/pnas.2036345100 PMID: 14718661
7. Ali M, Rincon-Arano H, Zhao W, Rothbart SB, Tong Q, Parkhurst SM, et al. Molecular basis for chromatin binding and regulation of MLL5. Proc. Natl. Acad. Sci. U. S. A. 2013; 110, 11296–301. doi: 10.1073/pnas.1310156110 PMID: 23798402
8. Pijnappel WW, Schaft D, Roguev A, Shevchenko A, Tekotte H, Wilm M, et al. The S. cerevisiae SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program. Genes Dev. 2001; 15, 2991–3004. doi: 10.1101/gad.207401 PMID: 11711434
9. Rincon-Arano H, Halow J, Delrow JJ, Parkhurst SM, Groudine M. UpSET recruits HDAC complexes and restricts chromatin accessibility and acetylation at promoter regions. Cell. 2001; 151, 1214–1228.
10. Sebastian S, Sreenivas P, Sambasivan R, Cheedipudi S, Kandalla P, Pavlath GK, et al. MLL5, a trithorax homolog, indirectly regulates H3K4 methylation, represses cyclin A2 expression, and promotes myogenic differentiation. Proc. Natl. Acad. Sci. U. S. A. 2009; 106, 4719–24. doi: 10.1073/pnas.0807136106 PMID: 19264965
11. Eissenberg JC & Shilatifard A. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Developmental Biology. 2010; 339, 240–249. doi: 10.1016/j.ydbio.2009.08.017 PMID: 19703438
12. Xiao B, Jing C, Wilson JR, Walker PA, Vasisht N, Kelly G, et al. Structure and catalytic mechanism of the human histone methyltransferase SET7/9. Nature. 2003; 421, 652–656. doi: 10.1038/nature01378 PMID: 12540855
13. Southall SM, Wong PS, Odho Z, Roe SM, Wilson JR. Structural Basis for the Requirement of Additional Factors for MLL1 SET Domain Activity and Recognition of Epigenetic Marks. Mol. Cell. 2009; 33, 181–191. doi: 10.1016/j.molcel.2008.12.029 PMID: 19187761
14. Ruthenburg AJ, Allis CD, Wysocka J. Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. Molecular Cell; 2007; 25, 15–30. doi: 10.1016/j.molcel.2006.12.014 PMID: 17218268
15. Shilatifard A. Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. Current Opinion in Cell Biology. 2008; 20, 341–348. doi: 10.1016/j.ceb.2008.03.019 PMID: 18508253
16. Li Y, Han J, Zhang Y, Cao F, Liu Z, Li S, et al. Structural basis for activity regulation of MLL family methyltransferases. Nature. 2016; 530, 447–452. doi: 10.1038/nature16952 PMID: 26886794
17. Palmer AG, Kroenke CD, Loria JP. NMR methods for quantifying microsecond-to-millisecond motions in biological macromolecules. Methods Enzym. 2001; 339, 204–238.
18. Horowitz S, Yesselman JD, Al-Hashimi HM, Trievel RC. Direct evidence for methyl group coordination by carbon-oxygen hydrogen bonds in the lysine methyltransferase SET7/9. J. Biol. Chem. 2011; 286, 18658–18663. doi: 10.1074/jbc.M111.232876 PMID: 21454678
19. Wu H, Zeng H, Dong A, Li F, He H, Senisterra G, et al. Structure of the catalytic domain of EZH2 reveals conformational plasticity in cofactor and substrate binding sites and explains oncogenic mutations. PLoS One. 2013; Dec 19; 8(12):e83737. doi: 10.1371/journal.pone.0083737 PMID: 24367611

20. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J. Appl. Crystallogr. 2007; 40, 658–674. doi: 10.1107/S0021889807021206 PMID: 19461840

21. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. 2010; 66, 486–501.

22. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. Sect. D Biol. Crystallogr. 2006; 62, 1002–1011.

23. Cowtan K. Recent developments in classical density modification. Acta Crystallogr. Sect. D Biol. Crystallogr. 2010; 66, 470–478.

24. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallographica Section D: Biological Crystallography. 2011; 67, 235–242.

25. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr. Protoc. Bioinformatics. 2014; 47, 5.6.1–5.6.32.

26. Zheng W, Ibanez G, Wu H, Blum G, Zeng H, Dong A, et al. Sinefungin derivatives as inhibitors and structure probes of protein lysine methyltransferase SETD2. J. Am. Chem. Soc. 2012; 134, 18004–18014. doi: 10.1021/ja307060p PMID: 23043551

27. Antonysamy S, Condon B, Druzina Z, Bonanno JB, Gheyi T, Zhang F, et al. Structural context of disease-associated mutations and putative mechanism of autoinhibition revealed by X-Ray crystallographic analysis of the EZH2-SET domain. PLoS One. 2013: 8(12):e84147. doi: 10.1371/journal.pone.0084147 PMID: 24367637

28. Justin N, Zhang Y, Tarricone C, Martin SR, Chen S, Underwood E, et al. Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. Nat Commun. 2016; Apr 28; 7:11316. doi: 10.1038/ncomms11316 PMID: 27121947

29. Lemak A, Yee A, Wu H, Yap D, Zeng H, Dombrovski L, et al. Solution NMR Structure and Histone Binding of the PHD Domain of Human MLL5. PLoS One. 2013; 8(10):e77020. doi: 10.1371/journal.pone.0077020 PMID: 24130829

30. Zhang P, Lee H, Brunzelle JS, Couture JF. The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. Nucleic Acids Res. 2012; 40, 4237–4246. doi: 10.1093/nar/gkr1235 PMID: 22266653

31. Patel A, Dharmarajan V, Cosgrove MS. Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. J. Biol. Chem. 2008; 283, 32158–32161. doi: 10.1074/jbc.C800164200 PMID: 18829459

32. Zhou P, Wang Z, Yuan X, Zhou C, Liu L, Wan X, et al. Mixed Lineage Leukemia 5 (MLL5) protein regulates cell cycle progression and E2F1-responsive gene expression via association with Host Cell Factor-1 (HCF-1). J. Biol. Chem. 2013; 288, 17532–17543. doi: 10.1074/jbc.M112.439729 PMID: 23629655

33. Kim T, Buratowski S. Dimethylation of H3K4 by Set1 Recruits the Set3 Histone Deacetylase Complex to 5' Transcribed Regions. Cell. 2009; 137, 259–272. doi: 10.1016/j.cell.2009.02.045 PMID: 19379692

34. Marchler-Bauer A, Bryant SH. CD-Search: Protein domain annotations on the fly. Nucleic Acids Res. 2004; 32:W327–31. doi: 10.1093/nar/gkh454 PMID: 15215404

35. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 2016; msw054.

36. Whelan S, Liò P, Goldman N. Molecular phylogenetics: State-of-the-art methods for looking into the past. Trends in Genetics. 2001; 17, 262–272. PMID: 11335036

37. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol. Biol. Evol. 2008; 25, 1307–1320. doi: 10.1093/molbev/msn067 PMID: 18367465

38. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution (N. Y). 1985; 39, 783–791.

39. Huelsenbeck JP, Ronquist F. MrBayes: Bayesian inference of phylogeny. Bioinformatics. 2001; 17, 754–5. PMID: 11524383

40. Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Res. 2007; 35(6):e45. doi: 10.1093/nar/gkm047 PMID: 17317681

41. Kabsch W. (2010) XDS. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 125–132.

42. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr. 2010; 66, 213–221.

43. Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN. REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr. Sect. D Biol. Crystallogr. 2004; 60, 2184–2195.

44. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. Sect. D Biol. Crystallogr. 2012; 68, 352–367.

45. Mayer M, Meyer B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angew. Chemie Int. Ed. 1999; 38, 1784–1788.

46. Pons JL, Malliavin TE, Delsuc MA. Gifa V. 4: A complete package for NMR data set processing. J. Biomol. NMR. 1996; 8, 445–452. doi: 10.1007/BF00228146 PMID: 20859778

47. Wu J, Xie N, Feng Y, Zheng YG. Scintillation Proximity Assay of Arginine Methylation. J. Biomol. Screen. 2012; 17, 237–244. doi: 10.1177/1087057111414903 PMID: 21821785

48. Sali A. MODELLER A Program for Protein Structure Modeling Release 9v4, r6262. Structure. 2008; 779–815.

49. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. J. Phys. Chem. A. 2001; 105, 9954–9960.

50. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. Journal of Computational Chemistry. 2005; 26, 1781–1802. doi: 10.1002/jcc.20289 PMID: 16222654

51. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B. 1998; 102, 3586–616. doi: 10.1021/jp973084f PMID: 24889800

52. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 1993; 98, 10089.

53. Kräutler V, Van Gunsteren WF, Hünenberger PH. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. J. Comput. Chem. 2001; 22, 501–508.

54. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J. Mol. Graph. 1996; 14, 33–38. PMID: 8744570

55. Tiwari SP, Fuglebakk E, Hollup SM, Skjaerven L, Cragnolini T, Grindhaug SH, et al. WEBnm@ v2.0: Web server and services for comparing protein flexibility. BMC Bioinformatics. 2014; 15, 427. doi: 10.1186/s12859-014-0427-6 PMID: 25547242

56. Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010; Jul; 38 (Web Server issue):W545–9. doi: 10.1093/nar/gkq366 PMID: 20457744

Appendix 9

Complete CV

CARVALHO, João Emanuel Marques
Date of birth: 12/11/1988 (28 years-old)
Nationality: Portuguese

Scientific Background

02/2013 – Current.	Ph.D. Fellow. Advisor: Dr. Michael Schubert, Laboratoire de Biologie du Développement de Villefranche-sur-Mer (LBDV), France. EvoInSiDe team. Funding: FCT, (Portuguese Foundation for Science and Technology).
09/2011 – 08/2012	Graduate researcher. Institut de Génomique Fonctionnelle de Lyon (IGFL), France. Molecular Zoology group.
09/2010 – 02/2011	Research intern. Max F. Perutz Laboratories, Computational and Structural Biology, Austria. Raible Lab.
09/2009 – 10/2009	Undergraduate researcher. University of Lodz, Poland. Department of Ecophysiology and Plant Development.
09/2008 – 06/2009	Undergraduate researcher. Instituto Gulbenkian de Ciência (IGC), Portugal. Plant Development Group.

Education

02/2013 – Current.	Ph.D. student. University of Pierre and Marie Curie (Paris 6), France. Ecole Doctorale Complexité du Vivant.
11/2009 – 07/2012	Master degree in Evolutionary and Developmental Biology. University of Lisbon, Portugal. Faculty of Science.
09/2011 – 07/2012	Erasmus placement (Master thesis research project). Ecole Normale Supérieure de Lyon, France.
09/2010 – 02/2011	Erasmus exchange (Genetics and Developmental Biology program). University of Vienna, Austria. Life Sciences Faculty
09/2006 – 07/2009	Bachelor degree in Biology. University of Lisbon, Portugal. Faculty of Science.

Conference presentations

06/2016	Talk: Retinoic acid signaling targets during amphioxus development: a transcriptomic and genomic perspective. Chromatin Meets South 2016, Nice, France.
09/2015	Talk: Big Data dans un contexte biologique: défis et perspectives. II Journée des Jeunes Chercheurs de l'OOV, Villefranche-sur-Mer, France.
04/2015	Talk and poster: Retinoic acid degradation in the amphioxus embryo: developmental functions and evolutionary implications. VI Journée André Picard, Villefranche-sur-Mer, France.
07/2014	Poster: Functional consequences of lineage-specific duplications: the example of retinoic acid degradation mechanisms in developing amphioxus. V Euro Evo-Devo meeting, Vienna, Austria.
07/2012	Poster: The evolutionary history of retinoid signaling: insights from endogenous retinoic acid degradation. IV Euro Evo-Devo meeting, Lisboa, Portugal.

Other activities and honors

09/2015	Organizer of interdisciplinary meeting. II Journée des Jeunes Chercheurs de l'OOV, Villefranche-sur-Mer, France.
09/2014	Organizer of interdisciplinary meeting. I Journée des Jeunes Chercheurs de l'OOV, Villefranche-sur-Mer, France.
02/2014	Grant award. Travel grant from the Société Française de Biologie du Développement (SFBD).
2013/2014	Volunteer work. Vice-president of the association Erasmus Student Network (ESN) Nice.
10/2012	Grant award. Ph.D. fellowship from FCT (Portuguese Foundation for Science and Technology).
09/2008	Grant award. Research fellowship from University of Lisbon and Amadeu Dias Foundation.

Publication list

1. Carvalho JE, Gancedo SJ, Marlétaz F, Lahaye F, Escriva H, Croce JC, Skarmeta JL, Schubert M. RNAseq and ATACseq data reveals new targets and mechanisms of action of the RA-signaling pathway in amphioxus development. In prep.

2. Carvalho JE, Croce JC, Schubert M. Bringing amphioxus to the bench-side. In prep.

3. Lahaye F, Carvalho JE, Croce JC, Schubert M. CYP26 function is involved in RA-degradation dependent patterning during the amphioxus development. In prep.

4. Carvalho JE, Lahaye F, Croce JC, Schubert M. Cephalochordates development: an integrative view. In prep.

5. Carvalho JE, Theodosiou M, Chen J, Chevret P, Alvarez S, de Lera AR, Laudet V, Croce JC, Schubert M. Lineage-specific duplication of amphioxus retinoic acid degrading enzymes (CYP26) resulted in sub-functionalization of patterning and homeostatic roles. BMC Evol Biol. Under revision.

6. Mas-y-Masa S, Barbona M, Teyssierd C, Déménéa H, Carvalho JE, Lebedev A, Fattorig J, Schubert M, Dumasa C, Bourgueta W, le Maire A. The human Mixed Lineage Leukemia 5 (MLL5), a sequentially and structurally divergent SET domain-containing protein with no intrinsic catalytic activity. PLoS One. 2016;11(11): e0165139.

7. Hirsinger E, Carvalho JE, Chevalier C, Lutfalla G, Nicolas J-F, Peyriéras N, Schubert M. Expression of fluorescent proteins in Branchiostoma lanceolatum by mRNA injection into unfertilized oocytes. J Vis Exp. 2015;(95).

8. Koop D, Chen J, Theodosiou M, Carvalho JE, Alvarez S, Lera AR de, Holland LZ, Schubert M. Roles of retinoic acid and Tbx1/10 in pharyngeal segmentation: amphioxus and the ancestral chordate condition. EvoDevo. 2014;5(1):36.

9. Holland LZ, Carvalho JE, Escriva H, Laudet V, Schubert M, Shimeld SM, Yu J-K. Evolution of bilaterian central nervous systems: a single origin? EvoDevo. 2013;4(1):27.

10. Carvalho JE, Schubert M. Retinoic acid: metabolism, developmental functions, and evolution. In: Dakshinamurti K, Dakshinamurti S, eds. Vitam-Bind Proteins Funct Consequences. CRC Press; 2013. 1–30.

11. Michard E, Lima PT, Borges F, Silva AC, Portes MT, Carvalho JE, Gilliham M, Liu L-H, Obermeyer G, Feijó JA. Glutamate receptor-like genes form Ca2+ channels in pollen tubes and are regulated by pistil D-serine. Science. 2011;332(6028):434–437.

Retinoic acid signaling in chordates:

The evolutionary history of a morphogen-dependent signaling

One of the most striking features of multicellular animals, the metazoans, is their amazing morphological diversity. Even though phylogenetic research has made remarkable progress towards revealing how the abundance and variety of animal life forms relates on the molecular, cellular, tissue, and organismal level, the alteration of developmental programs has been revealed as a key aspect in this process. During development, a rather limited number of signaling pathways has been shown to be instrumental for generating metazoan diversity. The retinoic acid (RA) signaling pathway is one of these instrumental signaling cascades.

A significant amount of time and work has been used to characterize the functions and roles of RA signaling during development, although further work is required to better understand the evolutionary history of the RA signaling network, from metabolism to signal transduction passing by the interactions with other signaling cascades and its developmental functions and how they evolve with time. In this context, model organisms with representative, vertebrate-like RA signaling cascades, such as the cephalochordate amphioxus, will be an important case-study in order to identify the blueprint of an ancestral RA network.

The amphioxus RA signaling pathway was initially studied about 20 years ago, even though not much is known about the bioavailability of RA during development. Moreover, the target genes of the RA signaling pathway and their hierarchical relationship during amphioxus development represent an interesting open question. Therefore, this work aimed at providing a detailed description of two fundamental aspects of the RA signaling pathway during amphioxus development: (1) the regulation of the bioavailability of RA in the developing embryo and (2) the target genes under the control of the RA signaling pathway together with their hierarchical regulatory relationship. To address these questions, the European amphioxus, Branchiostoma lanceolatum, was used as a model system.

During my research project, not only these questions were fundamental, but also the implementation of amphioxus as a reliable model system and thus the establishment of multiple aquaculture improvements as well as in vivo techniques, such as the microinjection of mRNAs into amphioxus eggs. Furthermore, to characterize the bioavailability of RA during development of amphioxus, I focused on the study of the enzymes that mediate the catabolism of RA endogenously, the CYP26 subfamily proteins. I thus described the evolutionary diversification of CYP26 genes in deuterostomes as well as their expression, their function and the mechanisms that govern the feedback loop controlled directly by RA during amphioxus development.

Additionally, to shed light on the target genes under the control of the RA signaling pathway during amphioxus development, I combined pharmacological treatments using retinoid-specific drugs with two different techniques of high throughput sequencing: RNAseq, that revealed the entire RNA profile and thus the genes being expressed at a given moment in time, and ATACseq (assay for transposase-accessible chromatin) that provided a global overview of accessible regions of the chromatin (i.e. open chromatin regions). By combining the data obtained by these techniques, I revealed a new set of genes that are under the control of the RA signaling pathway as well as new potential regulatory loops driving RAR-mediated expression. Moreover, I established a framework to characterize gene hierarchies during development that can be widely applied to other signaling pathways and organisms.