HAL
open science

Analyse fonctionnelle d'éléments cis-régulateurs lors du développement des Vertébrés

Elodie Thierion

To cite this version:

Elodie Thierion. Analyse fonctionnelle d'éléments cis-régulateurs lors du développement des Vertébrés. Biologie du développement. Université Pierre et Marie Curie - Paris VI, 2016. Français. NNT : 2016PA066347 . tel-01496865

HAL Id: tel-01496865

https://theses.hal.science/tel-01496865v1

Submitted on 28 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de Doctorat de l'Université Pierre et Marie Curie – Paris VI

École Doctorale Complexité du Vivant CDV (ED 515)

Présentée par

Élodie Thierion

Pour obtenir le grade de Docteur de l'Université Pierre et Marie Curie

---

Analyse fonctionnelle d'éléments *cis*-régulateurs lors du développement des Vertébrés

---

Soutenue publiquement le 26 septembre 2016
Devant un jury composé de :

Dr. Sylvie Schneider-Maunoury	Présidente
Dr. Olivier Cuvier	Rapporteur
Dr. Cristina Pujades	Rapporteur
Dr. François Spitz	Examinateur
Dr. Pascal Maire	Examinateur
Dr. Pascale Gilardi-Hebenstreit	Directrice de thèse
Pr. Patrick Charnay	Membre invité

Thèse de Doctorat de l'Université Pierre et Marie Curie – Paris VI

École Doctorale Complexité du Vivant CDV (ED 515)

Présentée par

Élodie Thierion

Pour obtenir le grade de Docteur de l'Université Pierre et Marie Curie

---

Analyse fonctionnelle d'éléments *cis*-régulateurs lors du développement des Vertébrés

---

Soutenue publiquement le 26 septembre 2016
Devant un jury composé de :

Dr. Sylvie Schneider-Maunoury	Présidente
Dr. Olivier Cuvier	Rapporteur
Dr. Cristina Pujades	Rapporteur
Dr. François Spitz	Examinateur
Dr. Pascal Maire	Examinateur
Dr. Pascale Gilardi-Hebenstreit	Directrice de thèse
Pr. Patrick Charnay	Membre invité

Institut de Biologie de l'École Normale Supérieure – INSERM U1024 – CNRS UMR 8197
46 rue d'Ulm, 75005 Paris

À ma grande tante Monique et mon grand-père Pierre,
tous deux passionnés de biologie.

Résumé

Les éléments *cis*-régulateurs contrôlent l'expression spatio-temporelle des gènes au cours du développement. La segmentation du rhombencéphale en rhombomères (r) établit l'organisation du cerveau postérieur des vertébrés. La spécification et la formation des segments r3 et r5 sont gouvernées par l'expression du facteur de transcription Krox20/Egr2, sous le contrôle de plusieurs éléments *cis*-régulateurs. Les éléments B (dans r5) et C (dans r3 et r5) sont actifs au cours du démarrage de l'expression de *Krox20* alors que l'élément A (dans r3 et r5) est un élément autorégulateur. En générant la délétion de l'élément A dans le génome murin, nous avons montré qu'il est nécessaire au maintien de l'expression de *Krox20* dans r3 et r5 et qu'en son absence les cellules ne sont pas spécifiées correctement et l'architecture neuronale est perturbée. De façon inattendue, la délétion de l'élément C n'impacte pas la phase de démarrage mais abolit le maintien de l'expression de *Krox20* dans r3 de manière analogue à celle de l'élément A. Un autre élément permet donc le démarrage de l'expression de *Krox20* dans r3, peut-être de façon redondante avec l'élément C. J'ai mis en évidence par différentes approches génétiques que l'élément C coopère en *cis* avec l'élément A, lors de la phase d'autorégulation. Plusieurs approches de conformation et d'accessibilité de la chromatine au locus *Krox20* indiquent que l'élément A interagit de façon constitutive avec le promoteur et que l'élément C est capable d'affecter l'accessibilité de l'élément A à distance. Parallèlement, j'ai montré que la délétion de l'élément C dans le génome du poisson-zèbre n'entraîne pas de modification de l'expression du gène *Krox20* contrairement à la souris malgré un patron d'expression similaire du gène. Ces approches complémentaires ont permis de caractériser le rôle fonctionnel de plusieurs éléments *cis*-régulateurs contrôlant l'expression d'un gène clef du développement des Vertébrés.

Abstract

Cis-regulatory elements ensure the specifity and the maintenance of the spatio-temporal gene expression pattern throughout the development of an organism. Segmentation of the hindbrain into rhombomeres (r) establishes the posterior brain organisation in vertebrates. The specification and formation of segments r3 and r5 are governed by the expression of the transcription factor Krox20/Egr2, controlled by three *cis*-regulatory elements. Elements B (in r5) and C (in r3 and r5) are active during the initiation of *Krox20* expression while element A (in r3 and r5) is an autoregulatory element. By studying the deletion of element A in the mouse genome, we have shown that *Krox20* expression is not maintained in absence of this element and that cells are improperly specified and neuronal architecture is disturbed, emphasizing the importance of the autoregulatory phase in the hindbrain patterning. Unexpectedly, the deletion of element C does not impair *Krox20* initiation but abolishes the maintenance of expression in r3, similarly to element A knock-out. This suggests that another element allows the initiation of *Krox20* expression in r3 and could be redundant with element C. I also showed by different genetic approaches that element C cooperates in *cis* with element A. Moreover, investigation of chromatin conformation and accessibility in the *Krox20* locus indicate that element A and the promoter are engaged in a constitutive interaction and that element C is able to affect the accessibility of element A at a distance. In parallel, the deletion of element C in zebrafish did not reveal any modification in *Krox20* expression, unlike in mouse and despite a similar expression pattern. These complementary approaches lead to the characterization of the functional role of several *cis*-regulatory elements driving the expression of a key developmental gene in vertebrates.

Remerciements

Je tiens tout d'abord à remercier les membres du jury : Sylvie Schneider-Maunoury, Olivier Cuvier, Cristina Pujades, François Spitz et Pascal Maire de m'avoir fait l'honneur de juger ce travail.

Je suis particulièrement reconnaissante envers Patrick Charnay qui m'a accueillie dans son laboratoire. J'ai apprécié sa disponibilité, ses précieux conseils et l'humour qu'il a distillé au travers de nombreuses pauses café. Je suis heureuse d'avoir pu répondre, avant même la fin de cette thèse, à sa question lors de mon entretien de M2 : l'avenir a tranché, je ferai donc un post-doc !

Je remercie chaleureusement Pascale, qui m'a encadrée et soutenue au cours de cette thèse, en me transmettant son expertise aussi bien de la paillasse que des raisonnements scientifiques. Elle m'a accompagnée au quotidien et a su gérer, avec brio, ma tendance au pessimisme tout au long de ce travail ! Je tiens à remercier ensuite celui qui fut en quelque sorte un doctorant « mentor », Yassine. Il m'a fourni le kit de survie du laboratoire, des expériences à la communication au sein de l'institut en passant par une grande rigueur scientifique et un sens des commérages. Un grand merci à Johan qui fut mon co-box, pour m'avoir évité quelques bourdes, pour son mode de fonctionnement particulier mais rafraichissant et nos discussions du soir. Parmi ceux ayant quitté l'équipe trop vite, je citerai Patricia et Charlotte, deux post-doctorantes pleines de conseils, et je suis heureuse de retrouver cette dernière à Cambridge. Comment ne pas mentionner Aurélie G., confidente et véritable coach de vie, elle m'a aidée à relativiser bien des fois, je la remercie de son soutien si précieux. Une dédicace spéciale à Alexandre, mon prédécesseur doctorant, qui en plus de ses sens artistiques et décoratifs certains nous a gratifiés de ses nombreuses blagues et jeux de mots plus ou moins réussis ! Un grand merci à présent à Patrick T., cette fois successeur d'une partie de mon travail, pour son non-conformisme qui m'a dans un premier temps déconcertée puis amusée et enfin ravie. Je remercie aussi Gaspard, qui fait toujours preuve de beaucoup d'autodérision, d'avoir si gentiment décoré mon bureau pendant mon absence !
Je remercie les autres membres de l'équipe : Piotr pour ses ingénieux outils de dissection, Carole pour son aide pour les injections, Fanny pour ses nombreux conseils concernant le séquençage, Renata pour son fort caractère et sa rigueur scientifique, Kasia qui a toujours un mot pour les autres et Graziella pour les petits services du quotidien.

Comment ne pas citer à présent nos chers voisins du labo Rosa ! Je commencerais cette liste par mon alter ego doctorante, Florence. Nous avons parcouru le chemin semé d'embûches de la thèse, au même rythme. Je la remercie pour nos nombreuses discussions aussi bien scientifiques que personnelles, pour sa grande ouverture d'esprit et bien sûr son soutien. Je remercie ensuite Aline, scientifique incroyable, qui m'a beaucoup aidée dans les derniers mois de thèse, Nicolas, extrêmement doué mais un brin râleur, Sylvain pour sa folie douce, Julien toujours de bon poil, Marina pour sa gentillesse, et tous les autres : Marika, Guillaume L., Fred et Brigitte.

Une petite pensée aussi pour les « Ciao » de Francesco et le sourire de Malik, tous deux anciennement à l'étage.

Je voudrais saluer les gestionnaires de l'équipe, d'abord Solange puis Rouben, qui dans des styles très différents ont facilité bon nombre de mes démarches.
Et un petit mot pour Aurélie J. qui nous a toujours reçus avec gentillesse dans la laverie, haut lieu des pauses café.

On monte d'un étage pour retrouver les bio-informaticiens. Amélie la première, ma coloc très complice, merci pour toutes nos discussions et fous rires. Une petite dédicace à Yves pour le « pôt » du vendredi soir et son coaching pour les interviews de post-doc. Je remercie ensuite mes collaborateurs : Morgane pour ses qualités pédagogiques, Céline pour sa gentillesse et Samuel pour sa redoutable efficacité.

Ayant passé beaucoup de temps à embêter le quatrième étage je me dois d'en remercier les fiers représentants. À commencer par Martin, qui est devenu un véritable ami et m'a prouvé que les plantes, ça pouvait être chouette ! Un grand merci à François pour ses nombreux conseils expérimentaux et son franc parler. Je remercie aussi Heni pour son enthousiasme ainsi que Yann, Erwann et Imen.

Je remercie nos amis de la section développement du 7ème étage, même si mon sujet de thèse m'a de plus en plus éloignée d'eux. Une pensée pour Guillaume O. et Florencia, les doctorants de ma génération.

J'adresse mes remerciements à tous les membres des plateformes d'animalerie : Lucia, Christophe, Dolores, Amandine, Éléonore, Guillaume, Firas et particulièrement Abder et Deborah qui se sont occupés de mes souris pendant 4 ans.

Je tiens à remercier Stéphane Nédélec qui a m'a accueillie à Evry pour m'apprendre les techniques de culture cellulaire.
Je suis très reconnaissante envers Daan Noordermeer qui m'a hébergeé dans son équipe à Gif-sur-Yvette quelques semaines pour m'enseigner la technique de 4C. Je le remercie pour le temps qu'il a et va encore consacrer à cette collaboration, sa patience et sa réactivité incroyables.

Enfin, cette thèse n'aurait pas été possible sans ma famille et mes amis. Je pense notamment à ma grand-mère qui a toujours cru en moi, à mes parents qui m'ont soutenue, ainsi que ma sœur et ma tante Isa qui ont toujours suivi mon parcours avec intérêt. Un grand merci à Mathieu qui a été à mes côtés lors de ces derniers mois en faisant preuve d'une grande curiosité pour le monde étrange de la recherche. Une dédicace très spéciale pour Élise, l'autre coloc du début de thèse, elle a été là de bout en bout avec une joie de vivre inébranlable et des gaffes inoubliables, merci. Une grande pensée pour tous les autres qui n'ont pas démérité : Cécile, Mathilde, Leslie, Élodie, Rébécca, Julia, Amaury, Virgile.

SOMMAIRE

INTRODUCTION **11**

1. LES GENES EUCARYOTES ET LA TRANSCRIPTION **12**
1.1 QU'EST-CE QU'UN GENE ? 14
1.2 COMMENT S'ORGANISENT LES GENES ? 14
1.2.1 Le promoteur *16*
1.2.2 Les enhancers et silencers *17*
1.2.3 Les insulateurs *17*
1.2.4 Les LCR et/ou super-enhancers *18*
1.3 LES ACTEURS DE LA TRANSCRIPTION 18
1.3.1 Les facteurs généraux de la transcription *18*
1.3.2 Les activateurs en trans ou facteurs de transcription *19*
1.3.3 Les co-activateurs *21*
1.3.4 Les répresseurs et co-répresseurs *21*

2. LA REGULATION DE L'EXPRESSION GENETIQUE CHEZ LES EUCARYOTES **23**
2.1 LA REGULATION EPIGENETIQUE 23
2.1.1 La méthylation de l'ADN et les îlots CpGs *23*
2.1.2 Le remodelage de la chromatine et les facteurs de transcription pionniers *25*
2.1.3 Les modifications d'histones et l'activation séquentielle des enhancers *27*
2.1.4 Les ARNs longs non-codants *29*
2.1.5 Le consortium ENCODE *31*
2.2 LES INTERACTIONS ENHANCERS-PROMOTEURS 31
2.2.1 Les protéines de structures modèlent l'architecture des boucles de chromatine ... *31*
2.2.2 La transcription des enhancers *34*
2.2.3 Les polymérases en pause *36*
2.2.4 La transcription bidirectionnelle des promoteurs *37*
2.3 L'ORGANISATION TRIDIMENSIONNELLE DU GENOME 38
2.3.1 Les domaines topologiques (TADs) *38*
2.3.2 Les domaines A&B et les usines de transcription *41*
2.3.3 Les territoires chromosomiques et les LADs *42*

3. LES ELEMENTS *CIS*-REGULATEURS AU CŒUR DES PROCESSUS DEVELOPPEMENTAUX, EVOLUTIFS ET DES MALADIES **47**
3.1 LES ELEMENTS *CIS*-REGULATEURS CONFERENT DE LA ROBUSTESSE PENDANT LE DEVELOPPEMENT 47
3.1.1 Redondance et précision transcriptionnelle *47*
3.1.2 Les sites de faible affinité aussi importants que les sites de forte affinité *48*
3.2 LES ELEMENTS *CIS*-REGULATEURS MOTEUR DE L'EVOLUTION 49
3.2.1 Évolution des enhancers et diversité animale *49*
3.2.2 Conservation d'expression sans conservation de séquences et vice versa *52*
3.3 LES VARIATIONS DES SEQUENCES NON-CODANTES SONT A L'ORIGINE DE MALADIES 53
3.3.1 Les petites variations génétiques modifient la fixation des TFs sur les éléments cis-régulateurs *53*
3.3.2 Les grandes variations structurales altèrent les domaines chromatiniens *55*

4. LA SEGMENTATION DU RHOMBENCEPHALE ET LE GENE *KROX20* 57
4.1 LA MISE EN PLACE DU RHOMBENCEPHALE 57
4.2 LA FONCTION DE *KROX20* DANS LE RHOMBENCEPHALE DEVOILEE PAR L'ETUDE DU MUTANT *KROX20-LACZ* 59
4.3 LES ELEMENTS *CIS*-REGULATEURS DU GENE *KROX20* DANS LE RHOMBENCEPHALE 61
4.4 UN MODELE TRANSCRIPTIONNEL INCOMPLET 63

RESULTATS 65

CHAPITRE 1 67
L'AUTOREGULATION DU GENE *KROX20* DANS LA SEGMENTATION DU RHOMBENCEPHALE 67
CHAPITRE 2 95
DIFFERENT MODES OF COOPERATION BETWEEN *CIS*-ACTING ELEMENTS GOVERN THE EXPRESSION OF THE *KROX20/EGR2* HINDBRAIN DEVELOPMENTAL GENE 95
CHAPITRE 3 123
DEVELOPPEMENT DE DEUX SYSTEMES POUR L'ETUDE DE LA REGULATION DU GENE *KROX20* CHEZ LES VERTEBRES 123
1. DERIVATION DE CELLULES SOUCHES EMBRYONNAIRES MURINES 125
2. DELETION D'UN ENHANCER CHEZ LE POISSON-ZEBRE 132

DISCUSSION 137

1. DELETION D'UN ELEMENT AUTOREGULATEUR CHEZ LA SOURIS 139
2. LA BOUCLE ENTRE L'ELEMENT A ET LE PROMOTEUR APPORTERAIT DE LA ROBUSTESSE 141
3. NOUVELLES QUESTIONS SUR LE ROLE DE L'ELEMENT C DANS L'EXPRESSION EN DEUX PHASES DU GENE *KROX20* 143
4. LES ELEMENTS *CIS*-REGULATEURS DU GENE *KROX20* DANS LE RHOMBENCEPHALE CHEZ LA SOURIS 145
5. LES ELEMENTS *CIS*-REGULATEURS DU GENE *KROX20* DANS LE RHOMBENCEPHALE CHEZ LE POISSON-ZEBRE 151
6. LIMITATIONS DE NOTRE SYSTEME EMBRYONNAIRE, VERS UNE ETUDE EN CELLULE UNIQUE ? 154

REFERENCES BIBLIOGRAPHIQUES 157

INTRODUCTION

Au cours du développement d'un animal, une cellule, le zygote, génère les différents types cellulaires et organes de l'organisme. Ces derniers acquièrent différentes morphologies et fonctions grâce à l'activité d'un ensemble de gènes. La première étape de l'expression des gènes est la transcription de l'ADN génomique à partir du promoteur en ARN par l'ARN polymérase II. Cependant la transcription seule est faible et requiert la présence de séquences régulatrices distinctes du promoteur que l'on appelle des éléments *cis*-régulateurs portés par l'ADN. Ces éléments permettent la régulation fine de l'expression spatio-temporelle des gènes notamment au cours du développement par la fixation de facteurs en *trans*. D'autres mécanismes de régulation complètent l'action des éléments *cis*-régulateurs et des facteurs en *trans* comme des modifications épigénétiques et des changements de conformation de l'ADN. Ce projet de thèse vise à caractériser les éléments *cis*-régulateurs contrôlant l'expression du gène *Krox20* dans le cerveau précoce en développement.

Je décrirai l'organisation des gènes et l'assemblage de la machinerie de transcription dans une première partie. Puis je rentrerai dans le détail des mécanismes de régulation de l'expression des gènes dans une deuxième partie. La troisième partie abordera l'importance des séquences *cis*-régulatrices au cours du développement, de l'évolution et dans certaines pathologies. Enfin, je ferai l'inventaire des connaissances sur le gène de développement *Krox20* et les arguments qui ont conduit à l'étude fonctionnelle des éléments *cis*-régulateurs de ce gène.

1. Les gènes eucaryotes et la transcription

Le génome est l'ensemble de l'ADN (acide désoxyribonucléique) d'un organisme qui constitue l'information génétique transmissible à la descendance. Il contient l'ADN chromosomique contenu dans le noyau chez les eucaryotes ainsi que l'ADN d'organelles tels les mitochondries ou les chloroplastes chez la lignée verte. La séquence de l'ADN permet de produire l'ensemble des protéines de l'organisme de façon appropriée dans le temps et l'espace grâce à une série de réactions telles que la transcription et la traduction. Le génome humain est composé de $3{,}3.10^9$ paires de bases (pb) pour 20 000 gènes, celui de la souris est du même ordre de grandeur avec $2{,}8.10^9$ pb pour 25 000 gènes. La définition d'un gène est complexe et a beaucoup évolué grâce au séquençage systématique des génomes. Je dresserai un bref historique sur l'évolution de cette définition, puis détaillerai l'organisation des gènes et enfin le démarrage de la transcription.

1.1 Qu'est-ce qu'un gène ?

Le mot gène est utilisé pour la première fois en 1909 par Wilhelm Johannsen (Roll-Hansen 1989) faisant référence au concept de Gregor Mendel (Mendel & Bateson 1866) : un gène est une unité discrète héréditaire. En 1915, Thomas Hunt Morgan publie 'le mécanisme de l'hérédité mendélienne' qui propose que les gènes sont arrangés de façon linéaire en un locus distinct (Morgan et al. 1915). Le gène est ensuite vu comme une molécule physique à la suite des expériences sur le pneumocoque de Griffith et d'Avery qui montrent que la virulence bactérienne est véhiculée par un substrat qui peut être détruit pas l'enzyme ADNase (Griffith 1928; Avery et al. 1944). La découverte de la structure tridimensionnelle de l'ADN par Watson et Crick en 1953 permet d'expliquer comment l'information génétique peut être copiée grâce à la complémentarité des bases (Watson & Crick 1953). En 1958, Crick résume les travaux récents par le 'dogme central' en génétique : de l'ADN à l'ARN (acide ribonucléique) puis la protéine (Crick 1958). De nombreuses exceptions à ce dogme redéfinissent le gène comme un code porté par un acide nucléique qui donne un produit fonctionnel. Les techniques de clonage et de séquençage dans les années 1970 apportent de nombreuses informations sur l'expression et la cartographie des gènes. Ils se définissent alors comme des séquences de cadre ouvert de lecture grâce à la connaissance du code génétique (Doolittle 1986). À l'ère du séquençage entier de génome, la définition de Pearson propose qu'un gène est une région repérable de séquence génomique, correspondant à une unité héréditaire, qui peut être associée à des régions régulatrices, des régions transcrites et/ou d'autres séquences fonctionnelles (Pearson 2006). La définition computationnelle considère les gènes comme des sous-programmes dans le système opérationnel qu'est le génome, exécutés à travers la transcription et la traduction (Searls 2002).

Cette définition de gène comme un locus génétique compact ne peut s'appliquer en termes de régulation à distance, de chevauchement des gènes au sein du génome, d'épissage alternatif, d'éléments mobiles et d'ARN non codants (cf. Partie 2). Une définition plus juste est proposée par Gerstein : le gène est l'union de séquences génomiques codant un ensemble cohérent de produits fonctionnels potentiellement chevauchants (Gerstein et al. 2007).

1.2 Comment s'organisent les gènes ?

La transcription des gènes par l'ARN polymérase II chez les eucaryotes est contrôlée par des éléments en *cis* portés par l'ADN (fig. 1A) et des facteurs en *trans* qui fixent l'ADN. Deux catégories d'éléments en *cis* se distinguent : le promoteur composé d'un promoteur cœur et d'éléments régulateurs proximaux, ainsi que des éléments régulateurs distaux

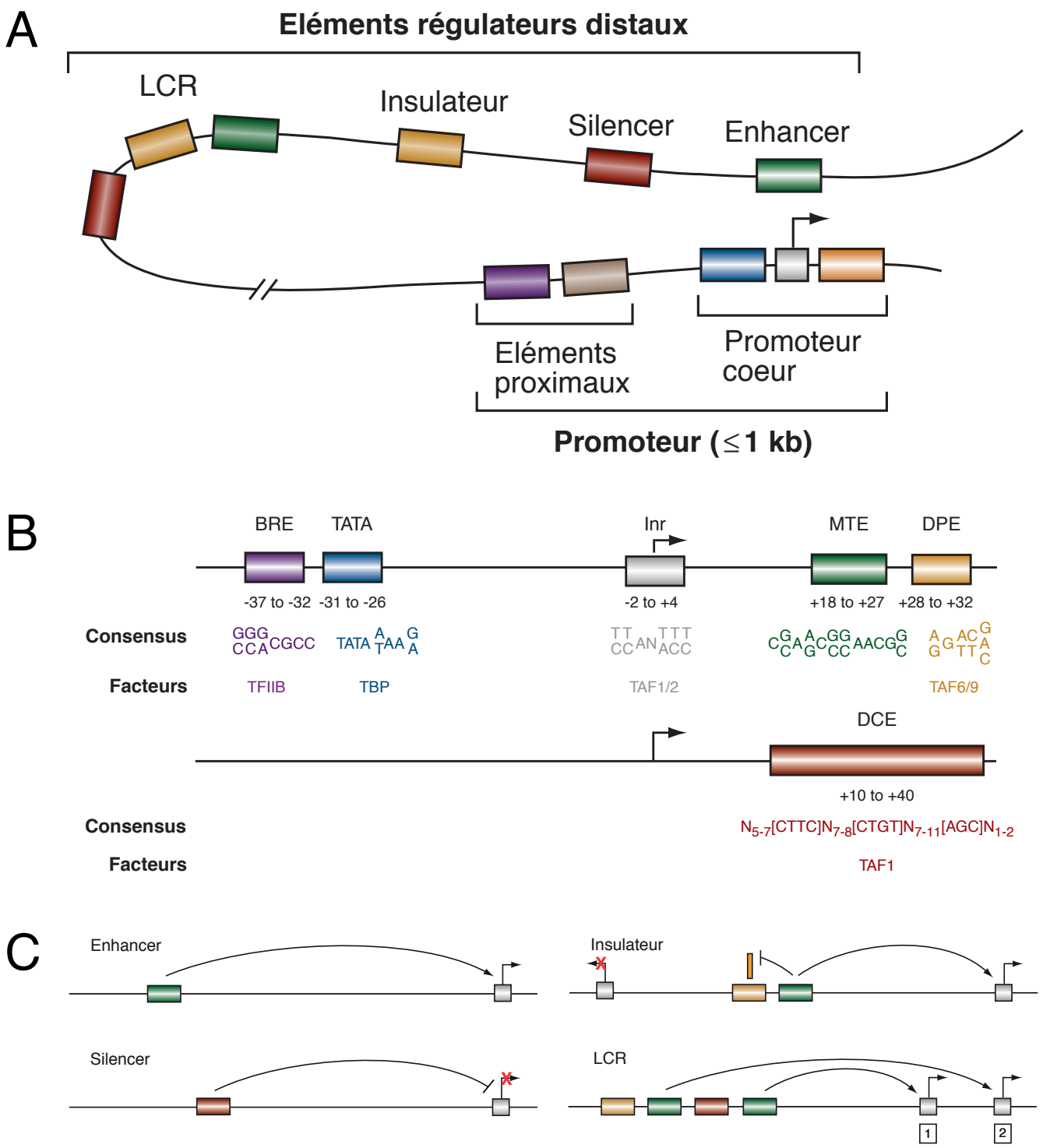


Figure 1 : Organisation en *cis* d'un gène

A. Les différents éléments d'un gène : le promoteur divisé en promoteur cœur et éléments proximaux, et les régulateurs distaux : LCR, insulateur, silencer et enhancer.

B. Les éléments du promoteur associés à leur position par rapport au +1 de transcription, les séquences consensus et les facteurs généraux de la transcription s'y fixant.

C. Les éléments distaux et leurs actions sur le promoteur. L'enhancer active l'expression du gène, au contraire du silencer qui réprime le gène. L'insulateur sert d'élément barrière à la transcription d'un gène proche et la LCR regroupe des éléments activateurs pour l'expression combinée de plusieurs gènes (ici nommés 1 & 2).

LCR : Locus Control Region, BRE : TFIIB Recognition Element, INR : INitiator Element, MTE : Motif Ten Element, DPE : Downstream Promoter Element, DCE : Downstream Core Element, TAF : TBP-associated Factor, TBP : TATA-binding protein.

D'après Maston et al., 2006.

composés des enhancers, des silencers, des insulators et des LCR (locus control region) et/ou super-enhancers (fig. 1A). Les facteurs en *trans* comprennent les facteurs généraux de transcription, les coactivateurs et les facteurs de transcription (TF) spécifiques à chaque type cellulaire. L'émergence des Vertébrés est associée à l'apparition de la pléiotropie, des gènes multifonctionnels et de fait à la multiplication, la diversification et la complexification des modules régulateurs.

1.2.1 Le promoteur

Le promoteur cœur est la région de démarrage de la séquence d'un gène où se fixe la machinerie basale de transcription et permet de définir le +1 de transcription ou TSS (transcription start site). Il comprend la TATA box (fig. 1B) où se fixe la protéine TBP (TATA-binding-protein) qui est une sous-unité du complexe $TF_{II}D$. Ensuite l'élément BRE est l'élément de reconnaissance de $TF_{II}B$ et l'élément d'initiation INR au niveau du TSS (fig. 1B). On peut y trouver l'élément cœur aval (DCE) et l'élément promoteur aval (DPE) qui contactent le cofacteur TAF1 et TAF6/8 respectivement, ainsi que le MTE (Motif Ten Element) (fig. 1B). Ces différents éléments du promoteur cœur ne sont pas exclusifs les uns des autres et sont parfois même complétement absents du promoteur.

Le promoteur proximal ou élément proximal est défini comme une région directement en amont de la TATA box, où se fixent de façon constitutive des facteurs communs à toute cellule et qui pourrait participer au rapprochement des éléments distaux. Une étude chez la drosophile propose que la combinaison d'éléments du promoteur cœur permet la sélection préférentielle des enhancers distaux (Zabidi et al. 2014). Cette sélection autorise alors l'expression ubiquitaire des gènes de ménage grâce à des enhancers constitutivement actifs contrairement à l'expression spatio-temporelle précise de gènes de développement par l'intermédiaire d'enhancers dont l'activité est modulable.

Une majorité des gènes codants présentent des promoteurs alternatifs chez les mammifères (Davuluri et al. 2008). Ces derniers génèrent des transcrits alternatifs dont l'expression spatio-temporelle peut varier. Par exemple, le promoteur sans TATA box du gène *HBG1* (Hemoglobin 1) est utilisé pendant le développement embryonnaire alors que le promoteur alternatif contenant une TATA box permet l'expression du gène après la naissance (Duan et al. 2002). Ceci suggère que la machinerie basale de transcription peut être recrutée au niveau de promoteurs cœurs différents de façon spécifique au cours du développement.

1.2.2 Les enhancers et silencers

Le terme d' « enhancer transcriptionnel » a été utilisé pour la première fois par Schaffner pour décrire une courte séquence d'ADN du virus SV40 capable d'augmenter significativement la transcription du gène humain de la ß-globine (Banerji et al. 1981). Le premier enhancer humain découvert fut celui de la chaîne lourde de l'immunoglobuline qui présente une activité tissu-spécifique (Banerji et al. 1983; Gillies et al. 1983). Les enhancers et les silencers sont alors décrits largement comme des séquences d'ADN de quelques centaines de paires de bases permettant de moduler la transcription de façon positive pour les enhancers (fig. 1C) et négative pour les silencers (fig. 1C). Cette modulation intervient quelles soit leur orientation ou leur distance du promoteur et de façon spatio-temporelle (Maniatis et al. 1987). Les enhancers et silencers sont composés de séquences dégénérées de quelques paires de bases qui correspondent aux sites de fixation pour des TFs activateurs ou répresseurs, respectivement. Ces séquences sont souvent fonctionnellement conservées entre différentes espèces lorsque l'expression spatio-temporelle d'un gène est identique. Cependant la conservation de séquence est beaucoup plus rare, rendant l'identification des enhancers par alignement de séquences peu efficace. Une combinaison de plusieurs enhancers peut être utilisée pour réguler un seul gène à la fois dans des tissus différents comme au sein du même tissu. Ces enhancers peuvent alors être redondants. Le terme de 'shadow enhancer' énoncé par Mike Levine (Hong et al. 2008) décrit un enhancer distant du promoteur, découvert après le premier élément plus proximal d'un gène et possédant des sites de fixation pour les mêmes TFs, ce qui sous-tend une même logique transcriptionnelle. Ils apporteraient de la robustesse et de la précision pour l'expression spatio-temporelle des gènes de développement notamment et constitueraient une source d'innovation évolutive (Lagha et al. 2012; Barolo 2012, cf. 3.1.1). Le terme est controversé car il fait référence à des notions déjà associées aux enhancers ; Barolo proposera de les appeler 'enhancers distribués'.

1.2.3 Les insulateurs

Les insulateurs, ou éléments barrières, empêchent les gènes d'être affectés par la transcription de gènes voisins (fig. 1C). Ils permettent de structurer le génome en domaine transcriptionnel. Ils ont deux propriétés principales : ils bloquent la communication entre enhancers et promoteurs et évitent la propagation d'hétérochromatine associée à une répression de la transcription. Chez les Vertébrés, le facteur CTCF (CCCTC-binding factor) fixe les insulateurs et a été associé aux deux propriétés précédentes (Kellum & Schedl 1991; Bell et al. 1999). Le rôle du facteur CTCF est ambivalent et j'y reviendrai.

1.2.4 Les LCR et/ou super-enhancers

Les LCR (Locus Control Regions) sont composés de plusieurs éléments régulateurs qui fonctionnent ensemble pour assurer l'expression spatio-temporelle d'un groupe de gènes proches (fig. 1C). Le premier LCR identifié est celui du locus humain de la ß-globine (Grosveld et al. 1987). Il comprend cinq sites hypersensibles à l'ADNase I qui contiennent de nombreux sites de fixation pour les TFs spécifiques de la lignée érythroïde (Goodwin et al. 2001). C'est un modèle d'interaction à distance entre promoteurs et enhancers.

Les super-enhancers ont été décrits dans des cellules souches embryonnaires (ES) murines de la façon suivante : 1) ces enhancers sont fixés par les facteurs de pluripotence tels Nanog, Oct4 et Sox2, 2) ces enhancers forment un concaténat sur une douzaine de kilobases, 3) un enrichissement fort de la fixation du Médiateur sur ces enhancers est observé par rapport aux autres régions alentour (Whyte et al. 2013). Un super-enhancer désigne donc un regroupement d'enhancers forts qui contrôlent des gènes clefs de la différenciation.

La pertinence du terme super-enhancer est controversée (Pott & Lieb 2015) car il inclut aussi les LCRs et l'archipel régulateur du groupe HoxD (Montavon et al. 2011). De plus la définition des super-enhancers n'est pas associée à une nouvelle notion de mécanistique des séquences régulatrices.

1.3 Les acteurs de la transcription

La transcription des gènes codants conduit à la production d'ARNm qui sera l'intermédiaire ribonucléique de l'information génétique avant la traduction en protéines. La transcription des gènes eucaryotes par l'ARN polymérase de type II implique trois types de facteurs : les facteurs généraux de la transcription (GTF), les activateurs spécifiques au niveau du promoteur et des enhancers ainsi que les co-activateurs. Ces facteurs forment des complexes multiprotéiques qui s'assemblent au niveau des éléments *cis*-régulateurs et dirige l'ARN polymérase vers le TSS (Maston et al. 2006).

1.3.1 Les facteurs généraux de la transcription

Les GTFs fixent le promoteur cœur en formant le complexe de pré-initiation (PIC). Ce dernier s'assemble de façon linéaire au niveau du promoteur. La protéine TBP qui fait partie du facteur $TF_{II}D$ fixe la TATA box auquel s'associe $TF_{II}A$ pour l'étape de nucléation. Vient ensuite le facteur $TF_{II}B$ qui s'associe au niveau du TSS et de BRE et crée le lien pour le recrutement de l'ARN polymérase II et le facteur $TF_{II}F$. L'étape de maturation est achevée par

la fixation de $TF_{II}E$ et $TF_{II}H$ au complexe préexistant (fig. 2A). Le PIC seul n'est pas capable d'activer la transcription *in vivo* (Poss et al. 2013).

1.3.2 Les activateurs en *trans* ou facteurs de transcription

Les activateurs de la transcription sont les facteurs de transcription qui fixent le promoteur proximal et les enhancers distaux. Ils sont composés d'un domaine de *trans*-activation où se lient d'autres facteurs pour réguler la transcription et un domaine de fixation à l'ADN pour la reconnaissance. Les facteurs de transcription sont regroupés par familles en fonction de leur domaine de fixation à l'ADN comme par exemple les homéodomaines des gènes *Hox* et les domaines à doigts à zinc du gène *Krox20*. Ils fixent des séquences nucléotidiques dégénérées qui peuvent être déterminées *in vitro* par exemple par EMSA (Electrophoretic Mobility Shift Assay) ou SELEX (Systematic Evolution of Ligands by EXponential Enrichment) et *in vivo* par ChIP (Chromatin ImmunoPrecipitation). Ces expériences permettent de générer des matrices de faible ou de forte affinité pour des recherches *in silico* de sites de fixation putatifs aux niveaux d'éléments *cis*-régulateurs. La combinaison des facteurs de transcription fixés à ces éléments définit la spécificité et l'activité spatio-temporelle de ce dernier. L'élément *cis*-régulateur comprend plusieurs sites de fixation pour des facteurs de transcription organisés selon un ordre, une orientation et un espacement particuliers que l'on peut assimiler à une « grammaire ». Typiquement, cette grammaire assure que les facteurs de transcription sont correctement agencés pour faciliter les interactions protéines-protéines, le recrutement de co-facteurs et la machinerie de transcription. La dissection fonctionnelle de plusieurs enhancers a conduit à proposer 3 modèles d'activité des enhancers par la fixation des facteurs de transcription (pour revue : Spitz & Furlong 2012) :

- Le modèle enhancéosome : les enhancers sont composés d'un ensemble serré de sites de fixation contactés par des facteurs de transcription selon un ordre précis dans le temps et dans l'espace. Les facteurs de transcription de l'enhancéosome agissent ensemble de façon synergique (Thanos & Maniatis 1995; Merika & Thanos 2001).
- Le modèle du « billboard » : les enhancers présentent des sites de fixation arrangés de façon flexible avec de plus faibles contraintes de distance entre les sites où les facteurs de transcription agissent de façon coopérative et additive. La mutation d'un site n'abolit pas complètement l'activité de l'enhancer suggérant que seulement quelques sites sont actifs à la fois (Lam et al. 2015; Arnosti & Kulkarni 2005).

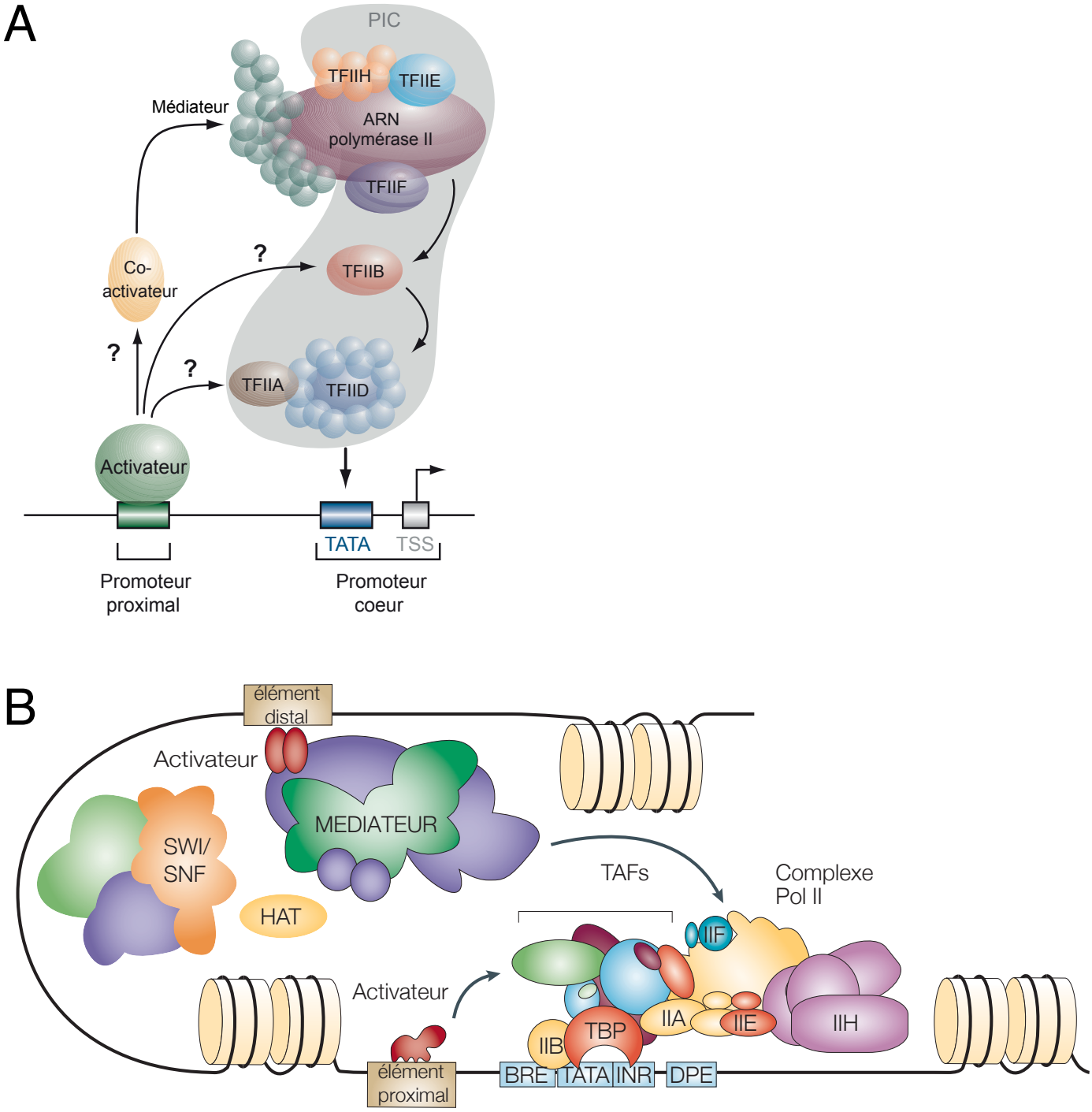


Figure 2 : L'assemblage de la machinerie de transcription

A. Interactions entre les facteurs en *trans* : le complexe de pré-initiation (PIC) comprenant TFIIA, TFIIB, TFIID, TFIIE, TFIIE, l'ARN polymérase II, les activateurs proximaux, les co-activateurs de la machinerie de transcription incluant le médiateur et les éléments en *cis* du promoteur cœur et du promoteur proximal. Les activateurs contactent directement ou indirectement grâce aux co-activateurs le PIC et le médiateur (points d'interrogation).

B. Schéma général de l'assemblage de la machinerie de transcription comprenant les facteurs généraux TBP, TFIIA,-B, -E, -F, -H qui fixent les éléments du promoteur (BRE, TATA box, INR et DPE), accompagnés de l'ARN polymérase II et les co-facteurs TAFs, les activateurs qui fixent les éléments proximaux et distaux, ainsi que les co-activateurs : le complexe SWI/SNF, les HATs et le médiateur.

BRE : TFIIB Recognition Element, INR : INitiator Element, DPE : Downstream Promoter Element, TAF : TBP-associated Factor, TBP : TATA-binding protein, HAT : histone déacétylase, SWI/SNF : Switch/Sucrose Nonfermetable, Pol II : ARN polymérase II.

D'après Maston et al., 2006 (A), Taatjes et al., 2004 (B).

- Le modèle collectif : plusieurs enhancers comprenant des sites de fixation plutôt relâchés recrutent de façon coopérative plusieurs facteurs de transcription identiques. Lorsque l'un des facteurs est inactivé, les autres ne parviennent pas à activer les enhancers (Junion et al. 2012).

1.3.3 Les co-activateurs

Les facteurs en *trans* contactent directement ou indirectement grâce aux co-activateurs les GTFs pour promouvoir l'initiation de la transcription (fig. 2A) et/ou le remodelage de la chromatine notamment grâce au complexe SWI/SNF (SWitch/Sucrose Non Fermentable et fig. 2B). Les co-activateurs ont le même rôle que les activateurs mais par l'intermédiaire de contacts protéine-protéine. Parmi les co-activateurs, deux catégories se distinguent : les facteurs associés à TBP ou TAFs et le Médiateur. Le rôle principal des TAFs est de relier les facteurs de transcription au GTFs pour faciliter l'assemblage du PIC. Cette fonction est notamment associée, d'une part, à la reconnaissance des marques d'histones au niveau du promoteur et, d'autre part, à leur activité de modification enzymatique des histones (acétylases ou HAT, fig. 2B). Le Médiateur est un complexe protéique gigantesque de 26 sous unités chez les mammifères. Cette molécule très flexible agit comme un pont pour la communication entre les facteurs de transcription et l'ARN polymérase II (fig. 2B). Ce rôle structural est indispensable dans les phénomènes de boucles d'ADN entre promoteur et enhancers distaux qui nous intéresserons particulièrement.

1.3.4 Les répresseurs et co-répresseurs

Les facteurs de transcription répresseurs exercent leur fonction par le recrutement de co-répresseurs. Il n'existe qu'une douzaine de complexes co-répresseurs et la plupart possèdent des activités désacétylases de type HDAC (pour revue : Payankaulam et al. 2010). L'acétylation des histones par les histones acétyltransférases (HATs) est associée à l'activation transcriptionnelle. Le recrutement de désacétylases ou l'interférence avec les HATs est donc le moyen le plus utilisé en répression transcriptionnelle. Les complexes les plus étudiés comprennent les complexes Sin3 et Groucho, exprimés de façon ubiquitaire, qui permettent le recrutement des HDAC directement grâce à des facteurs de transcription ou indirectement grâce à d'autres complexes co-répresseurs. Le faible nombre de ces complexes co-répresseurs est compensé par la grande diversification dans l'assemblage des différents facteurs notamment grâce à la présence de nombreux gènes paralogues chez les mammifères.

Les différents acteurs décrits précédemment sont cruciaux pour la transcription correcte des gènes. L'expression des gènes est cependant régulée par de nombreux mécanismes afin d'obtenir des patrons d'expression précis dans le temps et l'espace. Je détaillerai dans la partie suivante le rôle déterminant des facteurs de transcription à travers plusieurs niveaux de régulation transcriptionnelle.

2. La régulation de l'expression génétique chez les eucaryotes

Pendant le développement, un enchaînement d'étapes précises conduit, à partir d'une cellule œuf, à un organisme entier dont chaque cellule possède le même patrimoine génétique. Les différents types cellulaires mettent en œuvre des programmes génétiques spécifiques et stéréotypés, impliquant l'activation séquentielle de certains gènes au cours de la différenciation et pour le métabolisme correct de l'organisme adulte. Ceci sous-tend une régulation fine de l'expression des gènes au sein d'un même réseau génétique. Cette régulation, au niveau transcriptionnel, est essentiellement le fait de la présence des facteurs en *trans*. Cependant, plusieurs niveaux de modulation existent, comme les modifications épigénétiques, les interactions entre promoteurs et éléments *cis*-régulateurs et la conformation globale de l'ADN au sein du noyau.

2.1 La régulation épigénétique

Le terme « épigénétique » a été défini par Conrad Waddington en 1942 comme « la branche de la biologie qui étudie les interactions causales entre les gènes et leurs produits qui induisent le phénotype » (Waddington 1942). Cette première définition n'inclut pas la notion d'héritabilité que défendra bien plus tard Robin Holliday en décrivant l'épigénétique comme « l'étude des changements de la fonction d'un gène qui sont héritables par mitose ou méiose sans entraîner de modification de la séquence d'ADN » (Holliday 1994). Le terme d'épigénétique est depuis utilisé pour décrire, d'une part, l'impact de l'environnement sur le génotype et le phénotype qui en résulte, donc plutôt associé à l'écologie, et, d'autre part, la persistance d'un patron d'expression entre les différentes cellules et à travers les générations, en relation avec la génétique (Deans & Maggert 2015). Il semble que le terme soit donc utilisé de façon dichotomique dans l'une ou l'autre de ces branches de la biologie. Je m'intéresserais ici à la définition 'génétique' en considérant la régulation épigénétique de l'expression des gènes eucaryotes. Je présenterai différents phénomènes épigénétiques comme la méthylation de l'ADN, le remodelage de la chromatine qui suit la fixation des facteurs pionniers, les modifications d'histones et les ARN longs non-codants.

2.1.1 La méthylation de l'ADN et les îlots CpGs

La méthylation de l'ADN est un des mécanismes épigénétiques majeurs chez les mammifères pour la régulation de l'expression génétique tissu-spécifique, l'inactivation du

chromosome X, l'empreinte génomique et la répression des éléments transposables. Elle consiste au transfert d'un groupement méthyl sur une cytosine formant une 5-méthylcytosine. La majorité des cytosines méthylées précède une guanine, ces dinucléotides sont référencés comme des sites CpGs. Chez les mammifères, les sites CpGs sont sous-représentés dans le génome à cause de la désamination des 5-méthylcytosines en thymine (pour revue : Moore et al. 2012). Une exception existe cependant pour les îlots CpGs : de grandes régions à forte densité de CpGs le plus souvent non méthylés. Les îlots CpGs couvrent le TSS des promoteurs des gènes de ménage et des gènes régulés au cours du développement. La méthylation des îlots CpGs est associée à la répression stable de l'expression des gènes.

La méthylation de l'ADN peut contrôler l'expression des gènes de quatre façons :

- en empêchant la fixation d'un facteur de transcription sur sa séquence consensus méthylée
- en interférant avec des marques positives d'histones (cf 2.1.2)
- en recrutant des répresseurs qui fixent les groupements méthyles
- en affectant la disposition des nucléosomes.

Les méthyltransférases d'ADN (DNMTs) catalysent la réaction de méthylation. Deux types de DNMT assurent soit le maintien de la méthylation, soit la méthylation de *novo*. Le maintien de l'état de méthylation est perpétué par l'enzyme DNMT1 qui agit sur le brin non-méthylé après chaque réplication (Yoder & Bestor 1998). Les enzymes DNMT3A et DNMT3B (Okano et al. 1999), avec l'aide de DNMT3L (Bourc'his et al. 2001) ciblent des sites différents du génome et établissent un nouvel état de méthylation. Le retrait de la méthylation de l'ADN peut se faire de façon passive ou active. De manière passive, une inhibition de la DNMT1 conduit à la perte progressive de l'état de méthylation. Plusieurs mécanismes entraînant une déméthylation active sont encore débattus. Les deux stratégies consistent soit à la déamination active des méthylcytosines en thymine comme vu précédemment soit à la formation d'un intermédiaire par ajout d'un groupement hydroxyl menant à l'hydroxyméthylcytosine par les enzymes de la famille TET (Tahiliani et al. 2009). Ces deux composés sont ensuite pris en charge par d'autres enzymes pour la conversion en cytosine.

La méthylation est principalement identifiée par séquençage bisulfite qui consiste au traitement de l'ADN par du bisulfite qui convertit les cytosines en uraciles mais n'affecte pas les méthylcytosines. Le génome est d'abord fragmenté par digestion enzymatique pour la technique de RRBS (Reduced Representation Bisulfite Sequencing, Meissner et al. 2005) ou par sonication pour le WGBS (Whole Genome Bisulfite Sequencing, Lister et al. 2009) avant traitement au bisulfite et séquençage permettant une résolution au nucléotide près.

2.1.2 Le remodelage de la chromatine et les facteurs de transcription pionniers

La chromatine est l'organisation compacte et dynamique de l'ADN autour de protéines appelées histones, qui peut être divisée en euchromatine ouverte et en hétérochromatine condensée. L'unité de répétition de la chromatine est le nucléosome : un octamère d'histones (H2A, H2B, H3 et $H4)_2$ entouré d'environ 200 pb d'ADN. Les nucléosomes peuvent être organisés en fibres de 30 nm grâce aux histones de liaison de types H1 ou H5. La présence de nucléosomes au niveau d'un promoteur inactif empêche les GTFs et l'ARN polymérase II de fixer l'ADN. La transition vers un état actif se fait par un remodelage de la chromatine qui va rendre l'ADN accessible. Ce remodelage peut se faire de différentes façons (fig. 3A) :

- Par glissement : les octamères d'histones peuvent glisser le long de l'ADN pour rendre accessible une séquence particulière.
- Par ajustement : la distance entre les octamères d'histones peut être modifiée.
- Par éviction : un octamère d'histones peut être complétement déplacé, créant alors une zone libre de nucléosome.

Plusieurs complexes de remodelage ATP-dépendants existent chez les mammifères. Les plus connus sont le complexe activateur SWI/SNF et le complexe répresseur ISWI (cf 1.3.2). Un autre complexe INO80/SWR est impliqué dans l'échange des histones et notamment de l'histone variant H2A.Z (Kobor et al. 2004). Les nucléosomes contenant H2A.Z et/ou H3.3 sont plus instables et le dépôt au niveau des promoteurs et des enhancers est associé à l'activation de la transcription (Jin et al. 2009 et fig. 3C). Les complexes de remodelage de la chromatine sont recrutés à l'ADN par des facteurs de transcription « pionniers » qui fixent l'ADN indépendamment de la position des nucléosomes, rendant accessible la chromatine aux autres facteurs de transcription pour l'expression génétique à un stade plus tardif au cours de la différenciation. Ils peuvent rester fixés tout au long de l'activation de l'enhancer ou peuvent être remplacés par d'autres facteurs de transcription formant une grande plateforme de recrutement. Certains facteurs comme FOXA possèdent un domaine de fixation ressemblant à celui des histones de liaison qui permettent cette fois la décompaction de la chromatine par l'altération de la structure des nucléosomes (Cirillo et al. 2002; Iwafuchi-Doi et al. 2016). Le facteur FOXA1 est en plus capable de favoriser un changement épigénétique lors de sa fixation avec une déméthylation de l'ADN et la déposition de la marque H3K4me1 sur les histones (Sérandour et al. 2011). Chez la drosophile, le facteur Zelda est requis pour l'activation des nombreux enhancers à la transition maternelle-zygotique (Liang et al. 2008). Ce facteur possède deux propriétés inhabituelles. D'une part, il occupe plus de la moitié de ses sites potentiels de fixation et d'autre part il est

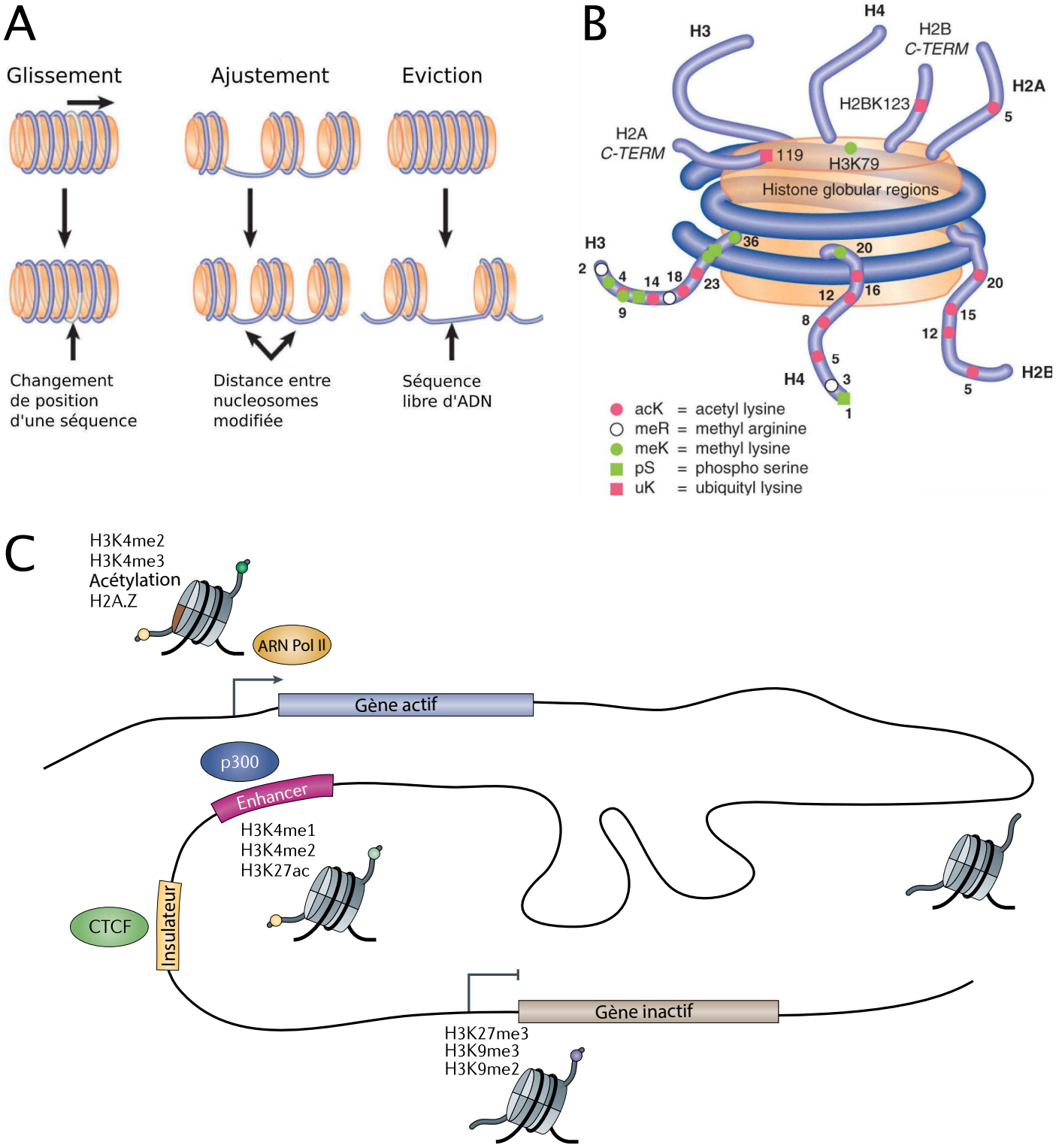


Figure 3 : Le remodelage de la chromatine et les modifications d'histones

A. Trois mécanismes de remodelage de la chromatine existent : le glissement des nucléosomes le long de l'ADN pour découvrir une séquence d'intérêt, le réajustement des distances entre nucléosomes ainsi que l'éviction d'un ou plusieurs nucléosomes découvrent l'ADN.

B. Les queues d'histones, dépassant des régions globulaires des histones, peuvent être acétylées, méthylées, phosphorylées et ubiquitinilées sur différents sites.

C. Les modifications d'histones délimitent les éléments fonctionnels du génome : les promoteurs, les enhancers et les insulateurs. Les promoteurs actifs sont marqués par H3K4me2, H3K4me3, acétylation et l'histone variant H2A.Z. Les promoteurs inactifs sont décorés par H3K27me3, H3K9me3, H3K9me2. Les enhancers sont enrichis par les marques H3K4me1, H3K4me2, H3K27ac et l'acétyltransférase d'histone p300. Le facteur CTCF fixe les insulateurs.

D'après Lewin's Genes X. _ 10th ed. (A et B), Zhou et al., 2011 (C).

étroitement lié à un changement de l'expression des gènes (Nien et al. 2011; Harrison et al. 2011). Deux études récentes ont montré que Zelda promeut l'accessibilité des enhancers à d'autres facteurs de transcription, par éviction des nucléosomes en association avec le facteur GATA et conduisant à l'activation du génome zygotique (Sun et al. 2015; Schulz et al. 2015). Ces résultats confirment le rôle de Zelda comme facteur pionnier qui, de façon surprenante n'est pas conservé chez les Vertébrés.

Les zones de décompaction de l'ADN peuvent être observées par le test d'hypersensibilité à l'ADNase I suivi de séquençage (Boyle et al. 2008), le FAIRE-seq (Formaldehyde Assisted Isolation of Regulatory Elements following by sequencing, Giresi et al. 2007) et plus récemment par l'ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing, Buenrostro et al. 2013). Ces techniques se basent sur la sensibilité de la chromatine à la fragmentation, par une endonucléase (Dnase I), par sonication après fixation en paraformaldéhyde (FAIRE-seq), ou par insertion d'étiquettes par une transposase pour une amplification par PCR (ATAC-seq). Ces zones de décompaction de l'ADN correspondent aux promoteurs actifs mais aussi aux éléments *cis*-régulateurs pour lesquels ces techniques sont un moyen d'identification à l'échelle génomique.

2.1.3 Les modifications d'histones et l'activation séquentielle des enhancers

Les queues d'histones dépassent de chaque côté du nucléosome de façon désordonnée et sont sujettes à des interactions avec d'autres protéines et des modifications post-traductionnelles. Toutes les histones peuvent être modifiées de façon covalente à différentes positions par méthylation, acétylation, phosphorylation, sumoylation ou ubiquitination (fig. 3B). La lysine (K) est la cible la plus commune pour ces modifications. La méthylation des lysines ne change pas la charge positive de l'acide aminé alors que l'acétylation des lysines apporte une charge neutre qui déstabilise l'interaction avec l'ADN chargé négativement et favorise donc la décompaction de la chromatine. De plus, l'acétylation des lysines est reconnue par des facteurs à bromodomaine alors que la méthylation est fixée par les facteurs à chromodomaine, les domaines Tudor et MBT. L'acétylation des histones est un équilibre entre l'activité enzymatique des HATs et des HDACs. Les HATS sont réparties en deux groupes. Les HATs du groupe B acétylent les histones néosynthétisées dans le cytoplasme et participent à l'assemblage des nucléosomes. Celles du groupe A agissent directement sur les histones de la chromatine et sont impliquées dans le contrôle de la transcription comme co-activateurs (cf 1.3.2). Le co-activateur le mieux décrit est p300/CBP (CREB binding protein, fig. 3C) qui acétyle majoritairement les histones au niveau des

enhancers. L'analyse pangénomique de la fixation de p300/CBP par immunoprécipitation de chromatine suivie de séquençage (ChIP-seq) a été notamment utilisée pour l'identification des enhancers (Visel et al. 2009). Les HDACs sont des répresseurs transcriptionnels (cf 1.3.3). Les HATs et les HDACs sont généralement peu spécifiques (pour revue : Verdin & Ott 2014). La méthylation des histones est plus complexe car elle peut être associée à l'activation ou la répression de la transcription. Les histones méthyltransférases (HMTs) ciblant la lysine contiennent un domaine SET ou appartiennent aux protéines DOT1 (Disruptor Of Telomeric silencing 1), celles ciblant l'arginine possèdent un domaine PMRT. La méthylation a longtemps été considérée comme irréversible à cause de son extrême stabilité. Depuis une dizaine d'années, plusieurs histones déméthylases (HDMTs) ont été découvertes comme LSD1 (Lysine-Specific Demethylase 1) et les protéines à domaine Jumonjy C (pour revue : Greer & Shi 2012).

La combinaison de marques spécifiques d'histones détermine l'état transcriptionnel d'une région génomique ; ce concept est appelé le code histone (Jenuwein & Allis 2001). Des modifications associées aux promoteurs, aux éléments *cis*-régulateurs et au corps d'un gène caractérisent l'état actif ou inactif d'un locus. On distingue ainsi (pour revue : Zhou et al. 2011 et fig. 3C) :

- Les marques H3K36me3 et H3K79me2 au niveau du corps des gènes transcrits et en particulier des exons.
- Les marques H3K4me3 et H3K27ac au niveau des promoteurs actifs, H3K9me3 et H3K27me3 pour ceux inactifs et les promoteurs bivalents par H3K4me3 et H3K27me3. La marque H3K27me3 est déposée par le complexe PRC2 (Polycomb Repressive Complex) grâce à la protéine EZH2 (Enhancer of Zeste Homolog 2) qui permet ensuite le recrutement de PRC1 qui à son tour ubiquitine H2A119 grâce à RING1 (Ring Finger Protein 1). La fixation des complexes Polycomb au niveau des promoteurs chez les mammifères empêche le recrutement des complexes de remodelage de la chromatine induisant la compaction de la chromatine et réduisant l'accès aux facteurs de transcription (Grau et al. 2011). La marque H3K9me3 déposée par SUV38H1 est fixée par la protéine HP1 (Heterochromatin Protein 1) qui permet la formation et le maintien de l'hétérochromatine (Zeng et al. 2010).
- Les marques H3K4me1, H3K4me2 et H3K27ac décorent les enhancers actifs alors que ceux inactifs le sont par H3K9me2 et H3K9me3. Un troisième état « poised » (prêt à être activé) est associé aux enhancers dans le cas des cellules souches embryonnaires avec la présence de H3K4me1 et H3K27me3 et constitue un état intermédiaire qui faciliterait l'activation rapide des enhancers pour la différenciation.

H3K4me1 et H3K27ac sont les signatures chromatiniennes utilisées pour l'identification des enhancers notamment dans le cadre du projet ENCODE (cf 2.1.5). Le dépôt des marques de chromatine participe à l'activation séquentielle des enhancers et promoteurs. La fixation des facteurs pionniers est concomitante avec la fixation des complexes de remodelage de la chromatine et l'incorporation des histones variants formant des zones de déplétion des nucléosomes. Les nucléosomes flanquants sont alors marqués par H3K4me1 par les HMTs Mll3/4 recrutées par les facteurs pionniers. Les facteurs de transcription tissu-spécifique fixent ces zones libres et mènent au recrutement de p300/CBP pour l'acétylation de H3K27. Cette activation des enhancers précède celle des promoteurs et coïncide avec l'éviction des facteurs Polycomb, la perte d'H3K27me3 et le maintien d'H3K4me3 par le remplacement de Mll1/2 par SET1A/B et le recrutement de la machinerie de transcription. De plus, enhancers et promoteurs peuvent être fixés par les complexes Cohésine et Médiateur qui participent à la formation de boucles chromatiniennes (Kondo et al. 2016).

2.1.4 Les ARNs longs non-codants

Les ARNs longs non-codants (lncRNAs) sont des ARNs de plus de 200 nucléotides, souvent polyadénylés et épissés mais dépourvus de cadres ouverts de lecture (ORFs) évidents. Les promoteurs des lncRNAs sont marqués par H3K4me3 contrairement aux ARNs issus d'enhancers (eRNAs, cf 2.2.2). Les lncRNAs servent de ligands pour des protéines et fixent l'ADN par complémentarité entre acides nucléiques sur le site cible en apportant les complexes de régulation transcriptionnelle. Les lncRNAs peuvent contrôler les gènes proches de leur site de transcription sur le même chromosome, ils agissent en *cis* (fig. 4B), ou bien ils activent ou répriment des gènes situés en *trans* (fig. 4A). Un exemple bien connu de lncRNA agissant en *cis* est Xist (X-inactive specific transcript) impliqué dans l'inactivation d'un des chromosomes X chez les mammifères par la formation d'hétérochromatine notamment grâce au recrutement du complexe Polycomb PRC2. Un autre lncRNA HOTAIR (HOX transcript antisens RNA) est transcrit du cluster HOXC et réprime le cluster HOXD en *trans,* là encore par recrutement de complexes de modification de la chromatine (pour revue : Fatica & Bozzoni 2014). Cependant les mécanismes de ciblage de ces lncRNAs, aussi bien la rétention sur le lieu de transcription pour ceux agissant en *cis* que l'adressage à longue distance pour ceux agissant en *trans*, ne sont pas connus.

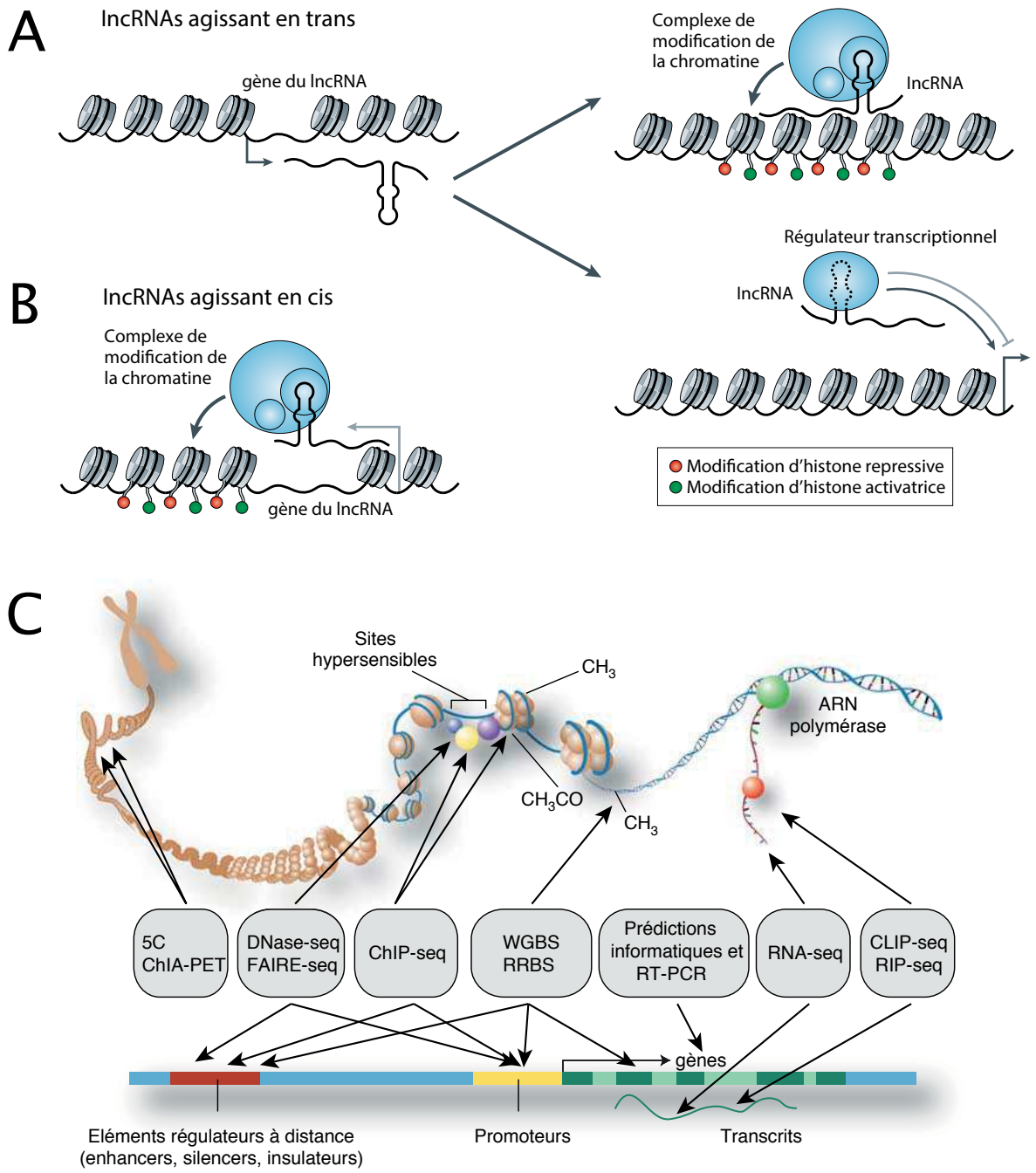


Figure 4 : Les modes d'action des ARNs longs non-codants et le projet ENCODE

A. Les lncRNAs peuvent agir en *trans* par le recrutement des complexes de modification de la chromatine ou directement des régulateurs transcriptionnels.

B. Les lncRNAs peuvent agir en *cis* par le recrutement des complexes de modification de la chromatine.

C. Le consortium ENCODE regroupe un grand nombre de méthodes biochimiques associées à du séquençage pour l'identification de gènes et d'éléments *cis*-régulateurs. Le 5C et le ChIA-PET pour capturer la conformation de l'ADN ; Dnase-seq et FAIRE-seq mettent en évidence les sites d'hypersensibilité correspondant aux zones de décompaction de l'ADN. La méthode de ChIP-seq permet d'identifier la position des marques d'histones ou des facteurs de transcription. Les sites de méthylation de l'ADN sont identifiés par WGBS ou RRBS. Les prédictions informatiques associées à des RT-PCR positionnent les gènes sur le génome. La technique de RNA-seq détermine le transcriptome de cellules ou tissus. Les méthodes de CLIP-seq et RIP-seq identifient les séquences de fixation de protéines associées au ARNs.

5C : Chromosome Conformation Capture Carbon Copy, ChIA-PET : Chromatin Interaction Analysis by Paired-End Tag Sequencing, ChIP/CLIP/RIP : Chromatin/Cross-Linking/RNA-binding protein ImmunoPrecipitation, FAIRE : Formaldehyde Assisted Isolation of Regulatory Elements, RRBS : Reduced Representation Bisulfite Sequencing, RT-PCR : transcription inverse puis PCR, WGBS : Whole Genome Bisulfite Shotgun, -seq : séquençage.

D'après Fatica & Bozzoni, 2014 (A et B) et le site : https://www.encodeproject.org/ (C).

2.1.5 Le consortium ENCODE

Le séquençage du génome humain achevé en 2001 a permis de révéler que les cadres ouverts de lecture représentent seulement 1,2% du génome (Lander et al., 2001 & Venter et al., 2001). Le reste du génome sera alors référencé comme 'junk DNA' d'après le terme utilisé par Ohno (Ohno 1972). Le projet ENCODE (Encyclopedia of DNA Elements) lancé en 2003, vise à référencer tous les éléments fonctionnels du génome humain et murin, comprenant les gènes, les transcrits, les régions régulatrices et les marques épigénétiques (Dunham et al. 2012; Feingold et al. 2004; Birney et al. 2007), en utilisant un grand nombre de techniques biochimiques reposant sur du séquençage à haut-débit (fig. 4C). En 2012, 80% du génome humain est assigné à une fonction biochimique, dont un grand nombre de séquences régulatrices potentielles ou correspondantes à des ARNs non-codants. Ainsi, le degré de complexité d'un organisme peut être mieux évalué par la quantité de transcription en ARN non-codant que par le nombre de gènes codants des protéines, même lorsque l'épissage alternatif et les modifications post-traductionnelles sont pris en compte (Taft et al. 2007). Ce projet constitue un catalogue de nombreuses données disponibles et facilement visualisables sur les navigateurs UCSC et Ensembl.

2.2 Les interactions enhancers-promoteurs

Chez les eucaryotes simples comme la levure, les éléments *cis*-régulateurs sont directement à proximité du promoteur autour d'1 kb. La drosophile est un organisme plus complexe dont le génome est plus vaste et comprend des régions régulatrices relativement proches du promoteur vers 10 kb. Chez les mammifères, les régions régulatrices sont plus diversifiées à la suite notamment de deux duplications complètes de génome. Les éléments *cis*-régulateurs peuvent alors être distants de plusieurs centaines de kb à plus d'un Mb du promoteur. La transcription des gènes cibles peut nécessiter la formation de boucles d'ADN qui rapprochent physiquement les enhancers du promoteur. Cette architecture en boucles d'ADN est modelée par des protéines de structure et conduit à des phénomènes de « pause » de l'ARN polymérase II, suivis de transcription souvent bidirectionnelle aussi bien au niveau des promoteurs que des enhancers.

2.2.1 Les protéines de structure modèlent l'architecture des boucles de chromatine

CTCF est un facteur de transcription à 11 doigts à zinc reconnaissant une séquence d'au moins 20 nucléotides non palindromique qui est conservée évolutivement chez la plupart

des animaux (Heger et al. 2012). Il a d'abord été décrit comme ayant un rôle de barrière à la propagation d'hétérochromatine puis comme bloqueur des contacts entre enhancers et promoteurs (cf 1.2.3). Ce second rôle a été déterminé par transgénèse mais peu d'exemples *in vivo* l'ont confirmé et des expériences plus récentes contredisent cet effet. L'utilisation d'une méthode dérivée du 3C (Chromosome Conformation Capture, Dekker et al. 2002) comme le 5C (Carbon-Copy 3C, Dostie et al. 2006) a montré que 79 % des interactions à longue distance entre des éléments *cis*-régulateurs distaux et promoteurs n'étaient pas bloquées par la fixation de CTCF (Sanyal et al. 2012). De plus, une partie de ces enhancers distaux coïncident avec un enrichissement de CTCF et/ou de marques d'histones activatrices telles H3K27ac, H3K4me1 et H3K4me2, supportant l'idée d'un rôle majeur de CTCF dans le ciblage des enhancers distaux aux promoteurs des gènes qu'ils régulent (Handoko et al. 2011). La topologie du génome au sein du noyau peut permettre de rationaliser les différents rôles de CTCF comme nous le verrons dans le paragraphe 2.3.1.

Le complexe cohésine forme un anneau composé des protéines SMC1A, SMC3, Rad21 et SA1 ou SA2 (fig. 5A) qui entoure l'ADN sans fixation directe par reconnaissance d'une séquence consensus. La cohésine est bien connue pour son rôle cohésif des chromatides sœurs après la réplication de l'ADN en phase S jusqu'en anaphase où elle est éliminée de la chromatine et les chromatides sœurs sont ségrégées dans les deux cellules filles. Cette fonction de la cohésine est essentielle et le complexe est très conservé au cours de l'évolution (pour revue : Nasmyth & Haering 2009). La cohésine est aussi un composant majeur de la chromatine des cellules interphasiques et coïncide avec la fixation du facteur CTCF (Wendt et al. 2008; Parelho et al. 2008). Le complexe cohésine régule donc l'activité des gènes grâce à la formation et la stabilisation des boucles à longue distance de la chromatine au niveau des sites de fixation de CTCF (Hadjur et al. 2009). De plus, il a été montré dans plusieurs types cellulaires que la cohésine et le médiateur interagissent physiquement et permettent de connecter les promoteurs et enhancers (Kagey et al. 2010, fig. 5B).

L'association de CTCF et de la cohésine dans les interactions à longue distance permet de réguler des *loci* complexes tels celui de la ß-globine dans les érythrocytes (Splinter et al. 2006) et des proto-cadhérines dans le cerveau (Remeseiro et al. 2012; Hirayama et al. 2012). Il a été montré que CTCF et la cohésine sont fortement enrichis au niveau des super-enhancers permettant de réguler l'expression des gènes par le regroupement spatial des enhancers (Ing-simmons et al. 2015) et que ces interactions interviennent entre deux sites CTCF co-occupés par la cohésine (Dowen et al. 2014). Plusieurs études ont mis en évidence que l'orientation convergente des sites de fixation de CTCF permet la formation des boucles de chromatine en association avec la cohésine (Vietri Rudan et al. 2015; Rao et al. 2014; Guo

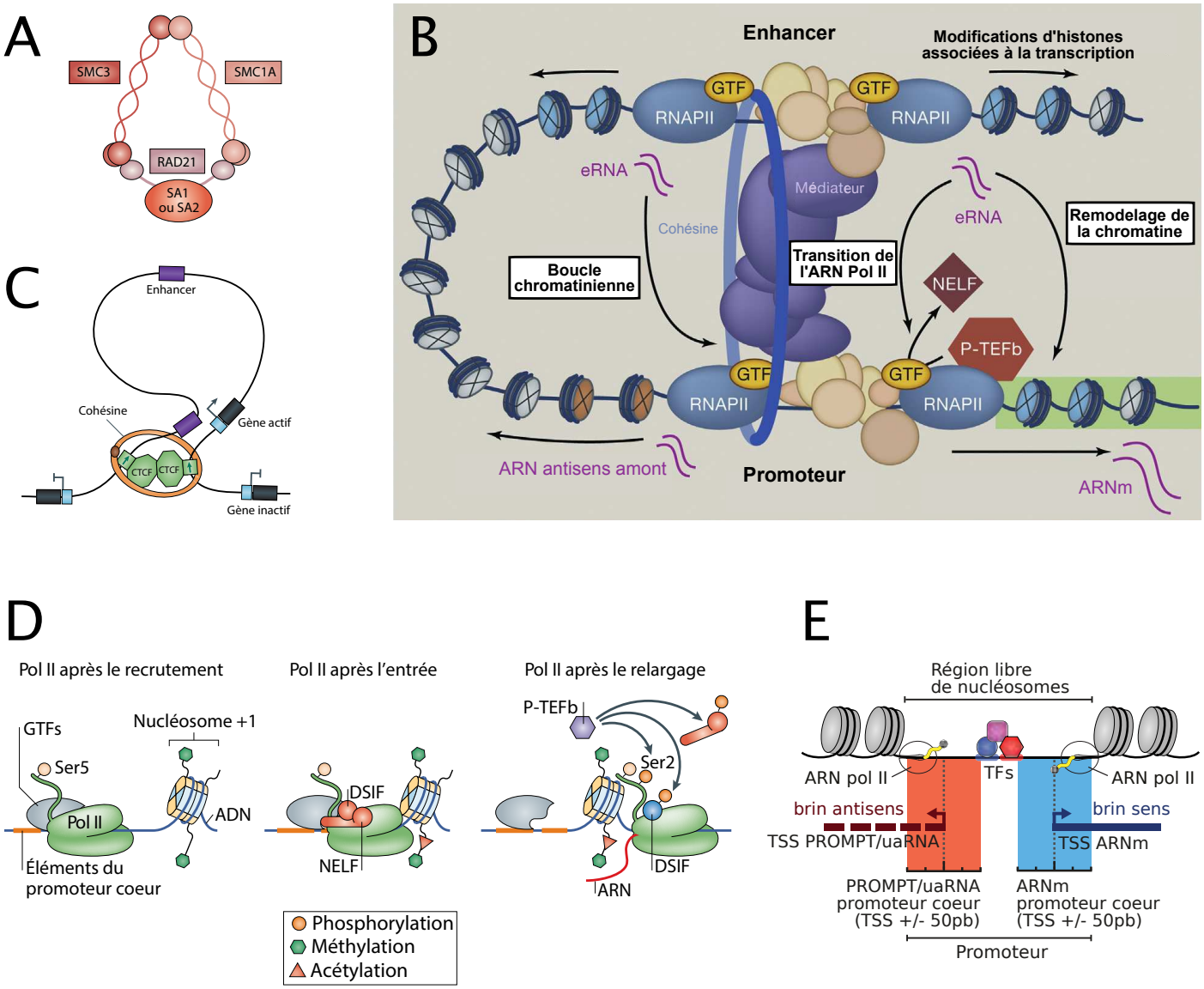


Figure 5 : Les boucles enhancers-promoteurs

A. L'architecture du complexe cohésine forme un anneau composé de SMC1A, SMC3, Rad21 et SA1 ou SA2 qui entoure l'ADN.

B. Les boucles enhancers-promoteurs sont structurées par la cohésine et le médiateur en association avec la Pol II et les GTFs. La transcription bidirectionnelle des enhancers, associée à la déposition de marques d'histones, produit des eRNAs qui agissent sur la boucle chromatinienne, le remodelage de la chromatine ainsi que sur le relargage de la Pol II en facilitant l'échange de P-TEFb contre NELF pour la production d'ARNm et d'ARN antisens au niveau du promoteur.

C. Boucle chromatinienne entre enhancer et promoteur d'un gène grâce à la fixation de facteurs CTCF et l'ancrage du complexe cohésine. L'orientation convergente des sites de fixation de CTCF est indiquée par les deux flèches vertes.

D. Recrutement de la Pol II et des GTFs au niveau des éléments du promoteur proximal, démarrage et entrée au +1 de transcription en association avec les facteurs DSIF et NELF et relargage de la Pol II pour l'élongation grâce recrutement du complexe de phosphorylation P-TEFb.

E. Un promoteur est divergent quand deux promoteurs cœurs sont opposés et séparés par une région libre de nucléosomes où se fixent les facteurs de transcription. Les Pol II fixent les TSS et s'arrêtent en pause en aval.

DSIF : DBR-Sensitivity-Inducing Factor; G(TFs) : General Transcription Factors; NELF : Negative Elongation Factor; Pol II et RNAPII: ARN polymérase II; PROMPT : Promoter Upstream Transcript; P-TEFb : Positive Transcription Elongation Factor-b; SMC1A/3 : Structural Maintenance of Chromosome proteins 1A/3; TSS : Transcription Start Site; uaRNA : upstream antisens RNA.

D'après Losada et al., 2014 (A), Kim and Shiekhattar, 2015 (B) Ong and Corces, 2014 (C), Jonkers and Lis, 2015 (D), Andersson et al., letter, 2015 (E).

et al. 2015; Wit et al. 2015 et fig. 5C et 6A). Cependant, l'utilisation du système d'édition du génome CRISPR/Cas9 a permis d'inverser l'orientation d'un site CTCF au niveau du super-enhancer de Sox2 et de montrer que la fixation de CTCF et de la cohésine n'est pas affectée, de même que l'expression du gène bien que la formation de boucles avec le promoteur soit abolie (Wit et al. 2015). De façon analogue, l'inversion d'un site CTCF au niveau de l'enhancer HS5-1 régulant les gènes du groupe Pcdhα (Protocadherin α) abolit les interactions avec les promoteurs de ce groupe alors qu'elles sont plus nombreuses pour ceux des groupes Pcdhβ et Pcdhγ. La fixation de CTCF reste identique mais celle de la cohésine diminue. Enfin l'expression des gènes du groupe Pcdhα est altérée mais il n'y a pas d'expression pour les autres groupes malgré les nouvelles interactions (Guo et al. 2015). Il semblerait donc que les interactions enhancer-promoteur ne soient pas indispensables à l'expression de certains *loci* qui serait plutôt la résultante de multiples acteurs. Par exemple, l'interaction entre le LCR et le promoteur de la ß-globine dépend aussi du facteur de transcription GATA1 et de la protéine de structure Ldb1. En l'absence de GATA1, le locus est inactif et déroulé. En adressant artificiellement Ldb1 au niveau du promoteur, l'interaction avec le LCR est restaurée et la transcription de la ß-globine réactivée (Deng et al. 2012). Le même groupe a montré de façon similaire que l'adressage de Ldb1 au promoteur du gène embryonnaire de la globine ßh1 permettant la réactivation de la transcription dans des érythroblastes primaires adultes, représentant une stratégie thérapeutique originale (Deng et al. 2014). Le rôle des protéines de structure dans les interactions à longue distance entre enhancers et promoteurs n'est pas encore totalement élucidé et dépend aussi probablement du locus considéré.

2.2.2 La transcription des enhancers

Deux études indépendantes de 2010 montrent que le recrutement de l'ARN Pol II et la transcription sont des caractéristiques pangénomiques des enhancers fonctionnellement actifs (Kim et al. 2010; De Santa et al. 2010). Les transcrits sont dénommés eRNAs (enhancer RNAs) et diffèrents des lncRNAs décrits précédemment. Les eRNAs sont transcrits à partir d'enhancers caractérisés par un niveau élevé de H3K4me1/2 par rapport à H3K4me3. Les eRNAs sont en général cappés (Lam et al. 2013) mais non-polyadénylés et sont donc de petite taille (<2kb). La transcription au niveau des enhancers est majoritairement bidirectionnelle sans biais contrairement aux promoteurs. Les eRNAs sont sensibles à la dégradation par les exosomes et sont transcrits à un faible niveau (pour revue : Lam et al. 2014). Leur détection par séquençage classique d'ARNs nécessite une grande profondeur de séquençage. Des

techniques ont été développées pour palier à cette limitation comme le CAGE (Cap Analysis Gene Expression) qui consiste en la capture des ARNs par la coiffe, d'une étape de transcription inverse et de PCR avant séquençage (Shiraki et al. 2003), le GRO-seq (Global Run-On sequencing) qui utilise l'incorporation de Br-UTP pour détecter les ARNs naissants suivie d'amplification et de séquençage (Core et al. 2008) ou la combinaison des deux techniques logiquement dénommée GRO-cap (Core et al. 2014). Il semblerait que l'induction des eRNAs soit un indicateur pertinent de l'activité des enhancers. Une analyse pangénomique de la transcription des enhancers de plusieurs types cellulaire chez l'humain a notamment montré une forte corrélation entre les enhancers transcrits, les marques H3K4me1/H3K27ac et les sites de sensibilité à l'ADNase I ainsi qu'une activité testée par transgénèse des candidats chez le poisson-zèbre (Andersson et al. 2014; Wu et al. 2014).

Trois hypothèses non exclusives ont été formulées pour expliquer le rôle physiologique de la transcription des enhancers :

- la transcription à bas bruit résulte de l'accessibilité des enhancers aux ARN polymérases II ;
- le processus de transcription *per se* conduit aux fonctions activatrices des enhancers ;
- ou le transcrit lui-même participe à l'activation des enhancers.

Plusieurs études ont confirmé les deux dernières hypothèses pour certains eRNAs (Melo et al. 2013; Lam et al. 2013; Li et al. 2013). Les mécanismes d'action sont encore particulièrement débattus pour les interactions entre promoteurs et enhancers. Pour certains *loci*, les eRNAs participent directement aux boucles chromatiniennes et à la déposition de H3K27ac au niveau du promoteur cible (Pnueli et al. 2015; Li et al. 2013 et fig. 5B) ou plutôt à l'activation de la transcription au locus déjà pré-conformé, en permettant notamment le relargage de la polymérase en pause (Melo et al. 2013; Schaukowitch et al. 2014 et fig. 5B). Une étude a montré que les eRNAs contribuent au remodelage de la chromatine et l'accès à la machinerie de transcription (Mousavi et al. 2013 et fig. 5B). Il a été mis en évidence récemment que le complexe Integrator, qui s'associe avec le domaine CTD de la Pol II, participe à la biogenèse des eRNAs et à la fonction des enhancers. Lorsque plusieurs sous-unités du complexe sont inactivées, la synthèse des eRNAs diminue et le contact enhancer-promoteur est aboli (Lai et al. 2015). Il semble donc que la transcription des enhancers et le transcrit agissent sur l'activation de la transcription des gènes cibles par des mécanismes différents, plutôt en *cis*, toujours en cours d'investigation, rappelant la diversité d'action des ncRNAs.

2.2.3 Les polymérases en pause

Chez les métazoaires, un grand nombre de gènes présentent des ARN polymérases II (Pol II) dites « en pause ». Cette polymérase est une forme active qui s'arrête entre 30 pb et 50 pb après le TSS. Des études pangénomiques suggèrent que les pol II en pause représentent une étape commune de régulation de la transcription des gènes de développement ou de réponse à des voies de signalisation (pour revue : Adelman & Lis 2012). Le phénomène de pause des pol II dépend du promoteur proximal auquel se fixent les TFs en association avec deux facteurs stabilisateurs : NELF (Negative Elongation Factor) et DSIF (DBR-Sensitivity-Inducing Factor). Le relargage de la pol II fait intervenir le recrutement par les TFs et co-facteurs du complexe P-TEB (Positive Transcription Elongation Factor-b) qui phosphoryle le domaine carboxyterminal de la pol II sur les serines 2 ainsi que NELF qui est évincé du promoteur et DSIF qui devient un facteur positif du processus d'élongation de la transcription (pour revue : Peterlin & Price 2006 et fig. 5D). Comme mentionné précédemment, la présence locale de eRNAs au niveau du promoteur facilite l'éviction du facteur NELF (Schaukowitch et al. 2014 et fig. 5B). Cependant, plusieurs études ont mis en évidence que les interactions à distance entre enhancers et promoteurs participent au relargage de la pol II. Le Médiateur sert de pont pour la communication enhancer-promoteur. Il interagit d'abord avec TFIID puis échange avec le complexe de super élongation (SEC), qui comprend notamment P-TEB, et facilite donc le relargage de la pol II et une élongation productive (Takahashi et al. 2011). La déplétion de la cohésine entraîne la diminution de la transcription des gènes fixés par le complexe, augmente la pause des pol II et diminue les contacts entre les enhancers et leurs promoteurs, indiquant le rôle de la cohésine dans le relargage de la pol II et l'importance des contacts enhancers-promoteurs dans l'activation de la transcription (Schaaf et al. 2013). Une étude à grande échelle chez la drosophile montre qu'une grande majorité des interactions entre enhancers et promoteurs sont présentes avant l'activation des gènes cibles et sont associées avec les polymérases en pause. Cet état de pré-conformation conditionnerait l'activation rapide et précise de la transcription par le recrutement de facteurs tissu-spécifiques et le relargage de la pol II (Ghavi-helm et al. 2014). Une étude récente montre que la sous-unité PC (ou son orthologue UBX) de PRC1 est physiquement associée avec les pol II en pause et qu'elle inhibe CBP pour l'acétylation de H3K27 (Tie et al. 2016). De plus, les interactions enhancers-promoteurs sont dépendantes de CBP chez la drosophile, suggérant que l'inhibition de CBP par PC participe à la régulation de l'expression en affectant la configuration chromatinienne (Kellner et al. 2012).

2.2.4 La transcription bidirectionnelle des promoteurs

Comme pour les enhancers, des études génomiques ont démontré que les promoteurs pouvaient être transcrits de façon bidirectionnelle. En effet, les TSS des ARNm des mammifères sont accompagnés de TSS amonts de direction opposée produisant des transcrits amonts du promoteur (PROMPTs) or ARN amonts antisens (uaRNAs) qui sont très instables, non codants et dégradés rapidement par les exosomes (Preker et al. 2008; Seila et al. 2008; Core et al. 2008; Core et al. 2014; Andersson, Chen, et al. 2015; Flynn et al. 2011). L'initiation de la transcription des ARNm et des PROMPTs/uaRNAs est caractérisée par un espacement de 110 pb à 250 bp entre les TSS, autour duquel un grand nombre de facteurs de transcription se fixent et qui est donc accompagné d'une déplétion des nucléosomes (Core et al. 2014; Duttke et al. 2015; Scruggs et al. 2015 et fig. 5E). Ces propriétés ont conduit à appeler ces promoteurs « divergents » ou « bidirectionnels ». La fonction des PROMPTs/uaRNAs n'a pas encore été élucidée. Quant à la proportion de promoteurs unidirectionnels par rapport aux bidirectionnels, un débat a opposé les auteurs de deux articles récents : les premiers qui défendent une large majorité de promoteur bidirectionnel (Andersson, Chen, et al. 2015) alors que les seconds proposent que les promoteurs sont par nature unidirectionnels (Duttke et al. 2015). Entre sémantique, techniques à haut débit, profondeur de séquençage et analyses bioinformatiques, les avis divergent et illustrent l'avancée rapide des technologies qui invite à une certaine prudence quant à l'analyse et l'interprétation de ces données à grande échelle.

Deux revues récentes pointent les similitudes entre enhancers et promoteurs aussi bien au niveau architectural que fonctionnel (Andersson, Sandelin, et al. 2015; Kim & Shiekhattar 2015). La capacité de transcription bidirectionnelle par le recrutement de la pol II et des GTFs est partagée entre ces deux éléments, de même que la synthèse efficace d'ARN sens et antisens. Ils mettent aussi en lumière la réciprocité suivante : les enhancers possèdent une activité promotrice alors que les promoteurs peuvent fonctionner comme des enhancers. Cependant, ceci sous-tend que les enhancers proximaux sont maintenant complètement rattachés à la notion de promoteur. Les modifications d'histones sont un autre aspect qu'ils rapprochent. Les trois états de méthylation de H3K4 (me1/2/3) pourraient être le reflet de la dynamique de la transcription, avec d'un côté des enhancers forts marqués par H3K4me3 et les promoteurs faibles par H3K4me1, rendant la distinction entre les deux types d'éléments encore plus floue. Bien que les propriétés allouées aux enhancers et promoteurs aient changé avec le séquençage de nombreux génomes, il n'en demeure pas moins qu'un promoteur dirige la transcription d'un gène, qu'il soit codant ou non codant, et qu'un enhancer facilite cette transcription précise de manière spatio-temporelle.

2.3 L'organisation tridimensionnelle du génome

Le noyau eucaryote est un organite délimité par une enveloppe nucléaire qui confine l'ADN génomique dans un espace tridimensionnel. Il s'organise en compartiments qui peuvent être spécifiques d'un type cellulaire et donc participer à l'expression correcte des gènes. Les avancées de techniques associées au séquençage à haut débit telles les méthodes dérivées du 3C et la microscopie à haute résolution ont révolutionné notre compréhension de l'organisation du génome et des mécanismes qui régulent l'activité des gènes. Outre les boucles enhancers et promoteurs, d'autres structures modèlent la chromatine : en domaines topologiques (TADs), en compartiments subchromosomaux actifs ou inactifs, en domaines associés à la lamina de l'enveloppe nucléaire (LADs) ou en territoires chromosomiques.

2.3.1 Les domaines topologiques (TADs)

Les domaines topologiques sont des domaines de l'ordre du Mégabase (Mb) chez les mammifères caractérisés par une plus forte proportion d'interactions avec les *loci* d'un même domaine qu'avec ceux des domaines adjacents. Ils ont été mis en évidence par les techniques de Hi-C et de 5C qui permettent d'obtenir une carte des interactions génomiques, dont la résolution dépend de la profondeur de séquençage entre 10 et 100 kb, chez les mammifères et la drosophile (Nora et al. 2012; Dixon et al. 2012; Sexton et al. 2012 et fig. 6A). Les TADs sont globalement invariants au cours de la différenciation cellulaire et ils sont généralement conservés en position entre l'homme et la souris (Dixon et al. 2012; Vietri Rudan et al. 2015). Cependant ils sont divisés en sous-domaines d'interactions (aussi appelés sub-TADs) qui varient fortement d'un type cellulaire à un autre et correspondraient à différentes conformations entre éléments *cis*-régulateurs et promoteurs (Dixon et al. 2012; Nora et al. 2012; Phillips-Cremins et al. 2013; Rao et al. 2014 et fig. 6A). Les frontières des TADs sont enrichies en gènes de ménage, gènes d'ARNt, séquences répétées du type SINE et caractérisées par la présence de CTCF et de la cohésine chez les mammifères (Dixon et al. 2012). Chez la drosophile, les frontières comprennent des gènes constitutivement actifs et une forte proportion de protéines architecturales telles CTCF, BEAF-32, Su(Hw) et Mod(mdg4) (Sexton et al. 2012; Hou et al. 2012). Plusieurs études ont montré que les deux frontières d'un TAD contiennent des sites CTCF orientés de façon opposée et sont engagées ensemble dans une interaction à distance (Wit et al. 2015; Gómez-Marín et al. 2015; Guo et al. 2015; Rao et al. 2014; Vietri Rudan et al. 2015 et fig. 6A). Chez les mammifères, 85% des sites de fixation du facteur CTCF sont localisés à l'intérieur des TADs (Dixon et al. 2012) où ces sites favorisent les interactions enhancers-promoteurs dépendantes de CTCF (Tang et al. 2015).

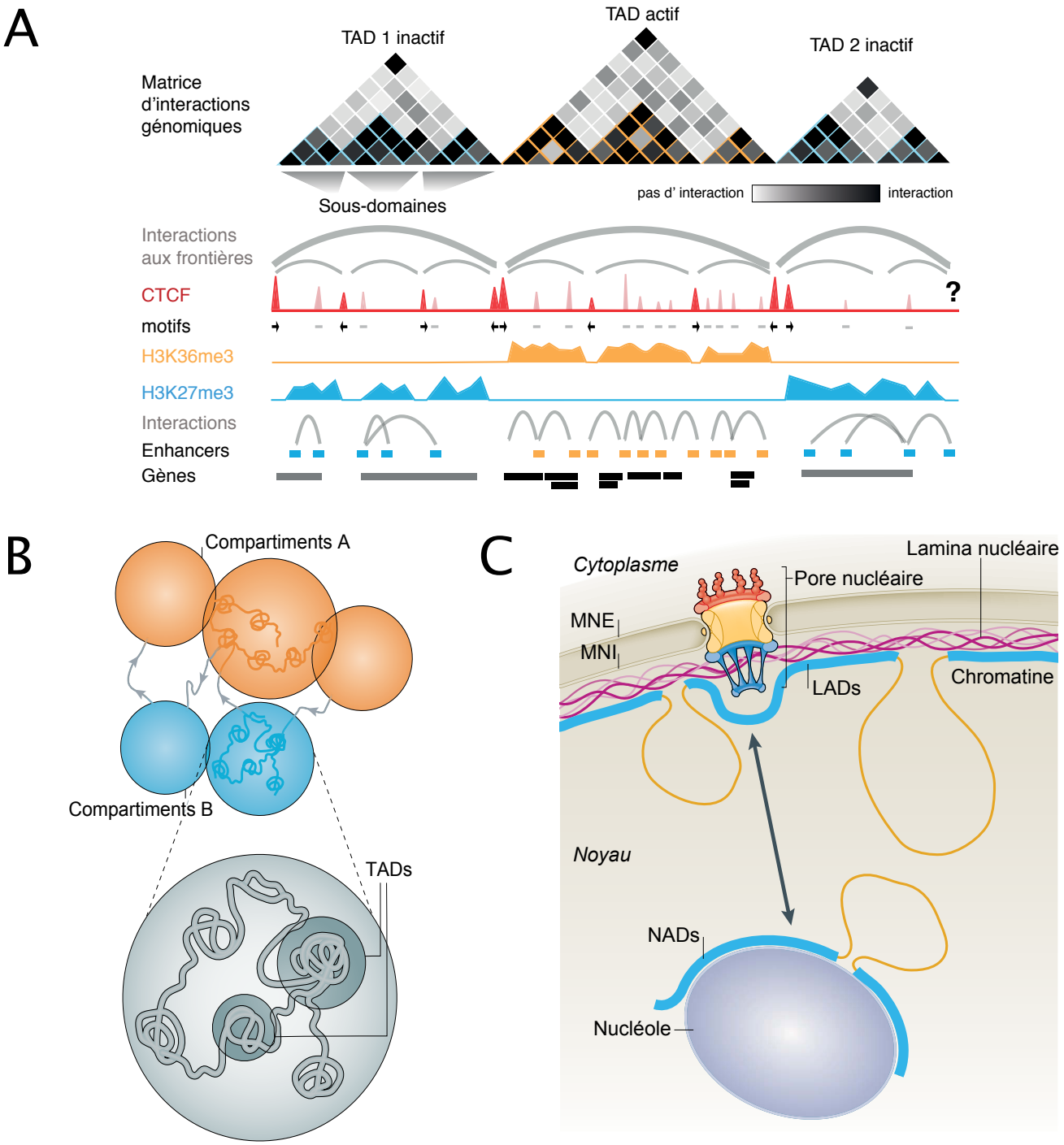


Figure 6 : L'organisation tridimensionnelle de la chromatine au sein du noyau

A. Les domaines topologiques (TADs) délimitent des zones d'interaction chromatiniennes séparées par des insulateurs. La matrice correspond à un Hi-C virtuel où les fréquences d'interaction sont représentées par des nuances de gris. Les TADs sont majoritairement délimités par des sites CTCF et sont divisés en sous-domaines d'interaction entre enhancers et promoteurs et sites CTCF dont l'orientation est indiquée par des flèches. Les TADs correspondent parfois à des domaines chromatiniens actifs (H3K36me3), riches en gènes et inactifs (H3K27me3), plus pauvres en gènes. Le point d'interrogation suggère que d'autres facteurs que CTCF peuvent participer aux interactions à distance.
B. Le génome est divisé en compartiments A (orange) et B (bleu) d'environ 5Mb qui s'associent préférentiellement entre eux au sein du noyau. Ces compartiments contiennent plusieurs TADs transcriptionnellement actifs (A) ou inactifs (B).
C. Les domaines associés à la lamina nucléaire (LADs) sont des régions chromatiniennes condensées accrochées à la lamina nucléaire alors que les régions ouvertes pointent vers l'intérieur du noyau. Certains LADs peuvent délocaliser autour du nucléole formant les domaines associés au nucléole (NADs). Le noyau est entouré d'une enveloppe nucléaire composée d'une membrane nucléaire externe (MNE) et interne (MNI) recouverte de lamines, et comprenant les pores nucléaires permettant la communication entre le noyau et le cytoplasme.
D'après Ali et al., 2016 (A), Dekker et al., 2013 (B) et Pombo et al., 2015 (C).

Plusieurs études ont visé à déterminer le rôle de la cohésine et de CTCF dans la formation et le maintien de cette structure. L'éviction de ces complexes entraînent un affaiblissement des frontières des TADs avec une augmentation des interactions entre éléments *cis*-régulateurs et promoteurs de différents TADs, quelques altérations au niveau de l'expression de gènes mais les TADs sont globalement conservés en forme (Seitan et al. 2013; Zuin et al. 2014; Phillips-Cremins et al. 2013; Narendra et al. 2015). Ces résultats suggèrent que la cohésine et CTCF participent à ce niveau d'organisation de la chromatine mais que d'autres facteurs ou mécanismes sont impliqués. Il est à présent possible de réconcilier les différentes fonctions du facteur CTCF au regard de son rôle de structure dans les frontières des TADs mais aussi fonctionnel au sein des TADs dans la communication entre éléments *cis*-régulateurs.

Plusieurs études suggèrent qu'il y aurait une spécificité limitée des interactions enhancer-promoteur et que le facteur déterminant quel enhancer régule quel gène serait leur localisation commune dans le même domaine chromosomique (Schwarzer & Spitz 2014). L'analyse de l'expression des gènes au cours de la différenciation des cellules ES en progéniteurs neuronaux montre que l'expression des gènes localisés dans le même TAD est plus corrélée que celle entre gènes de TADs adjacents (Nora et al. 2012). Une autre approche consiste à utiliser l'insertion aléatoire et non biaisée d'un transposon servant de capteur de l'influence des éléments *cis*-régulateurs agissant dans cette région à travers tout le génome. Une fois le site d'insertion du rapporteur déterminé, les auteurs ont observé que ceux situés dans le même domaine avaient une expression tissu-spécifique importante et que ces domaines corrèlent avec les TADs (Symmons et al. 2014). Un cas particulier est apporté par le groupe des gènes *HoxD* au cours du développement du membre chez les tétrapodes. En effet, les gènes sont à l'interface entre deux domaines topologiques et leur capacité à répondre à ces deux types de régulation dépend des contacts physiques avec les enhancers correspondants et leur position dans le groupe. Cette régulation bimodale permet le développement séquentiel du bras et de la main sous le contrôle de la colinéarité d'expression des gènes *HoxD* (Montavon et al. 2011; Andrey et al. 2013).

Les TADs sont des unités chromatiniennes actives avec la présence des marques H3K27ac et H3K36me3 ou inactives avec H3K27me3 et H3K9me3 correspondant parfois à des LADs (cf 2.3.3 et fig. 6A), et souvent corrélées au niveau d'expression des gènes au sein de ces TADs (Nora et al. 2012; Rao et al. 2014; Le Dily et al. 2014; Dixon et al. 2015). Ces états chromatiniens ne sont pas stables et peuvent être modulés en corrélation avec l'activation ou la répression des gènes notamment après l'induction d'hormones (Le Dily et al. 2014). Les TADs sont donc des domaines fonctionnels délimités par des frontières qui

empêchent la communication entre les enhancers et les gènes situés dans un TAD adjacent en association avec un état chromatinien actif ou inactif.

Un modèle d'organisation des chromosomes en interphase essaye de réconcilier les différents aspects de la communication entre éléments *cis*-régulateurs au sein des TADs évoqués précédemment. Ce modèle « d'exclusion de boucle » propose la formation active de boucles par un facteur de type cohésine capable d'exclure la chromatine de façon à créer des boucles de plus en plus larges en fixant un site CTCF à la frontière du TADs jusqu'à atteindre la suivante (Nichols & Corces 2015; Bouwman & de Laat 2015; Sanborn et al. 2015; Fudenberg et al. 2016). Une autre possibilité implique un complexe actif qui contacte et scanne l'ADN de façon directionnelle et ayant la capacité de former des boucles de chromatine. Le premier candidat est le complexe cohésine qui chez la bactérie et la levure peut glisser le long de l'ADN et condenser les chromosomes en formant des boucles (Kim & Loparo 2016; Ocampo-Hafalla et al. 2016). Le second est l'ARN polymérase II : il a été montré que des polymérases actives peuvent exclure un gène en formant une boucle (Larkin et al. 2013). Une autre étude a mis en évidence qu'au cours de la transcription, l'ARN polymérase II transcrivant peut rester associée avec le complexe enhancer-promoteur formant ainsi une boucle active avec le corps du gène (Lee et al. 2015). De plus, deux études suggèrent que les boucles enhancers-promoteurs sont directionnelles et que les promoteurs contactent préférentiellement les enhancers en amont, ceci en accord avec les deux modèles proposés ci-avant (Sanyal et al. 2012; Tang et al. 2015). La formation et la fonction des TADs restent encore peu explorées avec différents modèles à confirmer et des acteurs majeurs à identifier qui viendront renforcer notre compréhension de l'organisation du génome.

2.3.2 Les domaines A&B et les usines de transcription

Il a été largement montré que les zones ouvertes, riches en gènes et les domaines chromosomiques fermés, plus pauvres en gènes, sont spatialement séparés au sein du noyau (Fraser & Bickmore 2007). La technique de Hi-C a permis de révéler la présence de compartiments chromosomiques d'environ 5 Mb appelés A et B (Lieberman-Aiden et al. 2009; Zhang et al. 2012 et fig. 6B). Les *loci* regroupés dans le compartiment A sont généralement riches en gènes, actifs transcriptionnellement et sensibles à l'ADNase I (euchromatine) alors que ceux du compartiment B sont relativement pauvres en gènes, inactifs et insensibles à l'ADNaseI (hétérochromatine). Les TADs dans un même état chromatinien ont tendance à s'associer en domaines A et B qu'ils soient situés en *cis* ou en *trans*, de même que les compartiments A actifs et B inactifs interagissent préférentiellement

dans un schéma « qui se ressemble s'assemble » (Gibcus & Dekker 2013; Rao et al. 2014; Dekker & Heard 2015). Une étude récente utilise le recrutement mécanique de la méthyltransférase SuV39H1 (qui dépose H3K9me3) qui au niveau d'un sub-TADs d'un compartiment A, permet l'engagement de ce dernier dans un compartiment B. Cependant cette relocalisation nucléaire varie en fonction des loci ciblés, suggérant que l'état chromatinien n'est pas le seul déterminant (Wijchers et al. 2016). Ces résultats rejoignent l'observation que les TADs sont des blocs de structures invariants et que les différences entre cellules seraient le résultat de la relocalisation d'un TAD complet d'un compartiment à l'autre en fonction de la fréquence d'interactions intra-chromosomiques et des changements d'activité des gènes (Dekker & Heard 2015; Dixon et al. 2015).

Les caractéristiques du compartiment A ressemblent étonnamment à celles décrites pour les usines de transcription ('transcription factories') comme des regroupements de gènes actifs transcriptionnellement, parfois localisés sur des chromosomes différents, et qui sont associés en foci riches en ARN polymérases II phosphorylées et autres régulateurs transcriptionnels (Osborne et al. 2004). Une étude sur des érythrocytes de souris a montré que les gènes de la globine répondant au facteur Klf1 et situés sur des chromosomes différents co-localisent avec ce même facteur dans des usines de transcription en adaptant la technique de 4C avec du ChIP contre l'ARN polymérase II (Schoenfelder et al. 2010). De façon analogue, des expériences de 3C au niveau du promoteur du gène SAMDA4 ont mis en évidence le recrutement d'autres promoteurs de gènes induits par TNFα également transcrits (Papantonis et al. 2010). Plus récemment, deux études chez la souris et l'homme utilisent du Capture Hi-C qui capture les promoteurs et montrent que les promoteurs de gènes co-régulés interagissent de façon non aléatoire et corrélée à leur fonction biologique et leur niveau d'expression (Mifsud et al. 2015; Schoenfelder et al. 2015). Ces résultats suggèrent aussi que les interactions promoteurs-promoteurs au niveau des usines de transcription sont le reflet de la communication enhancer-promoteur qui concentrent les facteurs de transcription et l'ARN polymérase II.

2.3.3 Les territoires chromosomiques et les LADs

À la fin du XIXème siècle, Carl Rabl propose que les chromosomes en interphase conservent leur individualité et occupent des territoires distincts dans le noyau (Rabl 1885). Ce modèle s'oppose au modèle spaghetti où les chromosomes sont mélangés de façon aléatoire. En 1982, les frères Cremer mettent en évidence de façon indirecte l'existence des territoires chromosomiques à l'aide d'un laser UV qui induit des dommages à l'ADN

seulement dans une petite partie du noyau, et révèlent que peu de chromosomes sont marqués (Cremer et al. 1982). La technique de FISH (Fluorescent In Situ Hybridization) a permis de 'peindre' les chromosomes dans le noyau et de démontrer directement l'organisation des chromosomes en territoires (Pinkel et al. 1988; Lichter et al. 1988). La distribution radiale des chromosomes ne semble pas être aléatoire non plus comme le suggèrent des expériences de FISH à haute résolution avec notamment des régions denses en gènes (chromosomes 17, 19 et 20) au centre du noyau alors que celles plus pauvres (chromosomes 18, Y) se retrouvent à la périphérie (Croft et al. 1999; Cremer et al. 2001). Cette distribution est conservée chez les primates pour les régions orthologues du chromosome 19 et celles du chromosome 18 (Tanabe et al. 2002). L'introduction de la technique de Hi-C a permis de confirmer cette organisation des chromosomes en territoires où différents loci interagissent plus en *cis* qu'en *trans*, la proximité des régions chromosomiques riches en gènes et la ségrégation entre des zones ouvertes et fermées de chromatine (Lieberman-Aiden et al. 2009; Zhang et al. 2012). Les chromosomes voisins peuvent tout de même se mélanger, des boucles chromatiniennes d'un territoire contactant celles d'un autre territoire. Ces régions de mélange sont corrélées à des zones de translocations des chromosomes mais aussi à des usines de transcription, mettant en avant le rôle structural et fonctionnel de ce niveau d'organisation (Branco & Pombo 2006).

La membrane interne du noyau est recouverte de protéines, majoritairement des lamines, formant la lamina nucléaire. Les lamines appartiennent à la famille des filaments intermédiaires et maintiennent la forme et les propriétés mécaniques du noyau. La lamina nucléaire offre une large surface qui semble être une plateforme d'ancrage pour les chromosomes (van Steensel & Dekker 2010 et fig. 6C). Pour cartographier les contacts entre le génome et la lamine nucléaire deux techniques ont été utilisées : du ChIP-seq et de la DamID (DNA adenine methyltransferase identification) qui consiste à utiliser une protéine de fusion entre une lamine et une adénine méthyltransférase (Dam) qui s'incorpore dans la lamina nucléaire si bien que toute séquence génomique en contact avec la lamina sera méthylée *in vivo* et révélée par digestion enzymatique et amplification par PCR (Guelen et al. 2008). Ces techniques ont révélé chez les mammifères entre 1100 et 1400 domaines associés à la lamina nucléaire (LADs) variant de 10 kb à 10 Mb, représentant presque 40% du génome et fortement enrichis en marques répressives H3K27me3 et H3K9me2/3 et relativement pauvres en gènes associés à une faible activité transcriptionnelle (Pickersgill et al. 2006; Guelen et al. 2008; Amendola & Van Steensel 2014). Deux types de LADs existent. Les LADs constitutifs sont riches en A et T, stables parmi un grand nombre de types cellulaires et très conservés entre l'homme et la souris. Les LADs facultatifs, quant à eux, varient d'une cellule à l'autre et peuvent notamment être transféré au niveau du nucléole après mitose

formant les domaines chromatiniens associés au nucléole ou NADs (Meuleman et al. 2013; Kind et al. 2013 et fig. 6C). La perte de contacts de certains LADs avec la lamina nucléaire au cours de la différenciation est majoritairement associée à l'activation des gènes même si l'activation peut être plus tardive et les contacts *de novo* sont reliés à l'inactivation (Peric-Hupkes et al. 2010) ; en accord avec une réduction de l'expression de transgènes adressés artificiellement à la lamina nucléaire (Finlan et al. 2008). Les LADs correspondent à des TADs inactifs au sein de compartiments B à la périphérie du noyau au contraire des compartiments A occupant une place nucléaire plus centrale (Dixon et al. 2012; Dixon et al. 2015; Vieux-Rochas et al. 2015). Les LADs peuvent servir de squelette structural aux chromosomes interphasiques et influencent l'expression des gènes.

Nous avons vu que les gènes sont régulés par de nombreux mécanismes faisant intervenir des modifications de la chromatine, la transcription *per se*, et une conformation spatiale de l'ADN particulière. La fixation des facteurs de transcription sur leurs séquences régulatrices reste déterminante pour le contrôle spatio-temporel de l'expression des gènes. La troisième partie sera consacrée à la place centrale de ces éléments *cis*-régulateurs dans le développement, au cours de l'évolution et dans les maladies.

3. Les éléments *cis*-régulateurs au cœur des processus développementaux, évolutifs et des maladies

L'homme et la souris possèdent pratiquement les mêmes gènes mais sont phénotypiquement très différents l'un de l'autre. Le développement d'un organisme repose sur l'activation spatio-temporelle précise de ces gènes dont les combinaisons sont différentes d'une espèce à l'autre. La régulation de l'expression de ces gènes est cruciale et repose notamment sur les éléments en *cis* qui représentent environ 30 % du génome chez les mammifères. Des modifications des séquences régulatrices modulent le patron d'expression des gènes dans le temps et l'espace alors que les organismes évoluent. Ces modifications peuvent aussi être à l'origine de dérèglements temporisés ou non par l'organisme. Ce chapitre sera consacré au rôle des éléments *cis*-régulateurs au cours du développement, à l'évolution des séquences régulatrices à l'origine de la diversité animale, ainsi que les variations génétiques de ces séquences associées à des maladies.

3.1 Les éléments *cis*-régulateurs confèrent de la robustesse pendant le développement

Au cours du développement, des mécanismes de contrôle transcriptionnel assurent le patron d'expression stable et reproductible des gènes de l'embryon en dépit des fluctuations inhérentes liées aux processus de régulation, des variations du fond génétique et de l'exposition à l'environnement. Cette robustesse d'expression est notamment le fruit de la redondance liée aux éléments *cis*-régulateurs mais aussi à l'affinité et la syntaxe des sites de fixation des TFs qu'ils contiennent.

3.1.1 Redondance et précision transcriptionnelle

La précision transcriptionnelle comprend la formation de frontières nettes d'expression des gènes, le timing exact de l'activation des gènes, la coordination de l'expression d'un groupe de gènes au sein du tissu en développement et l'expression homogène dudit gène dans les cellules d'un même domaine. La redondance est un des mécanismes clefs de la précision transcriptionnelle. En plus de la redondance d'interactions génétiques, par exemple lorsque deux TFs reliés contrôlent les mêmes gènes en aval dans un tissu sans que les deux soient absolument nécessaires, la redondance d'éléments *cis*-régulateurs, particulièrement pour les enhancers, est fréquente. Les enhancers contiennent

plusieurs sites de fixation pour des TFs qui peuvent être redondants dans la mesure où des mutations au sein de certains sites n'altèrent pas le niveau d'expression du gène contrôlé par cet élément modifié. C'est le cas pour l'enhancer contrôlant *eve* dans la bande 2 (*eve* stripe 2) chez la drosophile qui contient un enhancer minimal et une extension contenant une répétition des mêmes sites de fixation. La transgénèse d'un BAC contenant l'enhancer et le gène sauvage dans un mutant *eve-/eve-* sauve le phénotype létal. Dans le cas de l'extension seule, les embryons ne survivent pas alors que dans sa version minimum de 500 pb au lieu de 700 pb, l'embryogenèse est normale dans des conditions optimales, mais la viabilité des embryons est réduite dans une situation de stress (température élévée ou fond génétique sensible, Ludwig et al. 2011 et fig. 7A). Cette redondance des sites de fixation de l'enhancer assure l'expression stable de *eve* dans la bande 2 dans des conditions de stress. Comme mentionné précédemment, de nombreux gènes cruciaux du développement sont régulés par plusieurs enhancers qui dirigent des patrons d'expression chevauchants selon une même logique transcriptionnelle et appelés « shadow enhancers » (Lagha et al. 2012 et fig. 7B). De nombreux exemples viennent d'études chez la drosophile grâce la simplicité des manipulations génétiques de ce modèle et du développement externe des embryons, cependant quelques cas de redondance d'enhancers chez les Vertébrés ont été rapportés. De façon intéressante, les deux enhancers du gène de la pro-opiomélanocortine (*Pomc*) sont redondants à l'âge adulte, contrôlant le niveau d'expression du gène, alors qu'ils agissent de façon synergique pendant l'embryogenèse chez la souris (Lam et al. 2015). En plus de conférer de la robustesse aux perturbations de l'environnement (Frankel et al. 2010), les shadow enhancers sont capables de tamponner les mutations (Cannavò et al. 2016) et aident à la formation de frontières franches d'expression de gène en réponse à de faibles changements du niveau des TFs (Perry et al. 2011 et fig. 7C).

3.1.2 Les sites de faible affinité aussi importants que les sites de forte affinité

La plupart des études sur les TFs se sont concentrées sur les sites de fixation de plus forte affinité. Cependant, de plus en plus d'études mettent en lumière l'importance des sites de fixation d'affinités variées, y compris ceux de faible affinité, pour la régulation de l'expression des gènes pendant le développement. L'information spatio-temporelle peut être modifiée par des variations de l'affinité des sites de fixation. Par exemple, un site de faible affinité pour PAX6 dans l'enhancer du gène murin *Neurogenin2* est requis pour restreindre les domaines d'expression à de forts niveaux de PAX6 dans la moelle épinière et le cortex. Le remplacement de ce site par un site de forte affinité entraîne une expression ectopique du

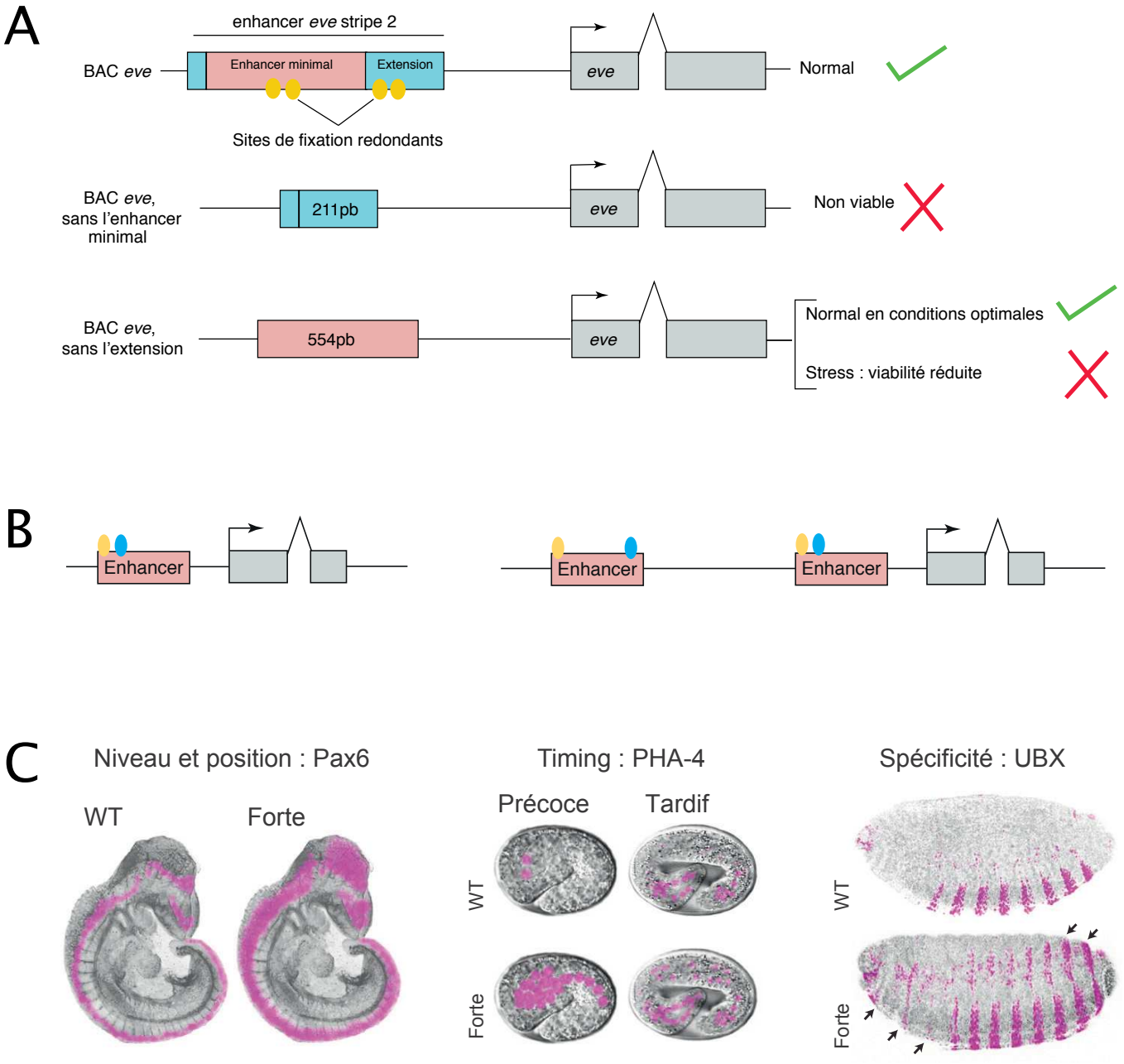


Figure 7 : Redondance et sites de faibles affinités au cours du développement

A. La redondance des sites de fixation d'un enhancer confère de la robustesse. L'enhancer stripe 2 du gène *eve* contient un enhancer minimal de 554pb (rose) et une extension de 211pb (bleue). En absence de l'enhancer minimal, les embryons ne sont plus viables. En revanche, l'enhancer minimal est suffisant pour le développement des embryons en conditions optimales mais la viabilité est réduite dans des conditions de stress. Les sites de fixation redondants sont en jaune.

B. Les shadow enhancers possèdent les mêmes sites de fixation que l'enhancer initial et assurent la robustesse d'expression. Les sites de fixation sont en bleu et jaune.

C. Les sites de faible affinité sont importants pour le développement. De gauche à droite : le remplacement des sites de fixation de faible affinité de PAX6 pour des sites de forte affinité conduit à l'expression ectopique du rapporteur sous contrôle de l'enhancer du gène *Neurogenin2* chez la souris ; le remplacement des sites de fixation de faible affinité de PHA-4 pour des sites de forte affinité entraîne l'expression précoce du rapporteur chez *C. elegans* ; le remplacement des sites de fixation de faible affinité de Ultrabithorax (UBX) pour des sites de forte affinité dans un enhancer de *shavenbaby* conduit à l'expression ectopique du rapporteur dans les régions antérieures de l'embryon de drosophile. D'après Lagha et al.,2012 (A & B) et Crocker et al., 2016 (C).

rapporteur (Scardigli et al. 2003 et fig. 7C). Les variations d'affinité des sites de fixation jouent aussi un rôle déterminant dans le timing de l'expression des gènes. Chez *C. elegans* le TF PHA-4 contrôle différents gènes cibles grâce à des sites d'affinités variables. Les modifications apportées à ces sites ont révélé que les sites de forte affinité dirigent l'expression des gènes précocement pendant le développement lorsque le niveau de PHA-4 est faible alors que c'est l'inverse pour les sites de faible affinité qui sont importants pour le développement tardif, lorsque le niveau de PHA-4 est fort (Gaudet & Mango 2002 et fig. 7C). Enfin les sites de faible affinité peuvent faciliter les interactions précises et spécifiques des TFs. Il a été montré chez la drosophile que des sites de faibles affinités pour le TF à homéodomaine Ultrabithorax (UBX) dans l'enhancer de *shavenbaby* confèrent la spécificité de fixation de UBX *in vivo*. En effet, le remplacement de ces sites par des sites de forte affinité entraîne l'expression ectopique dans l'embryon où UBX n'est pas exprimé, et ceci probablement suite à la fixation d'autres TFs à homéodomaine. De plus, le regroupement de ces sites de faible affinité pour UBX apporte de la robustesse d'expression puisque la mutation d'un seul site perturbe le développement dans des conditions de stress (Crocker et al. 2015 et fig. 7C).

Deux études récentes du même groupe mettent en évidence l'importance de ces sites de faible affinité grâce à une analyse à grande échelle de l'enhancer du gène *Otx2* chez l'ascidie (Farley et al. 2015; Farley et al. 2016). L'enhancer fonctionne en « sub-optimisation », c'est-à-dire que les sites de faibles affinités sont organisés en syntaxe sub-optimale qui lui confère sa spécificité. L'optimisation soit de l'affinité soit de la syntaxe conduit à une baisse de la spécificité observée par l'expression ectopique du rapporteur dans des tissus inappropriés (Farley et al. 2015). Les auteurs définissent un « code de régulation » basé sur la syntaxe des sites de fixation qu'ils testent *in silico* pour la recherche d'enhancers. Ils observent alors une corrélation entre sites de faibles affinités et syntaxe optimale et à l'inverse entre sites de fortes affinités et syntaxe sub-optimale (Farley et al. 2016). Il reste maintenant à découvrir si ce « code de régulation » s'applique à d'autres enhancers au cours du développement et chez d'autres espèces, mais c'est une piste prometteuse pour la recherche *in silico* d'enhancers.

Plusieurs hypothèses ont été avancées pour expliquer l'abondance de ces sites de faibles affinités. Tout d'abord, la spécificité des enhancers peut être augmentée en diminuant l'affinité des sites de fixation et ainsi la diversité des facteurs qui s'y fixent. La coopérativité de fixation des TFs peut être augmentée par la multitude de ces sites de fixation, de même que la robustesse. Enfin l'utilisation des sites de faible affinité peut définir la limite en affinité et spécificité comme pour les TFs à homéodomaine (pour revue : Crocker et al. 2016).

Les sites de faible affinité pourraient être caractéristiques de la plupart des enhancers car ces sites peuvent évoluer grâce à moins de mutations que les sites de forte affinité (Ramos & Barolo 2013). Cette observation rend cependant difficile l'identification *in silico* de sites de fixation pour des TFs qui est un des challenges actuels de la bio-informatique.

3.2 Les éléments *cis*-régulateurs moteur de l'évolution

Les mutations sont stochastiques et affectent le génome de façon aléatoire. Au sein d'une population, une mutation affectant l'un des individus sera perdue par sélection négative alors qu'elle peut être soit perdue soit fixée dans la population par sélection neutre. Si cette mutation est sous sélection positive, elle sera fixée plus rapidement dans la population (fig. 8A). La plupart des séquences sous sélection chez les mammifères sont en dehors des exons et la grande majorité sont des éléments régulateurs transcriptionnels. De plus en plus d'exemples de modifications sont associées à la régulation des gènes où le gain ou la perte de sites de fixation pour les TFs sont responsables de la divergence des patrons d'expression des gènes et l'altération des phénotypes. Ceci est en accord avec l'idée que des modifications de régulation transcriptionnelle sont un des mécanismes prédominant de l'évolution. Cependant il est aussi commun d'observer des divergences dans la séquence des éléments *cis*-régulateurs alors que leur fonction reste largement identique et inversement.

3.2.1 Évolution des enhancers et diversité animale

Un des mécanismes drastiques d'évolution des enhancers est la simple suppression qui entraîne la perte d'expression d'un gène dans une région particulière du corps. Un exemple célèbre concerne un enhancer du gène *Pitx1* chez le stickelback à trois épines (ou épinoche) qui vit dans les eaux douces des lacs nordiques. Le stickelback marin possède une nageoire pelvienne épineuse pour se protéger des prédateurs alors que celui d'eau douce en est dépourvu. Le développement de cette épine pelvienne dépend de l'activité du TF Pitx1 qui est exprimé dans la région pelvienne sous contrôle d'un enhancer spécifique chez le stickelback marin qui est absent chez celui d'eau douce. La réintroduction de l'enhancer en amont de *Pitx1* par transgénèse chez l'animal d'eau douce permet le développement d'une épine pelvienne. Il est intéressant de noter que la délétion de cet enhancer a eu lieu plusieurs fois de manière indépendante dans des populations isolées de stickelbacks d'eau douce (Shapiro et al. 2004; Chan et al. 2010 et fig. 8B).

Un autre moyen d'évolution d'enhancers est par mutations ponctuelles ou petites insertions-délétions qui peuvent créer ou éliminer des sites de fixation pour des TF au sein de

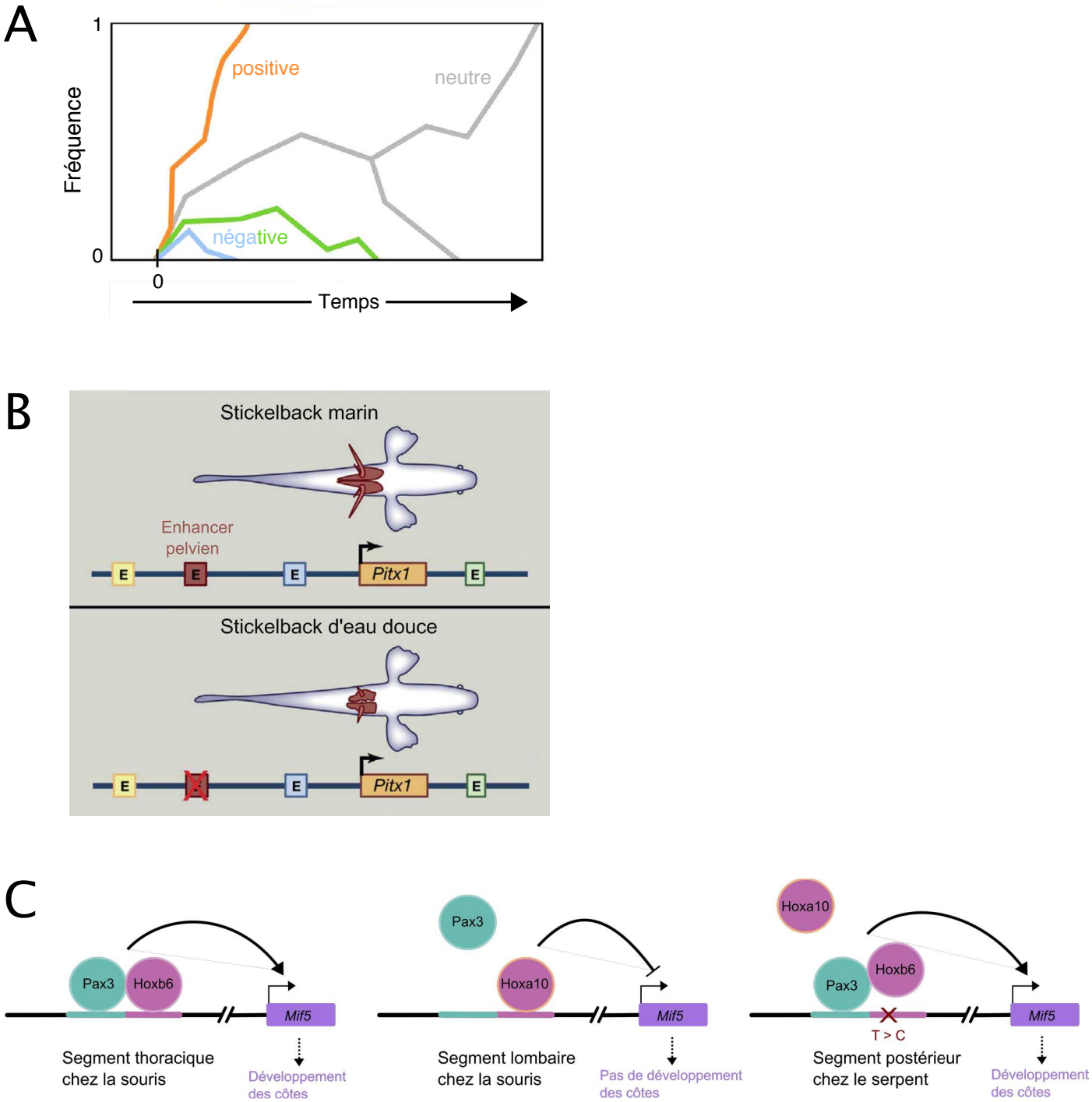


Figure 8 : Évolution des séquences régulatrices

A. Fréquence de mutations au sein d'une population en fonction du temps rencontrant une sélection neutre (courbe grise), faiblement négative (courbe verte), fortement négative (courbe bleue) ou positive (courbe orange). Une mutation chez un individu de la population apparaît au temps zéro.

B. Evolution de la nageoire pelvienne chez le stickelback. Le gène *Pitx1* est régulé par plusieurs enhancers dans différents tissus. La délétion de l'enhancer pelvien conduit à l'absence d'épine pelvienne chez le stickelback d'eau douce contrairement au stickelback marin.

C. Modèle de formation des côtes chez la souris et le serpent. Au niveau des segments thoraciques chez la souris, le gène *Myf5* est co-régulé par Pax3 et Hoxb6 pour le développement des côtes contrairement au segment lombaire où *Mif5* est réprimé par Hoxa10 qui conduit à l'absence des côtes à cet endroit. Chez le serpent, un polymorphisme dans la séquence empêche la fixation de Hoxa10 alors que Hoxa6 peut fixer l'enhancer grâce à sa co-fixation avec Pax3, entraînant la formation des côtes dans tout le segment postérieur malgré l'expression de *Hox10* dans le tissu.

D'après Siepel & Arbiza, 2014 (A), Shapiro et al., 2004 (B) et Mansfield, 2013 (C).

l'enhancer et modifier son activité. La perte d'activité peut être liée à l'accumulation de mutations ponctuelles de faible effet pour éventuellement empêcher les effets pléiotropiques causés par la perte complète de l'enhancer. Un exemple intéressant est apporté par l'étude des enhancers du gène *shavenbaby* chez l'espèce *Drosophila sechellia*. Cinq enhancers contrôlent l'expression de ce facteur de transcription qui gouverne la morphogenèse des trichomes. La fonction d'un de ces enhancers a évolué par de multiples substitutions uniques qui altèrent le niveau et le timing d'expression de *shavenbaby* par rapport à l'espèce *Drosophila melanogaster* (Frankel et al. 2011). De façon plus drastique chez les Vertébrés un polymorphisme dans l'enhancer du gène *Myf5* contribuerait à l'allongement de la cage thoracique chez le serpent comparé à la souris. En effet, au niveau des segments thoraciques chez la souris, le gène *Myf5* est co-régulé par Pax3 et Hoxb6 pour le développement des côtes. Au contraire dans le segment lombaire, *Mif5* est réprimé par Hoxa10 conduisant à l'absence des côtes à cet endroit. Chez le serpent, un polymorphisme dans la séquence empêche la fixation de Hoxa10 alors que Hoxa6 peut fixer l'enhancer grâce à sa co-fixation avec Pax3, entraînant la formation des côtes dans tout le segment postérieur malgré l'expression de *Hoxa10* (Guerreiro et al. 2013 et fig. 8C). Ces résultats suggèrent que le polymorphisme de l'enhancer répondant aux Hox et à Pax3 joue un rôle dans l'évolution de la colonne vertébrale en modulant la réponse aux Hox répresseurs et activateurs du développement des côtes.

Au cours de l'évolution, de nouvelles activités régulatrices proviennent de modifications d'enhancers existants mais aussi grâce à l'apparition de nouveaux enhancers. La distribution des CNE (Conserved Non-coding Elements) suggère que les nouveaux enhancers jouent un rôle dans plusieurs étapes évolutives des vertébrés (Lowe et al. 2011), au même titre que la co-option de transposons en nouveaux éléments *cis*-régulateurs associés à l'innovation reproductive des mammifères placentaires (Lynch et al. 2011). Une étude originale tire parti de l'utilisation des iPSCs (induced Pluripotent Stem Cells) en dérivant des cellules de crêtes neurales de l'homme et du chimpanzé pour comparer leur profil épigénétique et identifier les divergences d'éléments *cis*-régulateurs. L'implication de ces enhancers dans le développement cranio-facial est confirmée par transgénèse chez la souris. Les éléments *cis*-régulateurs spécifiques d'une espèce sont moins conservés que ceux actifs dans les deux espèces avec des variations dans les sites de fixations des TFs et sont enrichis en rétro-transposons. Ces différences de régulation seraient à l'origine de la divergence de la morphologie cranio-faciale entre l'homme et le chimpanzé (Prescott et al. 2015).

3.2.2 Conservation d'expression sans conservation de séquences et *vice versa*

Il y a deux grands types de mécanismes permettant à une région régulatrice de changer de séquence sans affecter la régulation transcriptionnelle du gène sous son contrôle. Le premier est le changement du facteur en *trans* pour compenser l'altération de la séquence d'ADN. Le second est le réarrangement de la séquence en *cis* dans une configuration différente mais fonctionnellement équivalente. Le premier mécanisme est beaucoup plus lourd de conséquences puisqu'il va affecter tout un ensemble de gènes. Le second mécanisme est majoritaire mais est difficile à mettre en évidence. Comme vu précédemment il y a de nombreux exemples de gain et perte d'enhancers ('turnover') qui peuvent être graduels, mais aussi de relocalisation ou inversion d'enhancers ('shuffling').

De nombreuses études mettent en évidence des conservations d'activité d'enhancers sans conservations de séquence, souvent par transgénèse ou par sauvetage génétique, par exemple entre différentes espèces de drosophile (Ludwig et al. 2005), entre le poisson-zèbre et l'homme (Fisher et al. 2006) ou entre la cione et l'oursin (Oda-Ishii et al. 2005). Une étude de génomique comparative utilise les marques de chromatines activatrices (H3K4me3 et H3K27ac) pour comparer l'évolution des promoteurs et enhancers dans le foie de vingt mammifères (Villar et al. 2015). De façon surprenante, il n'y pas de différence de fréquences entre le gain ou la perte d'enhancer et promoteur. La plupart des enhancers ont évolué rapidement, contrairement aux promoteurs, et la plupart sont le résultat de l'exaptation de séquences d'ADN existantes. Malgré le peu de conservation des enhancers, les 2/3 ont conservé leur activité parmi les dix mammifères avec la meilleure qualité d'alignement du génome (Villar et al. 2015).

Trois séries de duplication complète du génome sont survenues au cours de l'évolution des Vertébrés, deux séries à la base des Vertébrés et une troisième au niveau de l'embranchement des téléostéens. Chez le poisson-zèbre, une partie des gènes dupliqués ont été conservés au sein du génome ainsi que de nombreux éléments *cis*-régulateurs. Une étude s'est concentrée sur ces éléments dupliqués très conservés jusqu'aux mammifères et leur impact dans la régulation de la famille des gènes *Pax*. Ces éléments présentent beaucoup plus de différences d'activité que de ressemblances par transgénèse et de petites mutations sont suffisantes pour la divergence d'activité (Goode et al. 2011). Cependant, il faut garder à l'esprit que l'activité des enhancers testée par transgénèse n'est pas toujours représentative de l'activité *in vivo*. Avec l'arrivée des techniques d'édition du génome de type CRISPR/Cas9, l'analyse systématique de la fonction *in vivo* des enhancers se développe pour de nombreuses espèces.

Le turnover et le shuffling des enhancers peuvent être la résultante de schémas d'organisation qui assurent une fonction cohérente alors qu'ils facilitent les variations et l'acquisition de nouvelles fonctions, ce que François Jacob appelait le « bricolage évolutif » (Jacob 1977 : Evolution and tinkering).

3.3 Les variations des séquences non-codantes sont à l'origine de maladies

Les variants génétiques non-codants sont divisés en mutations ponctuelles ou SNVs (Single-Nucleotide Variants), en petites insertions et délétions de moins de 50 pb (indels) et en grandes variations structurales comprenant : les variants du nombre de copies (CNVs pour Copy Number Variants) pour les délétions et duplications, ou les variants neutres de types inversions ou translocations. Les études d'associations pangénomiques (GWAS) interrogent divers variants génétiques au sein d'une cohorte d'individus pour évaluer le lien avec un trait particulier. Globalement, 95 % des SNVs identifiés par GWAS sont retrouvés dans des régions intergéniques, dont 75 % sont associés à des sites de sensibilité à l'ADNase I, suggérant un lien fort avec les séquences *cis*-régulatrices (Maurano et al. 2012). Ces séquences régulent le patron d'expression des gènes et des SNVs ou CNVs affectant ces séquences peuvent conduire à l'apparition de pathologies. Il est cependant souvent difficile de relier les variants non-codants à un gène. Des approches de cartographie quantitative d'expression des loci (eQTL) peuvent permettre une interprétation fonctionnelle de ces variations. Il est aussi possible de relier des séquences *cis*-régulatrices à un ou plusieurs gènes grâce aux techniques de capture de conformation de chromatine (3C et ses variantes) et d'identifier notamment les altérations de conformation lorsqu'une séquence est affectée. Les deux paragraphes suivants visent à illustrer l'impact des petites et grandes variations affectant les éléments *cis*-régulateurs dans la fixation des facteurs de transcription et la structure chromatinienne qui peuvent engendrer des maladies génétiques souvent complexes et des cancers.

3.3.1 Les petites variations génétiques modifient la fixation des TFs sur les éléments *cis*-régulateurs

L'analyse de plus en plus de génomes a permis d'identifier des mutations ponctuelles affectant les éléments *cis*-régulateurs qui sont susceptibles de modifier la fixation des TFs sur ces séquences et d'être reliées à l'expression inappropriée de certains gènes.

Les SNVs peuvent entraîner la formation d'une nouvelle séquence consensus pour un TF sur un élément *cis*-régulateur et la fixation ectopique du facteur (fig. 9A). C'est le cas pour le

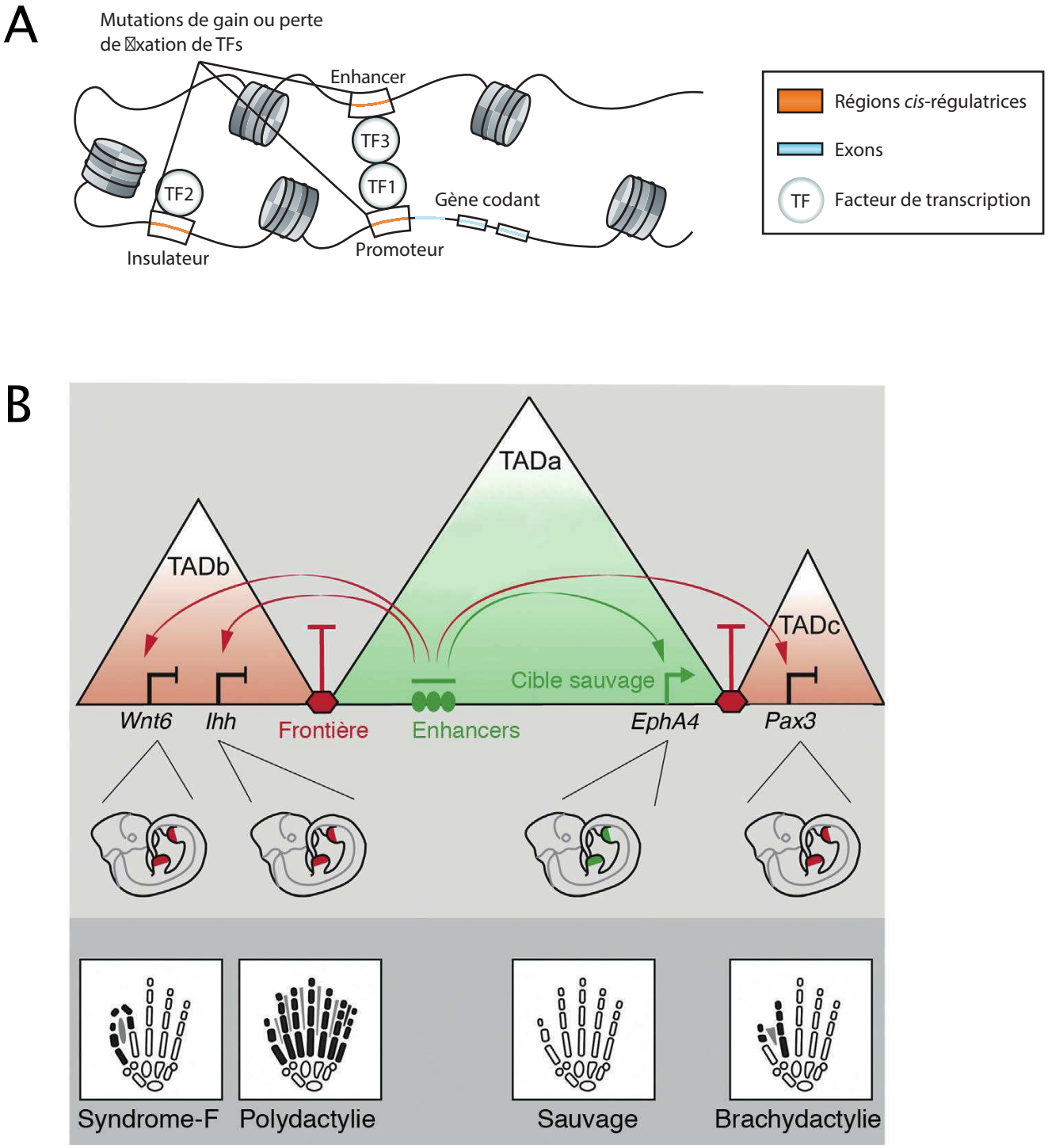


Figure 9 : L'altération des éléments *cis*-régulateurs à l'origine de maladies

A. Les petites variations génétiques impactant les éléments *cis*-régulateurs : enhancers, promoteurs et insulateurs modifient la fixation des facteurs de transcription (TFs).

B. De grandes variations structurales au locus *EphA4* entraînent plusieurs malformations des membres. Le TADa comprend le gène *EphA4* et un groupe d'enhancers régulant son expression dans le membre. L'interaction ectopique des enhancers avec *Pax3* dans le cas de la brachydactylie suite à une délétion, *Wnt6* dans le syndrome F après une inversion ou une duplication et *Ihh* dans la polydactylie suite à une duplication chez l'homme et une délétion chez la souris. En rouge, les frontières des TADs empêchent ces interactions ectopiques et favorisent les interactions entre le gène *EphA4* et les enhancers pour l'expression correcte. Le patron d'expression des gènes associé aux variations structurales est indiqué en rouge, alors que celui sauvage du gène *EphA4* est en vert et le phénotype associé au niveau des membres est représenté en dessous.

D'après Khurana et al., 2016 (A) et Lupiañez et al., 2015 (B).

promoteur du gène *TERT* où des mutations récurrentes ont été observées dans plusieurs types de cancers à l'origine de la création de motifs pour les facteurs ETS conduisant à la surexpression du gène (Huang et al. 2013; Horn et al. 2013). Dans certains cas de leucémie lymphoblastique aiguë à cellules T (T-ALL), des mutations somatiques dans une région non-codante créent un motif de fixation pour MYB formant ainsi un « super-enhancer » en amont du gène *TAL1* (T cell Acute lymphocytic leukaemia 1). La fixation de MYB sur cette séquence entraîne la surexpression du gène *TAL1* qui est un oncogène impliqué dans la différenciation des cellules érythroïdes, illustrant un gain de fixation d'un TF sur un enhancer. Dans le cas contraire, les SNVs abolissent parfois la fixation d'un TF indispensable à l'établissement correct du patron d'expression d'un gène. Une étude combinant du séquençage de génomes entiers et des données épigénomiques a identifié plusieurs mutations ponctuelles dans une région correspondant à l'enhancer du gène *PITF1A* (encoding pancreas-specific transcription factor 1a) qui abolissent la fixation de plusieurs facteurs de transcription importants pour le développement du pancréas et elles sont liées à des cas isolés d'agénésie du pancréas (Weedon et al. 2014). Une étude récente a identifié de fréquentes mutations ponctuelles dans les sites de fixation de CTCF co-occupés par la cohésine dans les cancers colorectaux (Katainen et al. 2015). Il est alors facile d'imaginer les implications de ces mutations sur l'expression des gènes adjacents en tenant compte du rôle du facteur CTCF et de la cohésine dans les frontières des TADs (cf. ci-après).

3.3.2 Les grandes variations structurales altèrent les domaines chromatiniens

L'organisation chromatinienne en TADs représente le squelette du génome qui facilite la reconnaissance des enhancers avec leur promoteur cible en les isolant en même temps des autres régions du génome grâce à leurs frontières enrichies en protéines de structure. Les variations structurales peuvent altérer l'architecture des TADs par délétion, inversion, duplication et repositionnement des frontières conduisant à l'activation ectopique des gènes par des enhancers des domaines voisins et causant une expression aberrante des gènes à l'origine des maladies.

Une étude élégante montre que de grandes variations structurales au locus *EphA4* sont associées à plusieurs malformations des membres (Lupiáñez et al. 2015). Une délétion comprenant le gène *EphA4* conduit à une brachydactylie (doigts courts) alors qu'une inversion ou une duplication d'une partie du TAD d'*EphA4* entraîne un syndrome F (fusion des doigts). De plus, une duplication chez l'homme et une délétion chez la souris entre le gène *Ihh* et *EphA4* conduit au même phénotype de polydactylie (au moins 7 doigts). En utilisant le

système CRISPR/Cas9, les auteurs ont reproduit ces variations chez la souris. Ils montrent alors qu'un groupe d'enhancers contrôlant l'expression d'*EphA4* dans le membre régule *Pax3* dans le cas de la brachydactylie, *Wnt6* dans le syndrome F et *Ihh* dans la polydactylie à la place d'*EphA4* (fig. 9B). Tous les phénotypes sont causés par une interaction ectopique entre le groupe d'enhancer et les autres gènes entraînant leur expression anormale et ce uniquement lorsqu'une frontière de TAD est affectée. Ces résultats confirment l'importance de l'intégrité du TAD et des frontières pour l'expression correcte des gènes. (fig. 9B). D'autres exemples de grandes variations structurales affectant les TADs ont été étudiés et associés au syndrome de Liebenberg (Spielmann et al. 2012; Seoighe et al. 2014), de Rett (Allou et al. 2012) et à des maladies congénitales (Ibn-Salem et al. 2014). La mise en évidence de la structure du génome en TADs met en lumière des études plus anciennes, notamment celle décrivant une inversion au locus *SHH* conduisant à l'expression ectopique du gène où les auteurs introduisent alors la notion 'd'adoption d'enhancer' (Lettice et al. 2011). D'autre part, une étude récente interroge la conformation du génome de cellules cancéreuses de prostate par Hi-C et montre que celui-ci comprend de plus petits TADs à cause de la formation de frontières supplémentaires. Ceci est notamment le cas pour une délétion au locus du suppresseur de tumeur *TP53* qui conduit à la division d'un TAD en deux. Ces subdivisions sont accompagnées de modifications des interactions entre enhancers et promoteurs et liés à un changement d'expression des gènes (Taberlay et al. 2016).

Longtemps considérés comme une partie du 'junk DNA' du génome, les éléments *cis*-régulateurs sont maintenant au cœur d'une intense recherche. L'identification et le rôle fonctionnel de ces séquences sont un enjeu majeur dans la compréhension des thèmes abordés dans cette partie. La dernière partie retrace les connaissances actuelles sur la régulation du gène de développement *Krox20* dans le cerveau postérieur des Vertébrés ainsi que les arguments qui ont conduit à la dissection fonctionnelle des éléments *cis*-régulateurs contrôlant son expression.

4. La segmentation du rhombencéphale et le gène *Krox20*

4.1 La mise en place du rhombencéphale

Au cours du développement, l'apparition de la plaque neurale marque le début de la neurulation pendant laquelle se met en place le système nerveux. La plaque neurale va s'invaginer et former deux bourrelets neuraux qui vont fusionner le long de l'axe antéro-postérieur de l'embryon établissant le tube neural. Peu après, les cellules les plus dorsales du tube délaminent, c'est-à-dire quittent le toit du tube neural, et donnent ainsi naissance aux cellules de la crête neurale, à l'origine du système nerveux périphérique, des mélanocytes, des os et des tissus conjonctifs de la face.

Le tube neural se divise dans un premier temps en trois vésicules : le prosencéphale, le mésencéphale et le rhombencéphale selon l'axe antéro-postérieur (fig. 10A). Dans un second temps, le prosencéphale dérive en télencéphale et le diencéphale, le rhombencéphale donnera quant à lui le métencéphale et le myélencéphale (fig. 10A). Le rhombencéphale est à l'origine du cervelet, du pont et du bulbe rachidien chez l'adulte (fig. 10A). Il joue un rôle clef dans le développement de la tête et représente un centre complexe de coordination des activités motrices, du rythme respiratoire et de nombreuses fonctions inconscientes. La crête neurale du rhombencéphale, quant à elle, fournit les cellules des ganglions crâniens (une majorité de neurones sensoriels, les cellules satellites et de Schwann) et des arcs branchiaux qui eux-mêmes dérivent en os, muscles et cartilages de la face. Le rhombencéphale est soumis à un processus de segmentation transitoire qui conduit à la formation de sept segments appelés rhombomères (r) (fig. 10B). Ces unités morphologiques préfigurent la différenciation neuronale, l'organisation des nerfs crâniens et la spécification des crêtes neurales (Lumsden & Keynes 1989). Les rhombomères sont aussi des compartiments de lignage cellulaire séparés par des frontières (Birgbauer & Fraser 1994) et des domaines d'expression de nombreux gènes (fig. 10C). Les gènes de segmentation sont définis par l'élimination ou la modification de leur domaine d'expression lorsqu'ils sont inactivés. Parmi eux, les gènes *Krox20* et *MafB* sont responsables de l'identité et du maintien des rhombomères en contrôlant directement l'expression d'autres gènes (fig. 10C). *Krox20* est un facteur de transcription à doigts à zinc (Chavrier et al. 1990) qui s'exprime dans les territoires non adjacents r3 et r5 (Wilkinson et al. 1989). *Krox20* permet l'acquisition d'une identité de position des rhombomères à travers l'activation de gènes *Hox* tels que *Hoxa2* (Nonchev et al. 1996) et *Hoxb2* (Sham et al. 1993) dans r3 et r5 ainsi que *Hoxb3* (Seitanidou et al. 1997) dans r5 (fig. 10C). Les gènes *Hox* appartiennent à une famille de facteurs de transcription à

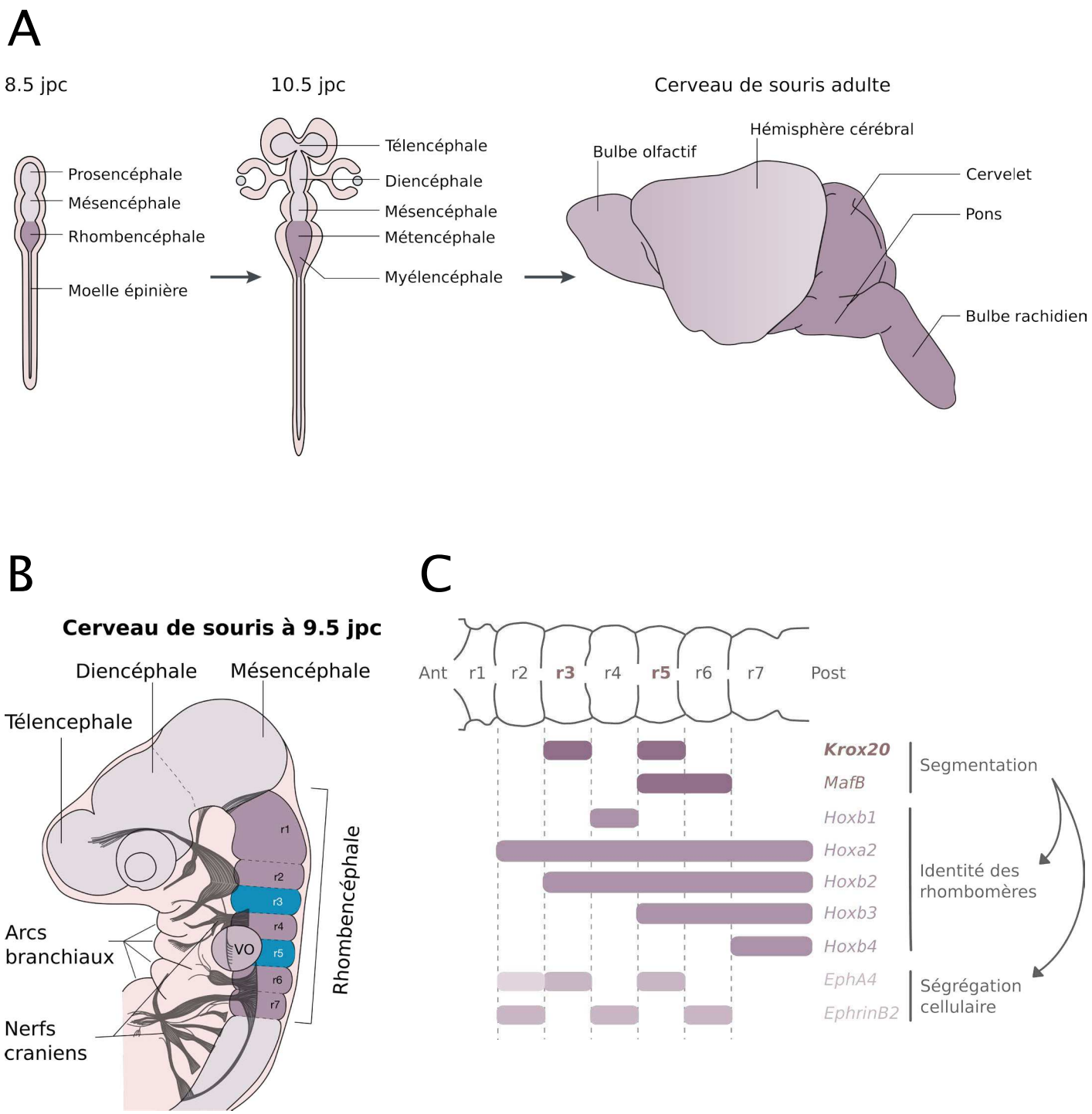


Figure 10 : La mise en place du rhombencéphale chez la souris

A. À 8,5 jpc, le tube neural forme trois vésicules dont le prosencéphale, le mesencéphale et le rhombencéphale, puis cinq vésicules où le rhombencéphale dérive en métencéphale et myélencéphale à 10,5 jpc. Ces structures sont à l'origine, à l'âge adulte, de trois parties postérieures du cerveau : le cervelet, le pons et la médulla.

B. La segmentation du rhombencéphale en sept rhombomères préfigure l'organisation des nerfs crâniens qui innervent les arcs branchiaux. En bleu les rhombomères r3 et r5, territoires d'expression du gène *Krox20*.

C. Les gènes de segmentation *Krox20* et *MafB* contrôlent l'identité des rhombomères via les gènes *Hox* et la ségrégation cellulaire via les gènes *Eph*. VO : vésicule otique.

D'après le manuscrit de Michel A. Wassef (A) et Guthrie et al., 2007 (B).

homéodomaine conservée chez de nombreuses espèces animales. Ils sont organisés en complexes de gènes dont l'ordre sur le chromosome coïncide avec leur domaine d'expression et leur fonction de spécification tissulaire le long de l'axe antéro-postérieur. Dans le rhombencéphale, ils sont exprimés de cette façon avec un décalage de deux rhombomères (fig. 10C). *Krox20* contrôle aussi l'expression de la molécule d'adhésion EphA4 dans les deux territoires impairs, participant ainsi à la ségrégation cellulaire grâce au système de récepteur Eph et ligand Ephrins exprimés dans les rhombomères complémentaires (Xu et al. 1999). La grande conservation évolutive de l'expression du gène *Krox20* suggère que son rôle est conservé chez l'ensemble des Vertébrés.

4.2 La fonction de *Krox20* dans le rhombencéphale dévoilée par l'étude du mutant *Krox20-lacZ*

Des études de perte de fonction du gène *Krox20* par knock-in (*Krox20-lacZ*, fig.11A) chez la souris ont été une étape clef dans la compréhension de son rôle dans la segmentation du rhombencéphale (Schneider-Maunoury et al. 1993, fig.11). L'observation de l'expression de *lacZ*, rapporteur de la transcription du locus *Krox20*, montre une perte d'expression rapide du gène chez l'homozygote par rapport à l'hétérozygote (fig. 11B-C). Deux hypothèses non-exclusives sont alors avancées pour expliquer ce phénotype : soit la mutation de *Krox20* entraîne l'élimination de r3 et r5, soit *Krox20* est nécessaire au maintien de sa propre expression. Les analyses qui suivront démontreront effectivement que la mutation de *Krox20* entraîne la perte de r3 et r5 (Schneider-Maunoury et al. 1993; Schneider-Maunoury et al. 1997). Des expériences de traçage génétique ont permis de suivre le destin cellulaire des territoires présomptifs de r3 et r5 (Voiculescu et al. 2001). Chez le mutant $Krox20^{LacZ/LacZ}$, les cellules qui auraient dû être spécifiées en r3 acquièrent une identité r2 et r4, et celles de r5 une identité r6 au stade 8-10 somites. Plus tard vers 20 somites, une vague d'apoptose conduit à la réduction de la taille du rhombencéphale associée à la disparition des rhombomères 3 et 5 (fig. 11D-E). De plus, la délétion de *Krox20* a pour conséquences la perte d'expression de gènes sous son contrôle comme *Hoxa2*, *Hoxb2* et *EphA4* dans r3 et r5 et l'activation de la follistatine dans r3 (Seitanidou et al. 1997). Enfin, la perte des rhombomères se reflète au niveau de l'organisation neuronale. En effet, le rhombencéphale segmenté donne naissance à des nerfs crâniens segmentés dont la mise en place est alors perturbée. Chez l'homozygote, on observe un rapprochement des nerfs V et VII/VIII ainsi que des nerfs IX et X (Schneider-Maunoury et al. 1993). Ces rapprochements peuvent être corrélés à la perte des rhombomères 3 et 5 qui perturbent la différenciation neuronale (Schneider-Maunoury et al. 1997). La délétion de *Krox20* n'est pas létale au cours de l'embryogénèse mais la majorité des

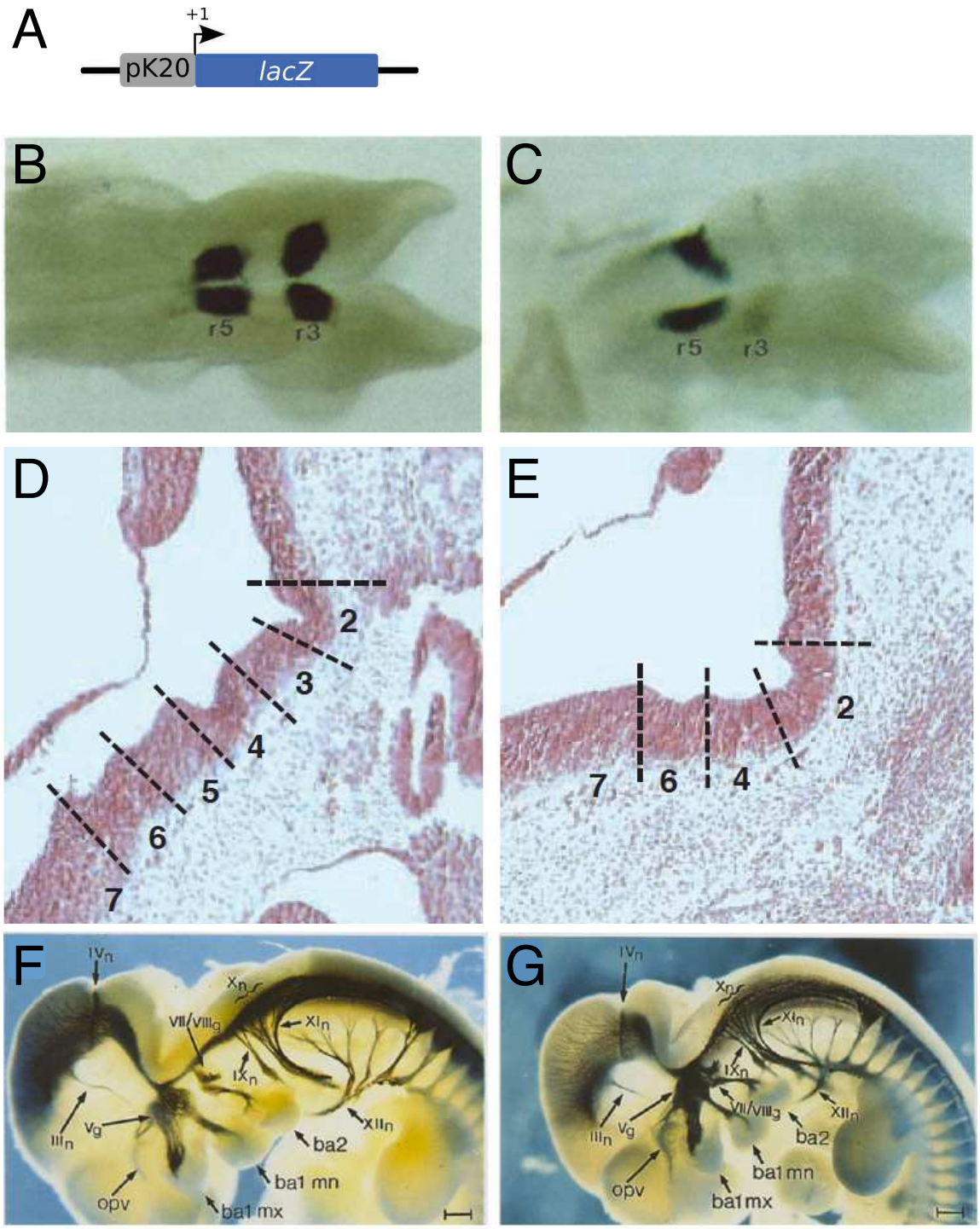


Figure 11 : Phénotypes associés au mutant Krox20-lacZ

A. Allèle muté *Krox20*^{lacZ} où la séquence codante du gène *Krox20* a été remplacée par knock-in par celle du *lacZ* qui servira de gène rapporteur.

B & C. Embryons entiers de souris hétérozygote (B) et homozygote (C) à 9 jpc colorés au X-gal.

D & E. Coupes sagittales en paraffine d'embryons sauvage (D) et mutant (E) à 10,5 jpc où les numéros des rhombomères sont indiqués.

F & G. Embryons entiers de souris hétérozygote (B) et homozygote (C) à 10,5 jpc marqués avec un anticorps contre les neurofilaments. La barre représente 0,25mm.

opv : vésicule optique, IIIn : nerf oculomoteur ; IVn : nerf trochléaire, Vg : ganglion trijumeau ; VII/VIIIg : ganglion facial et auditif ; IXn : nerf glosso-pharyngien ; Xn : nerf vague ; XIn : nerf accessoire ; XIIn : nerf hypoglosse ; ba1 mx, ba1 mn : composantes maxillaire et mandibulaire du premier arc branchial ; ba2 : second arc branchial.

D'après Schneider-Maunoury et al., 1993 & 1997.

souriceaux meurent 24h après la naissance et les autres n'atteignent pas plus de deux semaines ; ceci étant principalement dû à des défauts respiratoires associés à la mauvaise spécification des nerfs crâniens. Différentes expériences qui seront développées dans la partie suivante montrent aussi que *Krox20* est nécessaire au maintien de sa propre expression.

4.3 Les éléments *cis*-régulateurs du gène *Krox20* dans le rhombencéphale

La recherche des éléments *cis*-régulateurs de *Krox20* est une étape clef dans la compréhension de son patron d'expression. Afin de les identifier des expériences de transgénèse ont été réalisées chez la souris (fig. 12A-H) avec deux chromosomes artificiels bactériens (BAC) de poulet dont le génome est trois fois plus compact que celui de la souris, incluant *Krox20* et les séquences en amont et en aval du gène. Ensuite des expériences d'électroporation chez le poulet et de transgénèse chez la souris de sous-fragments issus de ces BAC, ont permis de caractériser trois enhancers actifs dans le rhombencéphale (Chomette et al. 2006). Ces derniers nommés A (situé à -217 kb du promoteur), B (-164 kb) et C (-144 kb) sont actifs dans r3/r5, r5 et r3-r5 respectivement. Alors que les éléments B et C sont actifs dès l'allumage du gène et ce même en absence de la protéine KROX20 (Chomette et al. 2006, fig. 12E-I), supposant ainsi leur rôle d'initiateur de l'expression de *Krox20*, l'élément A est totalement inactif en absence de la protéine (fig. 12D), suggérant un rôle autorégulateur. Une recherche *in silico* a effectivement permis d'identifier plusieurs sites de fixation de KROX20 au sein de l'élément A chez la souris et le poulet suggérant qu'il soit directement autorégulé par KROX20. De plus, des expériences d'électroporation chez le poulet ont montré que l'élément A fusionné à un gène rapporteur est activé en présence d'un plasmide d'expression de *Krox20* ainsi que des expériences de retard sur gel ont démontré la fixation de la protéine sur les sites identifiés et l'absence de fixation lorsque ces sites sont mutés, confirmant ainsi une action directe de KROX20 sur cet élément (Chomette et al. 2006).

Par ailleurs, le laboratoire a identifié différents facteurs permettant l'activation des éléments B et C comme Meis et Hox/Pbx (Wassef et al. 2008) sur l'élément C, ainsi que Hnf1B et MafB par coopération synergique sur l'élément B, par recherche *in silico* de sites de fixation pour des facteurs de transcription et validation par retard sur gel d'oligonucléotides sauvages ou mutés pour les sites en question (Chomette et al. 2006; Labalette et al. 2011). Les facteurs Pbx et Meis sont requis pour l'établissement du domaine d'expression antérieur du gène *Krox20* (Waskiewicz et al. 2002; Choe et al. 2002). Une indication de l'implication des facteurs Hox dans l'expression de *Krox20* a été obtenue chez le Xénope où la perte de fonction combinée de *HoxA1*, *HoxB1* et *HoxD1* conduit à l'absence totale d'expression de

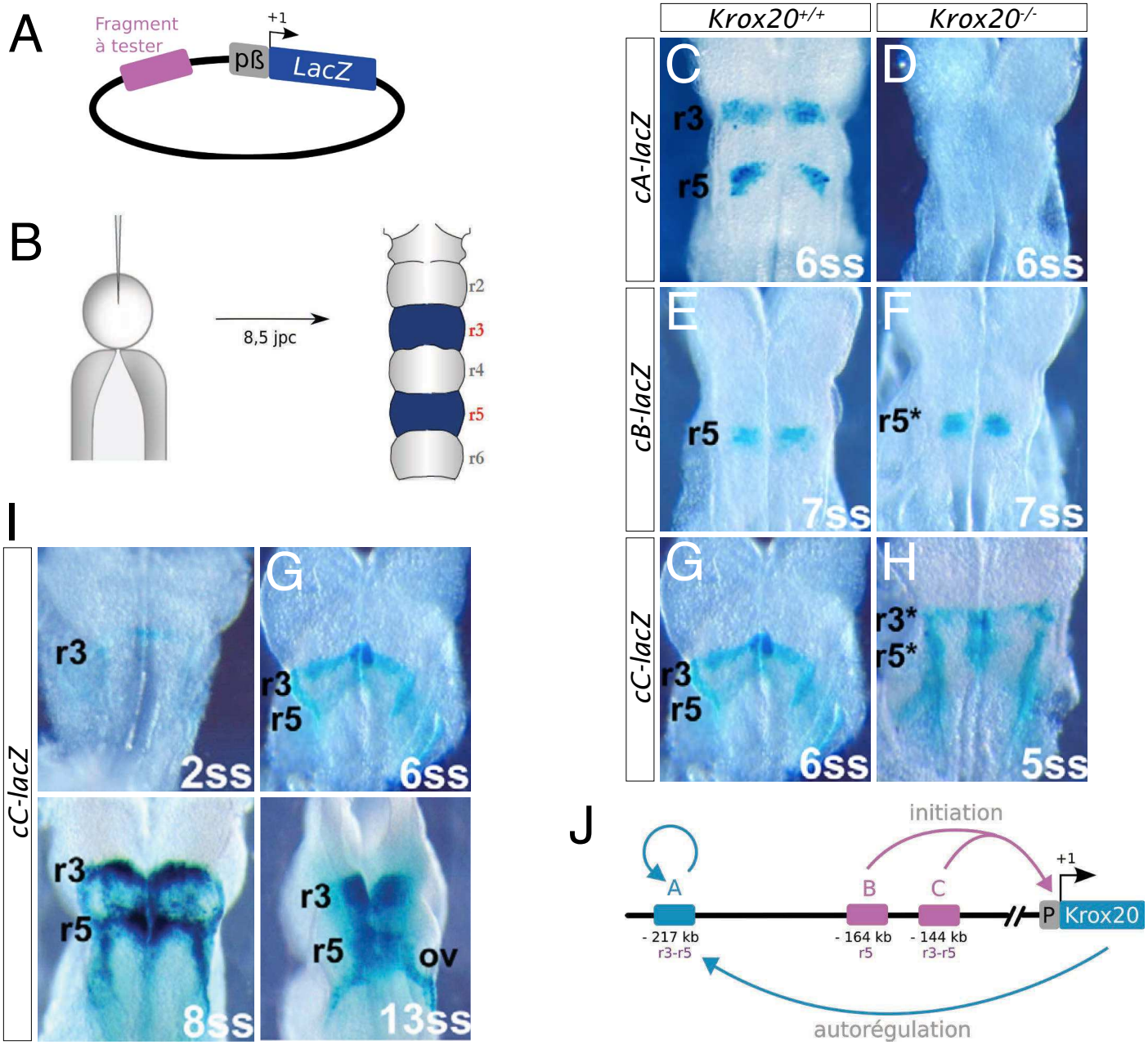


Figure 12 : Enhancers et régulation du gène *Krox20* dans le rhombencéphale

A. Vecteur utilisé pour tester l'activité de différents fragments issus de deux BACs de poulet grâce au gène rapporteur *lacZ* sous le contrôle du promoteur minimal de la *ß-actine*.

B. Transgénèse dans l'embryon de souris utilisant le vecteur présenté en A.

C & D. Embryons entiers de souris en fond sauvage (C) et homozygote nul (D) comprenant le transgène *cA-lacZ* à 8,5 jpc colorés au X-gal.

E & F. Embryons entiers de souris en fond sauvage (E) et homozygote nul (F) comprenant le transgène *cB-lacZ* à 8,5 jpc colorés au X-gal.

G & H. Embryons entiers de souris en fond sauvage (G) et homozygote nul (H) comprenant le transgène *cC-lacZ* à 8,5 jpc colorés au X-gal.

I. Cinétique d'activité de l'élément C : embryons entiers de souris en fond sauvage comprenant le transgène *cC-lacZ* à 2ss, 6ss, 8ss et 13ss colorés au X-gal.

J. Modèle de la régulation du gène *Krox20* dans le rhombencéphale : Les éléments B (à -164 kb) et C (à -144 kb) pour l'initiation du gène *Krox20* dans r5 pour B et r3 et r5 pour C. L'élément A est un élément autorégulateur situé à -217 kb dans r3 et r5.

ov : vésicule otique, pß : promoteur de la *ß-actine*, r : rhombomères, ss : stade somitique, kb : kilobase.

D'après le manuscrit de thèse de Yassine Bouchoucha (B) et Chomette et al., 2006 (C-I).

Krox20 dans le rhombencéphale (McNulty et al. 2005). Les facteurs Hnf1B et MafB sont connus pour être des facteurs de transcription en amont de l'activation de *Krox20* dans r5 (Wiellette & Sive 2003; Cordes & Barsh 1994). De plus, les éléments B et C répondent à la signalisation FGF (Fibroblast Growth Factor) (Labalette et al. 2011; Labalette et al. 2015), qui est requise pour l'initiation de l'expression de *Krox20* et détermine la taille des rhombomères 3 et 5 (Walshe et al. 2002; Aragon & Pujades 2009). Le facteur *Sp5* est capable de fixer l'élément C *in vitro* et serait l'intermédiaire de l'activité précoce de l'élément C par la signalisation FGF chez le poisson-zèbre (Labalette et al. 2015). Des expériences d'inhibition de cette signalisation dans l'embryon de poisson-zèbre permettent de moduler le niveau d'initiation de *Krox20* (Labalette et al. 2011) et suggèrent que la boucle d'autorégulation ne peut se mettre en place qu'une fois atteint un niveau seuil de protéine KROX20 dans la cellule. Ces arguments suggèrent que les élément B et C sont impliqués dans le démarrage de l'expression de *Krox20* dans le rhombencéphale.

Afin d'identifier de nouveaux facteurs de transcription impliqués dans le démarrage précoce de l'expression de *Krox20*, un RNA-seq sur embryons disséqués avant 4 somites a été réalisé pour le rhombencéphale et le mésoderme sous-jacent (Le Men et al., données non publiées). Cette étude récapitule les acteurs connus de la segmentation du rhombencéphale comme les gènes *Hox* ainsi que les facteurs cités précédemment. Parmi les gènes candidats exprimés dans le rhombencéphale, on peut citer *Olig3* dans r3, *Pou5f1* qui semblerait réguler positivement l'expression de *Krox20* dans r3 par l'intermédiaire de l'élément C.

L'expression du gène *Krox20* suit donc deux phases : le modèle de sa régulation serait la mise en place d'une phase d'initiation grâce aux éléments B et C activés par différents facteurs de transcription suivie d'une phase d'autorégulation grâce à l'élément A où, cette fois, la protéine KROX20 activerait sa propre production (fig. 12J). Ainsi l'organisation du rhombencéphale dépend à la fois de la phase d'initiation et de la boucle d'autorégulation qui joue un rôle déterminant dans la taille finale des rhombomères 3 et 5.

4.4 Un modèle transcriptionnel incomplet

Le modèle de la régulation de *Krox20* repose essentiellement sur la caractérisation spatiale et temporelle de l'activité des enhancers placés en amont d'un gène rapporteur et testés en transgenèse. Pour tester ce modèle de régulation transcriptionnelle et l'implication des éléments *cis*-régulateurs pour l'expression d'un gène, la délétion *in vivo* d'enhancers est une méthode de premier choix. Peu d'études publiées montrent la délétion d'éléments régulateurs car il y a un risque fort de redondance génomique avec d'autres séquences

régulatrices (Kurokawa et al. 2004; Lam et al. 2015). Afin d'étudier le rôle de l'initiation et de l'autorégulation du gène *Krox20* dans la segmentation du rhombencéphale, les délétions de l'élément C (seul élément actif précocement dans r3 en absence de protéine KROX20, fig. 12H) et de l'élément A, respectivement, ont été réalisées *in vivo* chez la souris.

La délétion de l'élément A confirme la nécessité de cet élément pour l'amplification et le maintien de l'expression du gène *Krox20* dans les rhombomères 3 et 5 puisqu'on observe une extinction prématurée de *Krox20* dans ces 2 rhombomères (Bouchoucha et al. 2013). L'analyse du fonctionnement moléculaire de cet élément démontre son rôle d'interrupteur bistable par la combinaison d'approches expérimentales et de modélisation. En d'autres termes, l'activation de l'élément A à l'échelle cellulaire aboutit à la formation de deux états stables : un état Krox20^{+} lorsque l'élément est activé, et un état Krox20^{-}. Cette bistabilité est due à une forte coopérativité de fixation des protéines KROX20 liant l'élément A. De ce fait, dans le cas où le niveau d'initiation est perturbé, le nombre de cellules qui s'engagent dans la voie d'un état Krox20^{+} est plus faible entraînant une réduction de la taille des rhombomères 3 et 5 comme observé précédemment (Labalette et al. 2011). L'élément autorégulateur permet donc la mise en place du patron de développement du rhombencéphale en définissant la dose de Krox20 nécessaire à l'activation de ces cibles transcriptionnelles comme *EphA4*. Je développerai dans la première partie des résultats les différents phénotypes associés à la délétion de cet élément.

De façon inattendue, la délétion de l'élément C n'affecte pas le démarrage de l'expression de *Krox20* dans r3, mais son maintien, de manière très analogue à l'élément A. Ceci nous a amené à envisager, d'une part, que les activités des éléments A et C peuvent coopérer pour assurer l'autorégulation du gène et, d'autre part, qu'il doit exister un autre élément initiateur pour l'expression dans r3. La régulation transcriptionnelle de *Krox20* est donc plus complexe qu'imaginée initialement. La deuxième partie des résultats visera à élucider les possibles interactions entre séquences *cis*-régulatrices ainsi que les changements chromatiniens dont l'accessibilité, la conformation et les marques de chromatine associées à l'expression du gène *Krox20*.

Enfin, ces nouvelles questions sur la régulation du gène *Krox20* nécessitent pour y répondre un grand nombre de cellules d'identité homogène, notamment pour la recherche systématique de nouveaux éléments régulateurs, et l'utilisation de nouvelles technologies. Je développerai, dans une troisième partie, deux approches pour essayer de pallier à ces limitations dont l'une nous apportera des informations évolutives sur la régulation du gène *Krox20*.

RÉSULTATS

- Chapitre 1-

L'autorégulation du gène *Krox20* dans la segmentation du rhombencéphale

De nombreux gènes impliqués dans le développement du système nerveux sont soumis à une autorégulation. Ce processus est crucial dans l'organogénèse et notamment pour la formation de frontières. Nous avons décidé d'utiliser *Krox20* comme gène modèle d'un phénomène d'autorégulation. Par délétion de l'élément *cis*-régulateur A dans le génome murin, nous montrons que *Krox20* est autorégulé de façon directe *via* cet élément dans le rhombencéphale. L'autorégulation permet l'amplification et le maintien de *Krox20* et d'atteindre la dose nécessaire à l'activation d'au moins un de ses gènes cibles (*EphA4*) pour la mise en place de r3 et r5. En son absence toutes les cellules ne sont pas spécifiées correctement et l'architecture neuronale est perturbée. L'analyse du fonctionnement moléculaire de l'élément A démontre son rôle d'interrupteur bistable par la combinaison d'approches expérimentales, *in vitro* et *in vivo* chez le poisson-zèbre, et de modélisation. En d'autres termes, l'activation de l'élément A à l'échelle cellulaire aboutit à la formation de deux états stables : un état Krox20^{+} lorsque l'élément est activé et un état Krox20^{-}. La séparation entre ces deux états dépend principalement du niveau d'initiation de *Krox20* qui doit atteindre un seuil pour enclencher la boucle d'autorégulation. A l'échelle moléculaire cette bistabilité est due à une forte coopérativité entre les protéines KROX20 liant l'élément A. L'extinction progressive de l'expression de *Krox20* peut être le fait de la déstabilisation de la boucle d'autorégulation par des facteurs répresseurs (*Nlz*). Ce travail a permis de comprendre au niveau moléculaire comment l'activation de *Krox20* définit le destin cellulaire et la taille des rhombomères.

Ce chapitre comprend un article (Bouchoucha et al. 2013) auquel j'ai participé pour l'étude du mutant murin de la délétion de l'élément A et un complément comprenant des expériences supplémentaires sur le rôle de l'autorégulation dans la spécification cellulaire et l'organisation des nerfs crâniens.

- Article -

Dissection of a Krox20 positive feedback loop driving cell fate choices in hindbrain patterning

Yassine X Bouchoucha[1,2,3,5], Jürgen Reingruber[1,2,3,4,5], Charlotte Labalette[1,2,3], Michel A Wassef[1,2,3,6], Elodie Thierion[1,2,3], Carole Desmarquet-Trin Dinh[1,2,3], David Holcman[1,2,3,4,7,*], Pascale Gilardi-Hebenstreit[1,2,3,7] and Patrick Charnay[1,2,3,*]

[1] Ecole Normale Supérieure, IBENS, Paris, France,

[2] INSERM, U1024, Paris, France,

[3] CNRS, UMR 9197, Paris, France,

[4] Group of Computational Biology and Applied Mathematics, IBENS, Paris, France,

[6] Present address: Institut Curie, Unité de Génétique et Biologie du Développement, 75005

* Corresponding authors.

Author contributions:

YXB conceived and performed a large part of the experimental work, carried out simulations and participated in the writing of the manuscript.

JR developed and analyzed the mathematical model, developed the simulation tools, carried out simulations and participated in the writing of the manuscript.

CL participated in the experiments involving the repression of the autoregulatory loop.

MAW generated zebrafish lines and initiated the project.

ET participated in the analysis of the mouse mutants.

CD-TD generated a mouse transgenic line.

DH participated in the development of the mathematical model and in the writing of the manuscript. PG-H co-led the experimental part and participated in the writing of the manuscript.

PC led the project and participated in the writing of the manuscript.

[5,7] These authors contributed equally to this work.

Molecular Systems Biology 9; Article number 690; doi:10.1038/msb.2013.46
Citation: ***Molecular Systems Biology* 9:690**
www.molecularsystemsbiology.com

OPEN ACCESS
SOURCE DATA
TRANSPARENT PROCESS

molecular systems biology

Dissection of a Krox20 positive feedback loop driving cell fate choices in hindbrain patterning

Yassine X Bouchoucha[1,2,3,5], Jürgen Reingruber[1,2,3,4,5], Charlotte Labalette[1,2,3], Michel A Wassef[1,2,3,6], Elodie Thierion[1,2,3], Carole Desmarquet-Trin Dinh[1,2,3], David Holcman[1,2,3,4,7,*], Pascale Gilardi-Hebenstreit[1,2,3,7] and Patrick Charnay[1,2,3,*]

[1] Ecole Normale Supérieure, IBENS, Paris, France, [2] INSERM, U1024, Paris, France, [3] CNRS, UMR 8197, Paris, France and [4] Group of Computational Biology and Applied Mathematics, IBENS, Paris, France
[5] These authors contributed equally to this work.
[6] Present address: Institut Curie, Unité de Génétique et de Biologie du Développement, 75005 Paris, France.
[7] These authors contributed equally to this work.
* Corresponding authors. P Charnay or D Holcman, Ecole Normale Supérieure; IBENS, 46 rue d'Ulm, 75005 Paris, France. Tel.: +33 1 4432 3607; Fax: +33 1 4432 3988; E-mail: patrick.charnay@ens.fr or david.holcman@ens.fr

Received 13.3.13; accepted 21.8.13

Although feedback loops are essential in development, their molecular implementation and precise functions remain elusive. Using enhancer knockout in mice, we demonstrate that a direct, positive autoregulatory loop amplifies and maintains the expression of Krox20, a transcription factor governing vertebrate hindbrain segmentation. By combining quantitative data collected in the zebrafish with biophysical modelling that accounts for the intrinsic stochastic molecular dynamics, we dissect the loop at the molecular level. We find that it underpins a bistable switch that turns a transient input signal into cell fate commitment, as we observe in single cell analyses. The stochasticity of the activation process leads to a graded input–output response until saturation is reached. Consequently, the duration and strength of the input signal controls the size of the hindbrain segments by modulating the distribution between the two cell fates. Moreover, segment formation is buffered from severe variations in input level. Finally, the progressive extinction of *Krox20* expression involves a destabilization of the loop by repressor molecules. These mechanisms are of general significance for cell type specification and tissue patterning.
Molecular Systems Biology (2013) **9,** 690; published online 24 September 2013; doi:10.1038/msb.2013.46
Subject Categories: simulation and data analysis; development
Keywords: Fgf; Krox20; rhombomere; stochastic model; transcriptional enhancer

Introduction

Cell fate specification is essential to metazoan development. It usually involves successive choices during which cells have the potential to commit to two distinct fates. Understanding the basis of cell fate specification can therefore be reduced to unravelling the molecular mechanisms underlying such choices (Graham *et al*, 2010 and references therein). Fate decisions can be induced by intrinsic cues, which may be asymmetrically distributed during cell division (Tajbakhsh *et al*, 2009; Graham *et al*, 2010), extrinsic factors that are provided by the cellular environment (Briscoe, 2009), or both. Once a choice has been made, it may become irreversible to preserve tissue integrity and directionality of the developmental process. Fate choices can be implemented by biochemical mechanisms involving feedback loops, which are maintained independently of the initial activating signal (Laslo *et al*, 2006; Graham *et al*, 2010). A single transcription factor that positively regulates its own expression constitutes the simplest genetic network that can generate a bistable switch underlying binary choices (Meinhardt, 1982; Ferrell, 2002; Tajbakhsh *et al*, 2009; Graham *et al*, 2010). The first autoregulatory transcription factor mechanism characterized experimentally was the lambda repressor, which underlies the choice between lysogeny and lytic cycle in bacteria. This example provided a paradigm for a genetic switch (Ptashne, 1986) and revealed the cooperative binding of the repressor to its DNA target as well as additional mechanisms ensuring robust efficiency. Although positive feedback loops are likely to be essential in many developmental processes, their dynamics, molecular implementation and precise functions remain elusive in vertebrates.

The establishment of hindbrain anterior–posterior (AP) identity involves a transient segmentation, which leads to the formation of seven to eight segments called rhombomeres (r) (Lumsden and Krumlauf, 1996; Briscoe, 2009). Rhombomeres constitute cell compartments and developmental units for neuronal differentiation and branchiomotor nerve organization (Lumsden and Keynes, 1989). The transcription factor Krox20 (also known as Egr2) is specifically expressed in r3 and r5, and is required for the formation and specification of these

© 2013 EMBO and Macmillan Publishers Limited

rhombomeres, in particular for establishing odd- versus even-numbered identity (Schneider-Maunoury *et al*, 1993; 1997; Swiatek and Gridley, 1993; Voiculescu *et al*, 2001). Three evolutionarily conserved transcriptional enhancer elements active in the hindbrain have been identified in the *Krox20* locus, termed A, B and C (Chomette *et al*, 2006; Wassef *et al*, 2008). Elements B and C drive the expression of reporter constructs in r5 and r3/5, respectively. Their activity is independent of Krox20, but is modulated by FGF signalling (Labalette *et al*, 2011). They are thought to be involved in the initiation of *Krox20* expression. Element A drives reporter expression in r3 and r5, and contains several Krox20-binding sites. The integrity of these sites are necessary for its activity, suggesting that element A is involved in a positive feedback loop (Chomette *et al*, 2006).

In the present study, to understand the mechanisms of a vertebrate autoregulatory loop, we took advantage of our extended knowledge of *Krox20* regulation and of the high conservation during vertebrate evolution of the molecular mechanisms governing its expression, including the activities of the *cis*-acting elements. This conservation allowed us to move between species (mouse, chick or zebrafish), taking advantage of the specificities of each experimental system. Using a knockout approach in the mouse, we established that element A is indeed required for Krox20 positive feedback in the developing hindbrain. As Krox20 DNA-binding, promoter activation, transcription, translation and degradation events are intrinsically subject to fluctuations, we developed a stochastic model to quantify and predict the role of these fluctuations in the determination of cell fate. Because of these fluctuations, a fixed input signal leads to variability in the cell fate choice. Using our stochastic model, we determine the bimodal cell fate distribution that a transient input signal induces into a homogenous population of cells. By combining quantitative data collected in the zebrafish with stochastic modelling, mathematical analysis and numerical simulations, we reach an unprecedented understanding of the molecular dynamics underlying a vertebrate patterning process at the cellular level.

Results

Element A controls Krox20 autoregulation in the mouse hindbrain

Element A activity is dependent on the direct binding of Krox20, shown in mouse and chick (Chomette *et al*, 2006), raising the possibility that this element is responsible for the autoregulation of the gene in the hindbrain. To establish this point, we generated a knockout of element A in the mouse. The details of the strategy are presented in Figure 1A. Two alleles were generated: *Krox20*$^{\Delta A}$, where element A is deleted and *Krox20*$^{A^*}$, where element A is replaced by element cA*. cA* is a version of the chick element A that contains specific mutations in the Krox20-binding sites that prevent binding of wild-type (WT) Krox20 and instead allow binding of a mutant Krox20 protein, Krox20* (Nardelli *et al*, 1991; Supplementary Figure S1A–C).

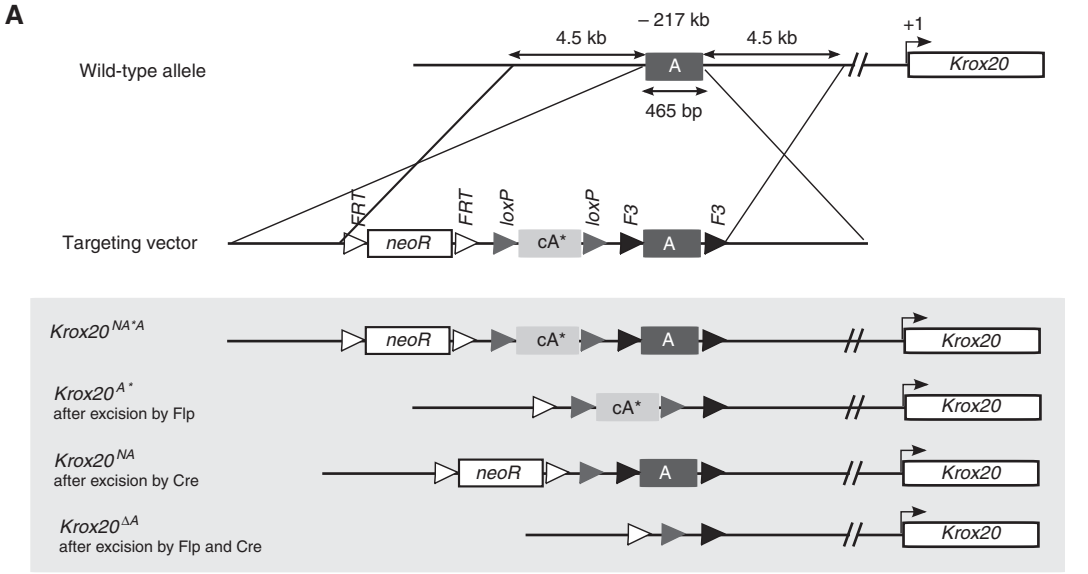
The consequences of element A deletion on *Krox20* expression were analysed by mRNA *in-situ* hybridization in homozygous mutant embryos (Figure 1B). No differences are observed between *Krox20*$^{\Delta A/\Delta A}$ embryos and their littermate *Krox20*$^{\Delta A/+}$ or WT controls until approximately six somites (s). At 6s, the level of *Krox20* mRNA and the extension of the r3 domain are slightly reduced in the homozygous mutants compared with that in controls (Figure 1B). At 8s, r3 expression is almost lost in *Krox20*$^{\Delta A/\Delta A}$ embryos, whereas it persists beyond 12s in controls. In r5, a reduction in the level of *Krox20* mRNA compared with that in controls is observed from 8s and expression is lost around 12s, whereas it persists beyond 16s in controls. These data indicate that the 465-bp sequence deleted in the mutant is required for both amplification and maintenance of *Krox20* expression and that element A is a key component of the *Krox20* autoregulatory loop.

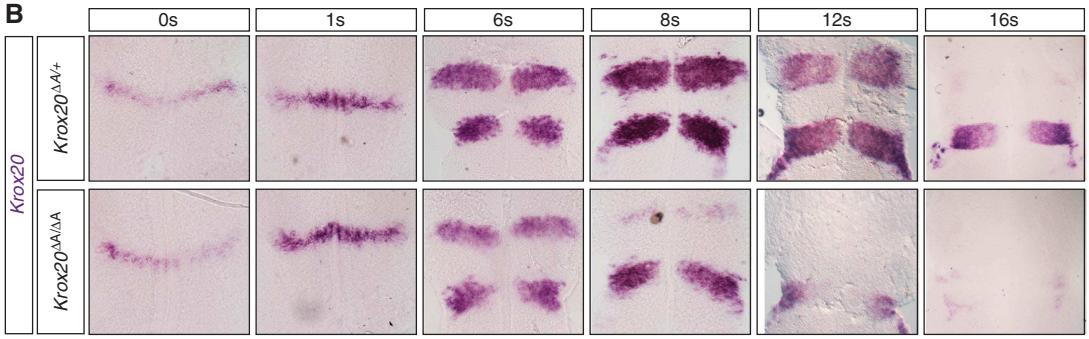
To investigate the consequences of this altered *Krox20* expression on hindbrain patterning, we analysed the expression of one of the Krox20 target genes, encoding the tyrosine kinase receptor EphA4 (Theil *et al*, 1998). *EphA4* is normally expressed at high relative levels in r3 and r5, and at low levels in r2 (Figure 1C; (Gilardi-Hebenstreit *et al*, 1992)). *EphA4* expression persists in r3 and r5 after *Krox20* is switched off (unpublished observations), indicating that at some point it becomes independent of Krox20 and is therefore a marker of commitment to the r3/r5 fate. In *Krox20*$^{\Delta A/\Delta A}$ embryos, the size of the domains of high *EphA4* expression is markedly reduced after 8s (Figure 1C). This suggests that transient expression of *Krox20* is sufficient to drive a limited number of cells into an r3/r5 fate, but that the Krox20 autoregulatory loop is required for obtaining odd-numbered rhombomeres of normal size. This observation was confirmed by direct analysis of the r2 and r4 territories, using an alkaline phosphatase reporter transgene specifically expressed in r2 (Studer *et al*, 1996) and *in-situ* hybridization against *Hoxb1* to reveal r4 (Figure 1E). This analysis shows the persistence of a reduced r3 territory at embryonic day 9. We also observed an increase in r4 size (Figure 1E), suggesting that lack of Krox20 autoregulation may lead to re-specification of cells normally fated to belong to odd-numbered rhombomeres.

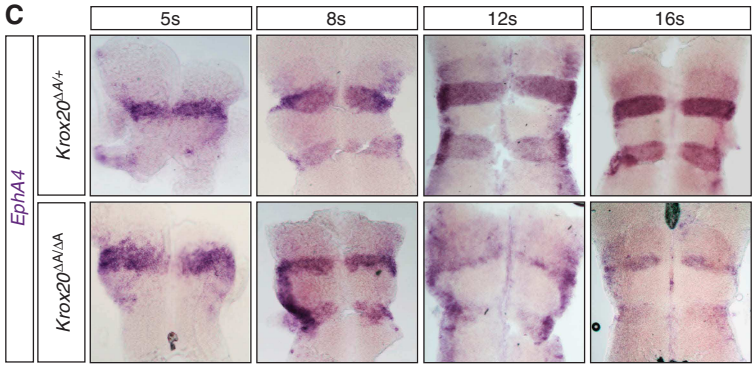
To demonstrate that the *Krox20*$^{\Delta A/\Delta A}$ phenotype specifically originates from a defect in *Krox20* autoregulation, we attempted a rescue by re-establishing a positive feedback loop based on the specific interaction between Krox20* and cA* (Supplementary Figure S1D–F). We verified that *Krox20*$^{A^*/A^*}$ and *Krox20*$^{\Delta A/\Delta A}$ embryos display similar phenotypes (Figure 1D). We then generated a transgenic mouse line, *Tg(cA:Krox20*)*, carrying the *Krox20** coding sequence under the control of the chick element A, which responds to mouse Krox20 and is active in r3 and r5 (Chomette *et al*, 2006). In *Tg(cA:Krox20*);Krox20*$^{A^*/A^*}$ embryos, hindbrain expression

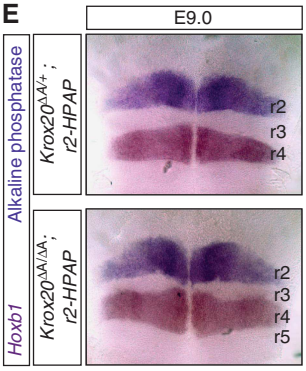
Figure 1 Element A is required for the maintenance of mouse *Krox20* expression. (**A**) Knockin into the A element and deletion strategy. The different alleles obtained after homologous Cre (targeting loxP sites) or Flp (targeting FRT and F3 sites) recombination are presented. cA* is a mutant form of chick element A that can only be bound by a mutant version of Krox20, termed Krox20* (see Supplementary Figure S1A–C). *Krox20* (**B**) and *EphA4* (**C**) *in-situ* hybridization performed on *Krox20*$^{\Delta A/+}$ (control) and *Krox20*$^{\Delta A/\Delta A}$ embryos at the indicated stages. (**D**) *EphA4 in-situ* hybridization performed on *Krox20*$^{A^*/A^*}$ embryos carrying or not the *Tg(cA:Krox20*)* transgene. (**E**) *Hoxb1 in-situ* hybridization (revealing r4) and alkaline phosphatase staining (revealing r2) performed on *Krox20*$^{A^*/A^*}$; *r2-HPAP* embryos at day 9 of embryonic development (25s approximately).

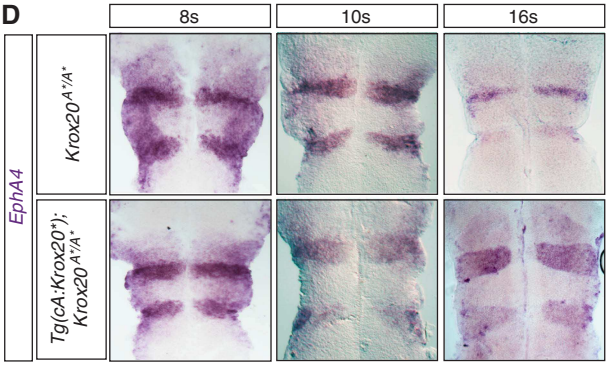
© 2013 EMBO and Macmillan Publishers Limited











© 2013 EMBO and Macmillan Publishers Limited

of *EphA4* was largely rescued (Figure 1D). In this situation, Krox20 is normally initiated from the endogenous locus and activates the exogenous element A, driving the expression of Krox20*; in turn, Krox20* activates element A* on the endogenous locus, leading to further production of Krox20 (Supplementary Figure S1F). The damaged, endogenous loop is rescued by a novel, indirect autoregulatory loop. This analysis demonstrates that the phenotype associated with the *Krox20*$^{\Delta A/\Delta A}$ mutation is exclusively due to a lack of *Krox20* autoregulation. Furthermore, our data establish that the activation of element A involves direct binding of Krox20, as mutations in the *cis*-acting element can be rescued by compensatory mutations in the *trans*-acting factor.

Krox20 autoregulation is conserved in zebrafish

To quantitatively analyse the features of Krox20 autoregulation, we used the zebrafish embryo. We first investigated the existence of an autoregulatory loop in this species. A fish line carrying a point mutation in the *krox20* coding sequence that abolishes Krox20 function (*krox20*^{fh227} allele (Monk *et al*, 2009)) and, therefore, autoregulation was used to perform a loss-of-function analysis. The mutation does not prevent activation of *krox20* expression in r3 and r5, but rapidly leads to its extinction from 6s in r3 and 10s in r5, whereas expression is maintained in both rhombomeres beyond 14s in WT embryos (Figure 2A–F and A′–F′). This phenotype indicates that Krox20 is required for maintaining its own expression. We then engineered a transgenic fish line, *Tg(hsp:mKrox20*$_{HA}$*)*, in which a HA-tagged murine version of *Krox20* is under the control of a heat-shock (HS) promoter. Using this system, we can modulate the amount of *mKrox20* produced in all embryonic cells by modifying the temperature or duration of the HS. HSs performed at increasing temperatures lead to progressive activation of endogenous *zkrox20* in r2, r4 and r6 (Figure 2G–J). Although expression of the transgene occurs in the entire embryo, efficient activation of the endogenous gene is essentially restricted to the hindbrain, presumably due to additional and unknown regulatory mechanisms. Together, these data establish the existence of a Krox20 autoregulatory loop in the zebrafish.

We were unable to find a fish orthologue of element A by an *in-silico* search based on nucleotide conservation. However, we identified a 1-kb DNA fragment that can drive specific expression of a reporter gene in r3 and r5 in the zebrafish (Supplementary Figure S2). This fragment contains a cluster of five potential Krox20-binding sites, is located upstream of the *krox20* gene, at a position corresponding approximately to that of the chick element A, and has been shown to co-immunoprecipitate with the H3K4m1 histone modification, which usually marks the enhancers (Aday *et al*, 2011). It is therefore likely to contain zebrafish element A.

Analysis of the molecular mechanisms governing *Krox20* expression

To evaluate the respective contributions of initiation and autoregulation to *krox20* expression, we measured *krox20* mRNA levels in WT and *krox20*$^{fh227/fh227}$ zebrafish embryos by

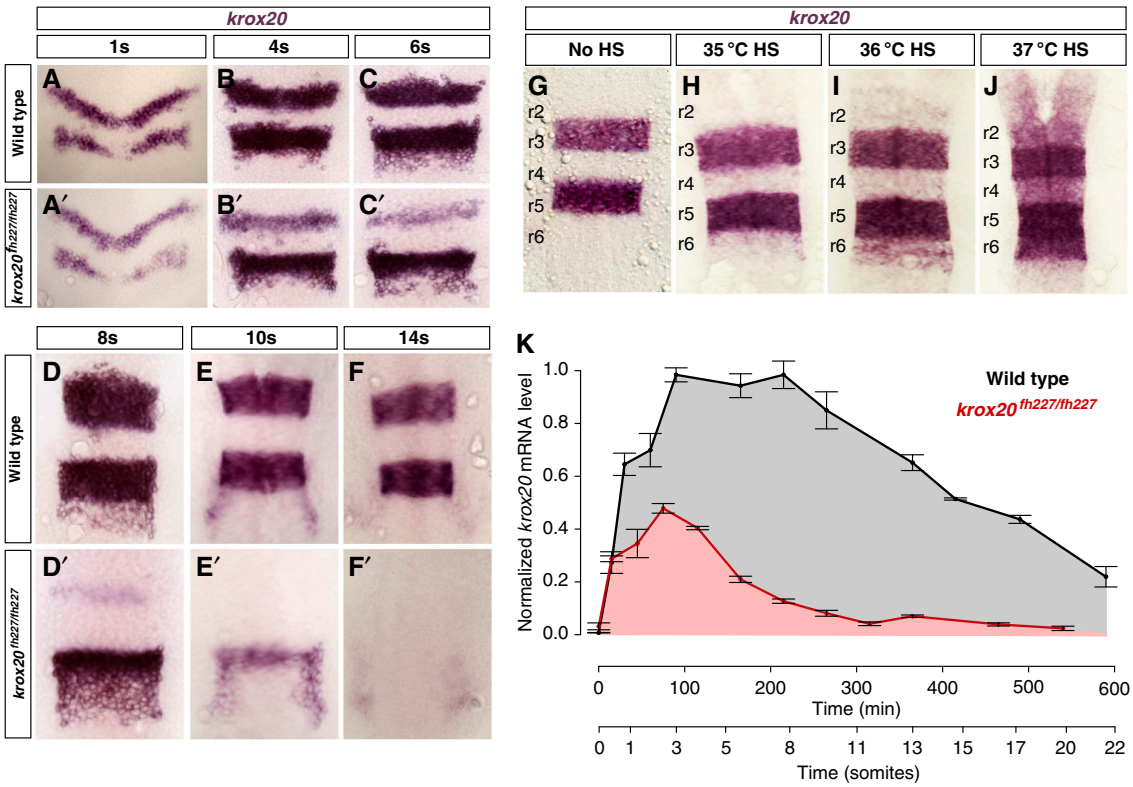


Figure 2 *Krox20* autoregulation in the zebrafish embryo. (**A–F** and **A′–F′**) *krox20 in-situ* hybridization performed on WT and *krox20*$^{fh227/fh227}$ embryos at the indicated stages. (**G–J**) *krox20 in-situ* hybridization on 15s *Tg(hsp:mKrox20*$_{HA}$*)* embryos after 10 min HS performed at 1s at the indicated temperatures. Rhombomeres (r) are indicated. (**K**) Time course of *krox20* mRNA level measured by RT–qPCR in WT (black curve) and *krox20*$^{fh227/fh227}$ (red curve) embryos. The data are normalized with the value of the WT plateau. Source data for this figure is available on the online supplementary information page.

© 2013 EMBO and Macmillan Publishers Limited

reverse transcriptase–quantitative PCR (RT–qPCR), between 100% epiboly and 22s (Figure 2K). During this period, *krox20* expression is restricted to the hindbrain and the measurements therefore correspond to the added levels of r3 and r5. Mutant expression reflects only the initiation process, whereas WT expression corresponds to the combined output of initiation and autoregulation. In the mutants, *krox20* mRNA rapidly accumulates, peaks around 3 s and then decays. The presence of the autoregulatory loop leads to a twofold increase in the maximal mRNA level and to an extension of the expression period, with a plateau between 3s and 8s, followed by a linear-like decline. Therefore, the initiator elements provide only a short pulse of *krox20* expression that is necessary to trigger the autoregulatory loop, which results in 4.7-fold higher dose of *krox20* mRNA during the 0–22s period (compare areas under the curves in Figure 2K).

Transcription systems often rely on cooperative transcription factor binding to DNA and on synergistic activation (also known as concerted recruitment) of the transcriptional machinery by multiple transcription factors (Georges *et al*, 2010). We investigated whether element A activity involves such mechanisms. *In-silico* analysis of the 416-bp chick element A revealed the presence of seven putative Krox20-binding sites (Figure 3A and (Chomette *et al*, 2006)). *In-vitro* binding of Krox20 to each of the sites was assessed by competitive electromobility shift assay (EMSA). Three of the sites were of high affinity and one of medium affinity (sites K2, K5, K7 and K6, respectively; Supplementary Figure S3). The

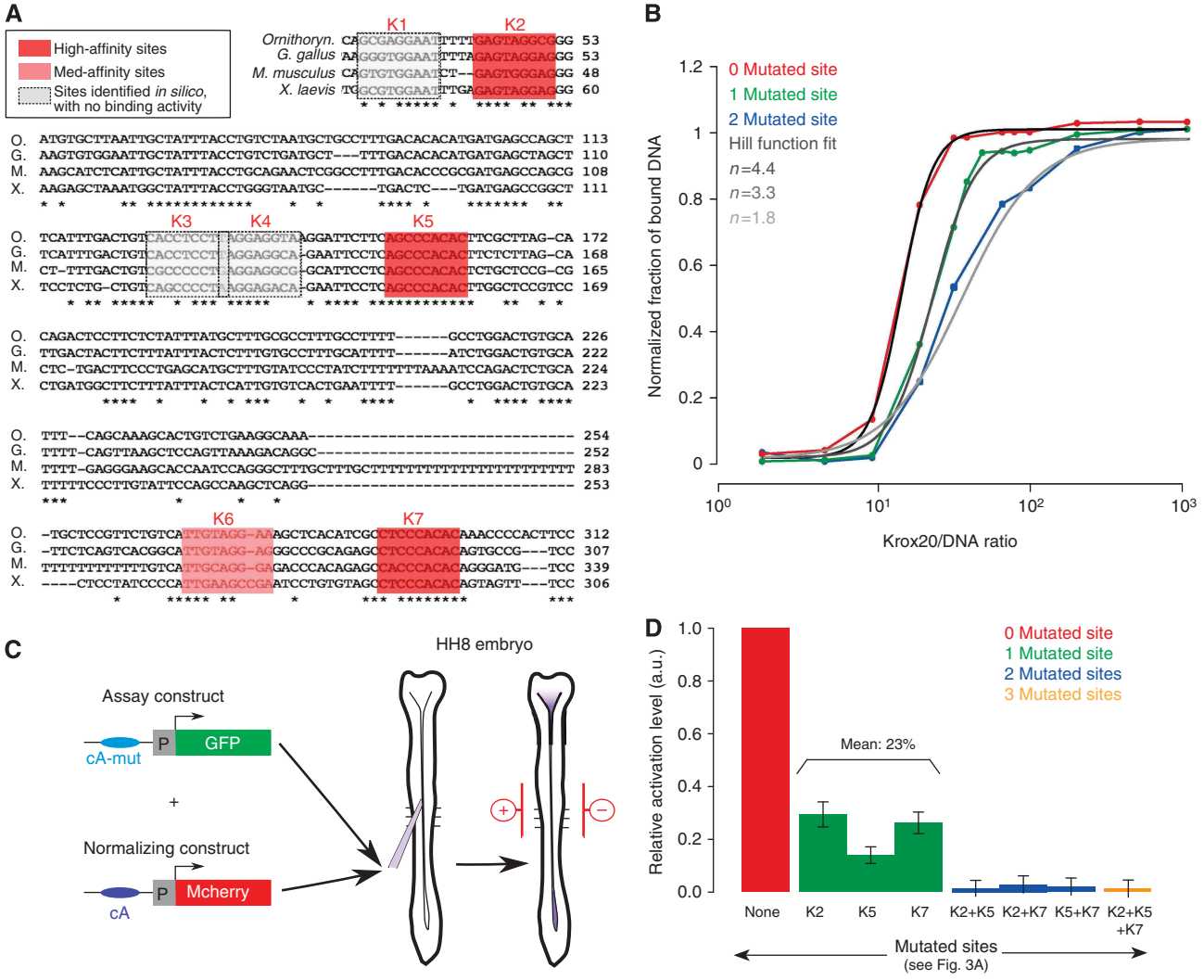


Figure 3 Binding cooperativity and synergy in transcriptional activation. (**A**) Alignment of element A nucleotide sequences from different vertebrate genomes (O, *Ornithorynchus*; G, *Gallus gallus*; M, *Mus muscullus*; X, *Xenopus laevis*). Stars indicate conserved nucleotides. Bona fide Krox20-binding sites in the chick enhancer are highlighted in red or pink. (**B**) Saturation curves of Krox20 binding on chick element A obtained by EMSA experiments. The experiment was performed on WT element A (red) and mutant variants carrying either one mutated Krox20 site (site K2, green) or two mutated sites (sites K2 and K5, blue). Fitting with a Hill function is shown (grey curves). (**C**) Experimental design of the chick electroporation assay. Two constructs were co-electroporated in the chick embryo neural tube at stage HH8: the assay construct, where WT or mutated element A drives *GFP* expression, and the normalization construct, where WT element A drives *mcherry* expression. In all experiments, the level of GFP was normalized with the level of mCherry fluorescence. (**D**) Quantification of the electroporation experiments. GFP levels of the different element A mutants (green, blue, orange columns) are displayed relative to WT (red column). Source data for this figure is available on the online supplementary information page.

© 2013 EMBO and Macmillan Publishers Limited

other putative sites had very low or no binding activity and were no longer considered. We introduced deleterious mutations into the three high-affinity Krox20-binding sites, either alone or in combination, and performed saturation experiments, where a fixed concentration of element A is incubated with increasing amounts of protein (Figure 3B). The data demonstrate strong cooperative binding. Fitting each curve with a Hill function (Figure 3B) provided Hill coefficients of 4.4 for WT element A, 3.2 and 3.3 for single mutants and 1.8 for the double mutant. These data are consistent with cooperative binding involving each of the four sites (Ma *et al*, 1996; Burz *et al*, 1998).

To evaluate synergy in transcriptional activation, we measured the relative activity of each element A mutant in a co-electroporation assay in the chick hindbrain, which allows an easy and quantitative comparison of the steady-state dynamics in the hindbrain of exogenous enhancer elements driving the expression of reporter genes (Chomette *et al*, 2006; Wassef *et al*, 2008). In this assay, expression of a reporter driven by element A is essentially restricted to r3 and r5 (Chomette *et al*, 2006). Two reporter plasmids were co-electroporated: a construct in which WT or mutant element A drives GFP expression and a normalization construct, in which mcherry is driven by WT element A (Figure 3C). Eighteen hours after electroporation, for each mutant the normalized fluorescence level associated with GFP in r3 and r5, relative to WT, was taken as a measure of the relative activity of the element. Single mutations reduce element A transcriptional activity to approximately one-fourth of the control, whereas combinations of two or three mutations abolish it completely (Figure 3D). The non-additive contributions of the binding sites demonstrate that Krox20 molecules bound to the enhancer activate transcription in a synergistic manner.

A stochastic model for *Krox20* transcriptional regulation

To study the mechanisms governing *Krox20* expression, we developed a stochastic mathematical model based on molecular dynamics. We modelled the dynamics of Krox20 mRNA and proteins, cooperative binding/unbinding of Krox20 proteins to the four binding sites of element A and synergy for transcriptional activation. We also modelled the transient initiation phase that allows activation of *Krox20* expression. As shown below, only a few mRNA molecules are involved in this activation and a stochastic approach was required to correctly account for large fluctuations.

The model is schematically represented in Figure 4A and is fully described in the Supplementary Information. *Krox20* expression is initially zero. From time $t = 0$ to t_I, in addition to autoregulation, an initiation process leads to the production of *Krox20* mRNA by a mechanism that is independent of autoregulation. This mechanism can be either endogenous (through the activity of initiator *cis*-acting elements), or exogenous (upon HS in the *Tg(hsp:mKrox20$_{HA}$)* line). *Krox20* mRNA is produced by this process at a Poissonian rate Φ_I. This Krox20-independent initiation process is responsible for an initial production of Krox20 that can binds to element A and activates the autoregulatory loop. Krox20 protein production due to a single mRNA occurs at a Poissonian translation rate ϕ. mRNA molecules are degraded with a Poissonian rate Ψ and proteins with a rate ψ. Proteins bind to element A, thereby also activating and modulating mRNA production. Element A is modelled with four equivalent binding sites and can be in five states ($s = 0, 1, 2, 3, 4$), depending on the number of bound Krox20 proteins. We implemented cooperativity in unbinding from element A using the state-dependent unbinding rates $\mu_s = \mu/\gamma_s$, where μ is an overall rate constant and the numbers γ_s describe the modulations due to the state of element A (see section 2.8 in Supplementary Information). In contrast, we assumed that binding is not affected by the state of element A and used a single forward binding rate λ. We implemented transcriptional synergy in element A activity by using a state-dependent mRNA production rate $\Phi_{A,s} = \Phi_A \xi_s$, where Φ_A is the maximal production rate when all four sites are bound and the numbers ξ_s describe the modulation by the state of element A. Finally, the model takes into account the two *Krox20* alleles, the parameters Φ_I and $\Phi_{A,s}$ corresponding to a single allele.

The model is characterized by three dynamic variables: the number *m* of *Krox20* mRNA molecules, the number *n* of proteins and the state *s* of element A. To compute these variables, we derived the Master equations (equation (1) in Supplementary Information) to obtain the joint probability $p_s(m,n,t)$ to find *m* mRNA molecules, *n* proteins and element A in state *s* at time t. To separate initiation from autoregulation, we distinguished the numbers of Krox20 mRNAs and proteins produced by initiation, Krox20^I, and by element A, Krox20^A. The model depends on 14 parameters listed in Supplementary Table S1. Section 2 of the Supplementary Information describes how 10 of these parameters were determined experimentally and 2 derived from the literature. To obtain the remaining two parameters, we developed an indirect parameterization approach, where we compared simulations with experimental data.

This approach allowed us to derive both the dynamics of *Krox20* expression and the fluctuations from elementary molecular events. The fraction of cells that eventually activate element A depends on the initiation period and therefore cannot be computed from steady-state analysis. To study the time evolution of the system, we numerically solved the Master equation to obtain $p_s(m,n,t)$, from which we computed all statistical properties. In addition, to study the evolution of a single cell, we used the Gillespie algorithm (Gillespie, 1976; Tajbakhsh *et al*, 2009; Graham *et al*, 2010) to simulate the molecular reactions underlying the Master equations. We refer to these single cell traces as 'molecular dynamics simulations'.

The model accounts for the dynamics of *krox20* activation

We first implemented the model to study the dynamics of *krox20* expression. In Figure 4B, we compare the measured time course of *krox20* mRNA expression in WT and *krox20$^{fh227/fh227}$* embryos (black curves) with numerical simulations (red curves). In the WT, the activation of element A is saturating (see part 2.3 of Supplementary Information) and we used the plateau value as a reference for the maximal mRNA level to normalize the experimental data (Figure 2K).

© 2013 EMBO and Macmillan Publishers Limited

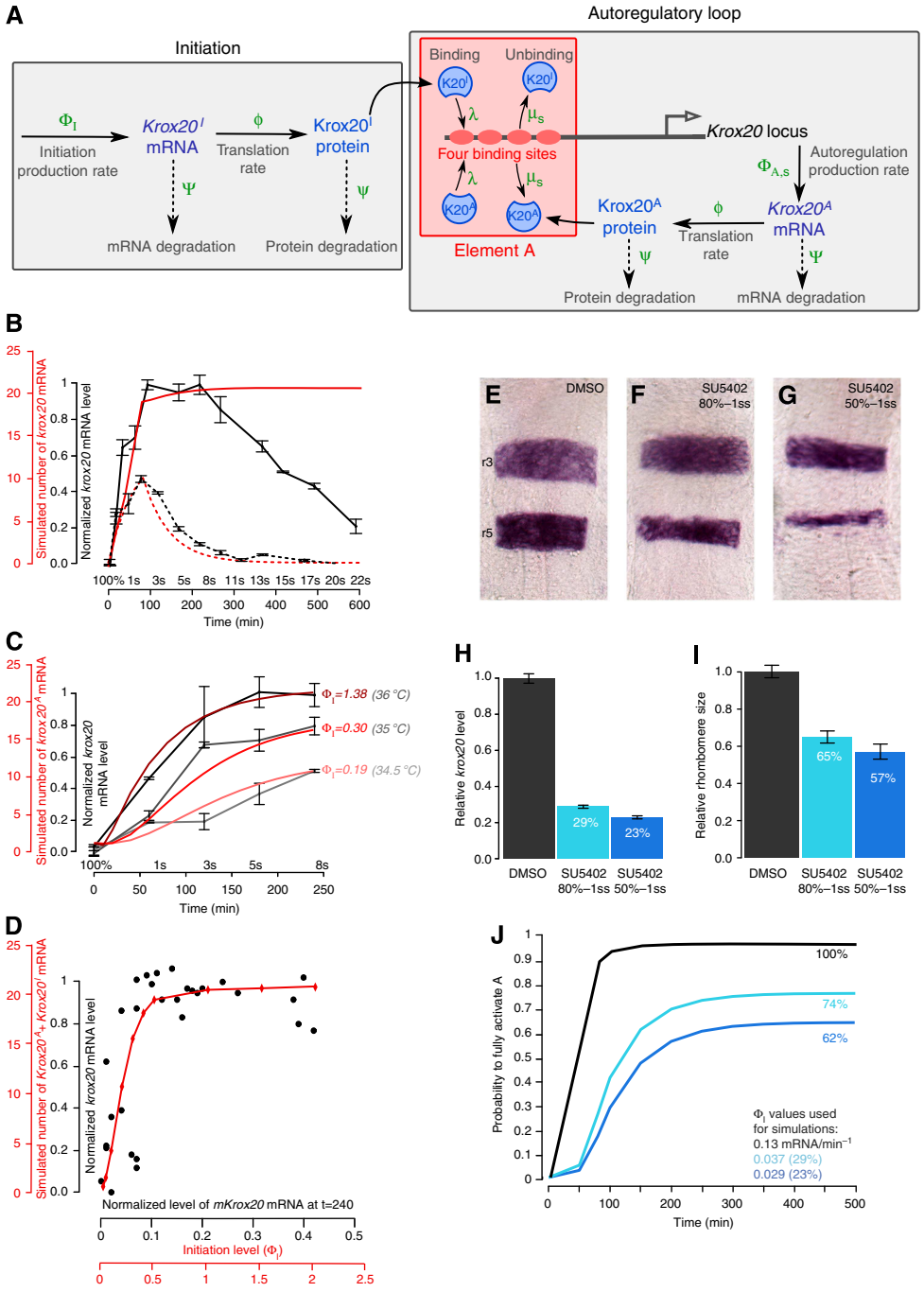


Figure 4 A quantitative model of *Krox20* expression. (**A**) Schematic representation of the molecular events used to model *Krox20* expression. *Krox20* expression proceeds in two phases: during a first phase that lasts for t_I minutes, Krox20 production involves both a Krox20-independent initiation process and the autoregulatory loop, whereas only autoregulation is maintained afterwards. For further details, see the text. (**B**) Comparison of simulations (red) and experimental data (black) of *Krox20* expression time courses in developing embryos with (solid curves) and without (dashed curves) activity of the autoregulatory loop. For the simulations, we used 80 min initiation time and rates $\Phi_I \approx 0.7\Phi_A = 0.13$ mRNA/min. (**C**) Time courses of *krox20* expression obtained from experimental data (grey curves) and simulations (red curves) following HS at the indicated temperatures. To obtain the experimental level of *krox20* mRNA present in even-numbered rhombomeres, we subtracted the level of *krox20* mRNA present in non-transgenic fish, as determined in **B**. The HS simulations were performed with 10 min initiation time and the indicated initiation rates. (**D**) Experimental (black dots) and simulated (red curve) input/output relationship of element A activation extracted from the HS data. Each point (black dot) corresponds to the levels of *mKrox20* and *zkrox20* RNA measured by RT–qPCR 4 h after HS in a single embryo. Embryos were subjected to 10 min HS ranging from 33 to 39 °C. (**E**–**G**) *krox20 in-situ* hybridization on 15s WT zebrafish embryos treated under two conditions with the FGFR inhibitor SU5402 (**F**, **G**) or mock-treated (**E**). (**H**, **I**) Diagrams showing the relative *krox20* mRNA levels in *krox20*$^{fh227/fh227}$ embryos measured by RT–qPCR (**H**) and the relative area of r3 + r5 in WT embryos treated under the two SU5402 conditions compared with mock treatment (DMSO). (**J**) Simulations of the probability to fully activate A (p_4) when the initiation rate is altered. The simulations were performed with 80 min initiation time, and the black curve (labelled 100%) was obtained with the control rates $\Phi_I = 0.13$ mRNA/min. For the other two conditions, Φ_I was reduced to 29 and 23% of the control value, corresponding to the experimental measurements in H. The steady-state level for $\Phi_I = 0.13$ mRNA/min is defined as 100%. Source data for this figure is available on the online supplementary information page.

© 2013 EMBO and Macmillan Publishers Limited

Figure 4B was obtained from numerical simulations of the joint probability $p_s(m,n,t)$ (equation 1 in the Supplementary Information), from which we computed the time course of the mean number of mRNA per cell. To compare experiments with simulations, we scaled the number of mRNA/cell with the steady-state number of mRNA/cell provided by a fully activated element A. For the WT simulations, we used two alleles per cell; we estimated $\Phi_I \approx 0.7\Phi_A = 0.13$ mRNA/min for each allele, and we assumed that initiation in r3 and r5 was the same and lasted for 80 min.

We found that the model accounted for the dynamics *krox20* expression in the mutant and its upregulation and early plateau phases in the WT (Figure 4B). However, it did not account for the decrease observed after 8s in the WT. This difference will be discussed in detail below.

Our experimental measurements of *krox20* mRNA levels by RT–qPCR were obtained from whole embryos and, therefore, correspond to the summation of the levels in r3 and r5. The dynamics of *krox20* expression in the absence of autoregulation are shifted in time between r3 and r5 (Figure 2A′–F′). Therefore, our model with a single input represents a simplification of this situation. However, we have performed numerical simulations based on our model with two shifted inputs of the same Φ_I value for r3 and r5, and the differences with simulations performed with a single input were marginal (Supplementary Figure S4). For simplicity, we therefore performed all following simulations with a single input.

To investigate *Krox20* expression in non-saturating conditions, we analysed the *Tg(hsp:mKrox20$_{HA}$)* transgenic line before the decline of endogenous *zkrox20* expression. In this case, the level of input *mKrox20* mRNA can be tuned experimentally by varying the HS temperature and can be discriminated from the level of *zkrox20* mRNA by RT–qPCR. We performed 10-min HSs at three temperatures (34.5, 35 and 36 °C) at 100% epiboly and measured the level of *mKrox20* mRNA at time $t = 0$ (end of the HS) and the level of *zkrox20* mRNA in even-numbered rhombomeres from $t = 0$ to 240 min (Figure 4C, grey curves). Experimental mRNA levels were normalized with the saturating level obtained at 36 °C. We next compared the experimental data with model predictions where we varied the initiation production rate Φ_I. As the experimental initiation rates Φ_I are not known, we first estimated one reference value for Φ_I (0.19 mRNA/min) by fitting the experimental curve at 34.5 °C (Figure 4C). The Φ_I values at 35 °C (0.30 mRNA/min) and 36 °C (1.38 mRNA/min) were computed such that the ratios of Φ_I values equal the ratios of experimental input values. Both simulations obtained with these calculated Φ_I values were also in agreement with the experimental data (Figure 4C), showing that the model correctly predicts *krox20* expression in both saturating (36 °C) and non-saturating conditions (34.5 °C and 35 °C).

Finally, we used the HS data to extract a dose–response curve giving the levels of *zkrox20* mRNA as a function of the levels of *mKrox20* mRNA (Figure 4D, black dots). Both mRNA levels were measured 240 min after HS, i.e., close to the steady state according to the kinetics shown in Figure 4C. A simulated dose–response curve (in red in Figure 4D) was obtained by plotting the number of *Krox20^A* mRNA molecules as a function of the initiation rate Φ_I. The agreement between the simulation and the experimental data shows that the model correctly predicts the input–output relationship of this system.

The feedback loop underpins a bistable switch and induces a bimodal cell distribution

We used the model to analyse the autoregulatory loop. We first investigated in which conditions positive feedback leads to a bistable switch. We found that a stable state with a high *Krox20* expression level exists when the product of the production rate Φ_A and the ratio $\beta = \lambda/\mu$ exceeds a threshold value (Supplementary Figure S5 and S6 and section 3.2 of Supplementary Information). With one or two alleles and a production rate $\Phi_A = 0.18$/min (Supplementary Information, section 2.4), this results in the minimal values $\beta_{min} = 0.13$ or $\beta_{min} = 0.065$, respectively. We estimated a value of 0.20 for β (Supplementary Information, section 3.2.2), which is largely above the minimal values. Therefore, stable maintenance of the loop is guaranteed in both WT and heterozygote.

We then studied how the initiation phase affects *Krox20* expression. We find that the fraction of the cells that commit to a Krox20-positive fate can be modulated by varying either the initiation strength Φ_I or its duration t_I (Supplementary Figure S7). Furthermore, the gradual splitting of the original homogeneous cell population leads to two different homogeneous populations, as the stochastic properties of the initiation process do not introduce additional variability (Supplementary Figure S7D, H). We illustrate these findings in Figure 5 by dissecting the dynamics of *Krox20* activation for values of Φ_I corresponding to the HS temperature conditions presented in Figure 4C. The results presented in Figure 5A–F were obtained from numerical simulations of the joint probability $p_s(m,n,t)$ (equation 1 in the Supplementary Information). The time course of the probability $p_s(t)$ of finding element A in state *s* (referred to as 'state probability', Figure 5A–C) reveals that element A is either fully activated ($s = 4$, four Krox20 proteins bound) or fully deactivated ($s = 0$, no protein bound) and the probability of the intermediate states are negligible. As a consequence, the steady state value of $p_4(t)$ for large time periods yields the fraction of cells that select a Krox20-positive fate. Figure 5A shows that an average initial production of about four mRNA molecules per cell ($0.19\,\text{mRNA/min/allele} \times 10\,\text{min} \times 2$ alleles ≈ 4 mRNA) induces the Krox20-positive fate with 60% probability (see also Supplementary Figure S7I). Thus, the critical region that determines cell fate involves the action of only a few mRNA molecules, leading to a highly stochastic regime. The heat maps of the time-dependent probability for the number of Krox20 proteins within a cell show two strands corresponding to the bimodal distribution of Krox20-positive versus Krox20-negative cells (Figure 5D–F). Increasing the initiation rate Φ_I changes the occupancy of the large strand, but not its location nor width, demonstrating that the initiation does not affect the characteristics of Krox20-positive cells. For large Φ_I, almost all cells express *Krox20* and the probability distribution is essentially unimodal (Figure 5F). Figure 5G–I show the time evolution of *Krox20* expression in individual cells, obtained from the molecular dynamics simulations. Cells either evolve toward a stable high-level *Krox20* expression or the expression vanishes. For cells that have reached high-level expression, this state is

© 2013 EMBO and Macmillan Publishers Limited

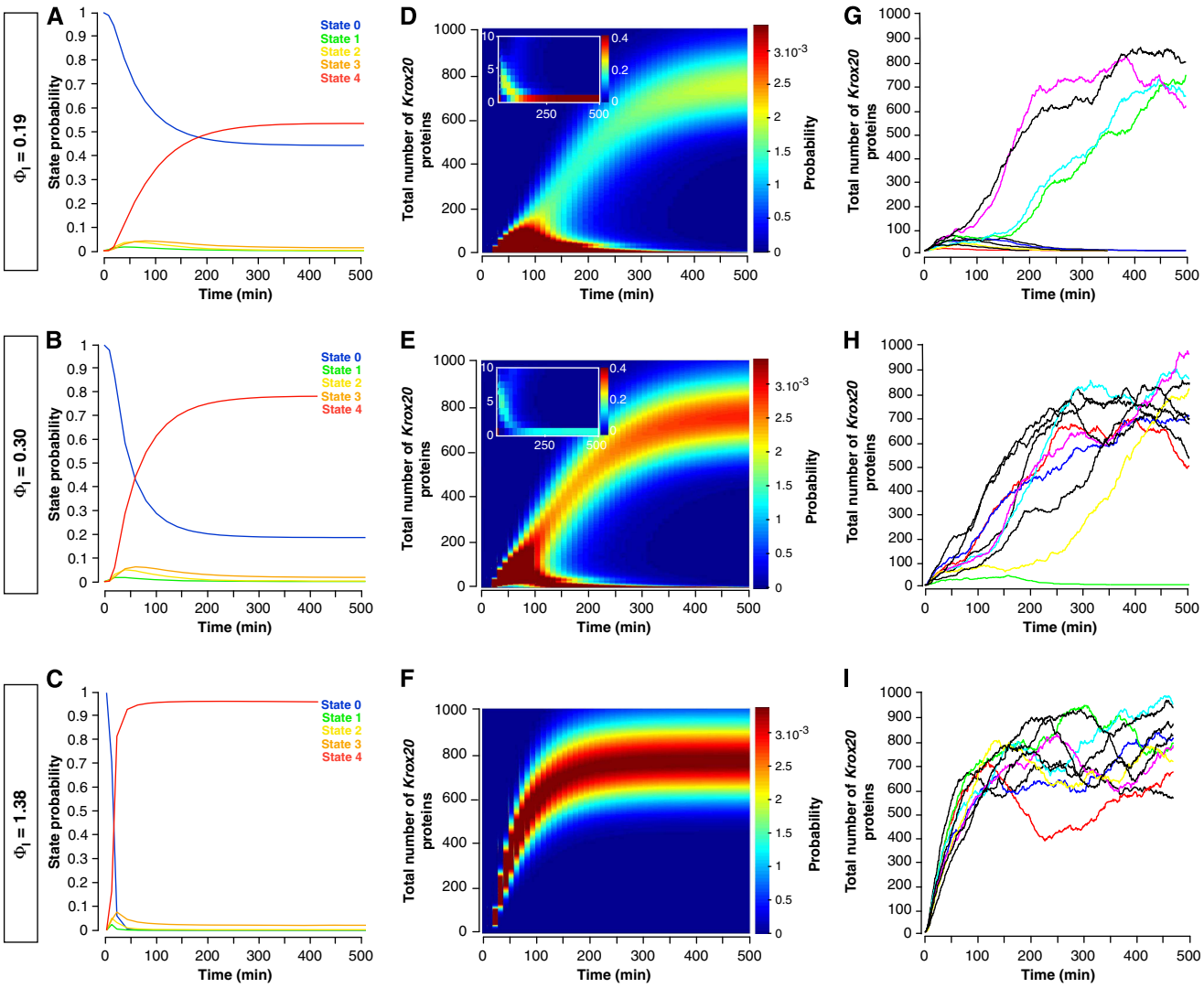


Figure 5 Dynamics of *Krox20* activation derived from numerical simulations. (**A–C**) Time-dependent probability of the different states of element A ($s = 0, 1, 2, 3$ or 4) for three different levels of initiation (Φ_I), corresponding to the HS conditions presented in Figure 4C with 10 min initiation time. (**D–F**) Heat maps for the evolution of the probability distribution for the total number of Krox20 proteins in a cell with initiation as in **A–C**. Insets are zooms for low protein numbers. (**G–I**) Molecular dynamics simulations for the evolution of the stochastic number of Krox20 proteins in single cells with initiation as in **A–C**. Each panel shows the time course in 10 cells. The results shown in panels **A–F** were obtained by integrating the Master equation 1 in the Supplementary Information. The molecular dynamics simulations shown in **G–I** are obtained from the Master equation 1, using the Gillespie algorithm.

maintained. In the critical region of initiation values, the evolution is highly stochastic, leading to a large variability in the transition time to the final state (Figure 5G and H).

Finally, we explored how changes in binding cooperativity and synergistic promoter activation affect the bistable behaviour of the system. We find that both are required to obtain the following properties (see section 3.2.5 of Supplementary Information and Supplementary Figure S8): (1) efficient activation of the loop; (2) existence of a bimodal state with either fully activated or fully deactivated promoter; (3) the fraction of cells that commit to a Krox20-positive fate is sensitive to and gradually modulated by low initiation levels.

In conclusion, our model shows that an initial fluctuating Krox20 signal is converted by the autoregulatory loop into a bistable behaviour conditioning cell fate choice, and the distribution between the two fates is determined by the strength of the initiation signal and depends on cooperative binding and synergistic production.

Experimental demonstration of bimodality

Bimodality is predicted by the model for low-level inputs. To verify this experimentally, we used a zebrafish transgenic line carrying a reporter construct in which a histone h2b-mcherry fusion protein is placed under the control of chick element A, *Tg(cA:h2b-mcherry)*. In this condition, *mcherry* is expressed in r3 and r5, and localizes in the cell nucleus. This reporter line was crossed with the *Tg(hsp:mKrox20$_{HA}$)* line (Figure 6A). As expected, in the absence of HS, we observe specific and homogeneous nuclear mCherry fluorescence in r3 and r5 cells, in about half of the embryos (Figure 6B and C). These positive

© 2013 EMBO and Macmillan Publishers Limited

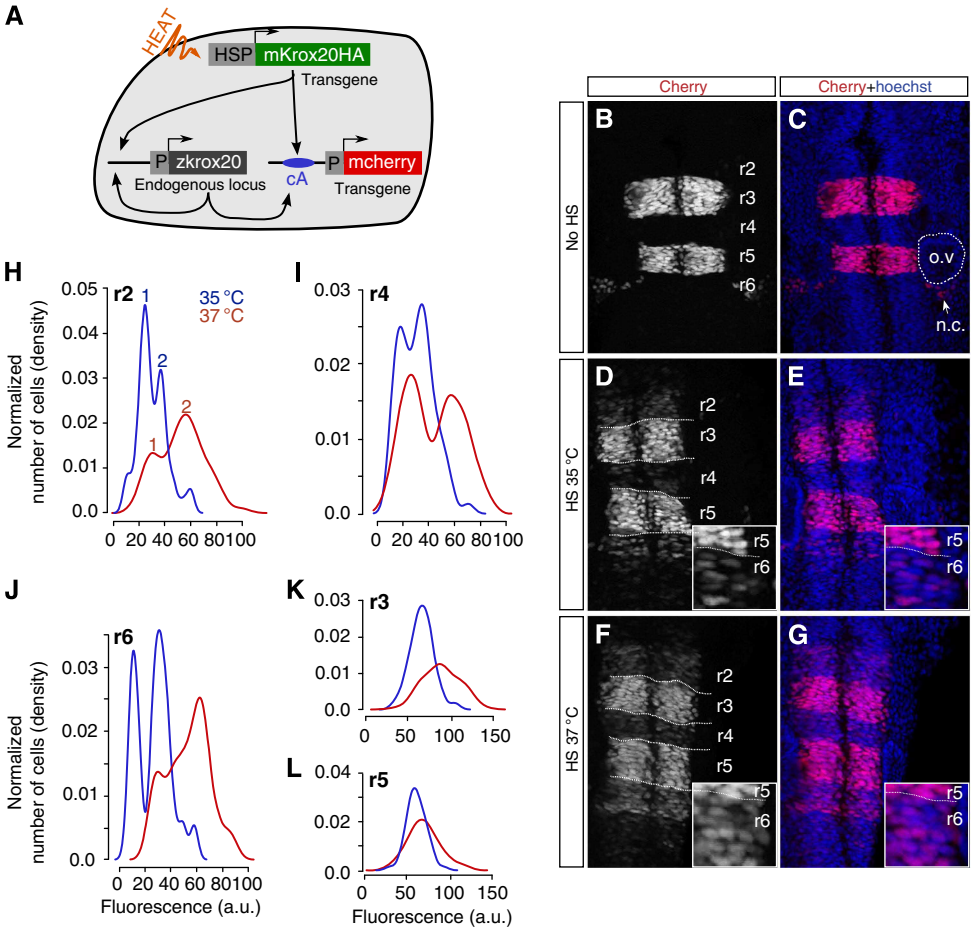


Figure 6 Experimental manifestation of bimodality. (**A**) Schematic representation of the system designed to measure element A activity at the single-cell level. HSs in double transgenic *Tg(hsp:mKrox20$_{HA}$);Tg(cA:h2b-mcherry)* embryos produced an ectopic activation of element A. This activation was quantified by measuring nuclear mCherry fluorescence in single cells. (**B**–**G**) Confocal pictures showing mCherry and Hoechst 33342 fluorescence in double transgenic embryos that did not undergo HS (**B**,**C**) or were HS treated for 10 min at 35 °C (**D**,**E**) or 37 °C (**F**,**G**) at 5s. (**H**–**L**) Distribution of the number of cells according to their level of mCherry fluorescence in r2 (**H**), r4 (**I**), r6 (**J**), r3 (**K**) and r5 (**L**) measured in the 35 °C (blue curve) and 37 °C (red curve) HS experiments; o.v., otic vesicle; n.c., neural crest.

embryos correspond to those having inherited the *Tg(cA:h2b-mcherry)* transgene. Approximately 25% of the HS embryos (35 °C or 37 °C) showed mCherry fluorescence 4 h later in even-numbered rhombomeres (Figure 6D–G). They correspond to embryos having inherited both transgenes. The level of fluorescence in each nucleus was quantified and the number of cells within each rhombomere was plotted according to the level of fluorescence (Figure 6H and I). No significant modification in the level and distribution of fluorescence was observed in r3 and r5 upon HS and the distribution was found unimodal in these rhombomeres (Figure 6K and L; bimodality tests: $P>0.1$ for r3 and r5 at 35 °C and 37 °C, respectively). In contrast, in even-numbered rhombomeres the distribution of fluorescence revealed the existence of two peaks (Figure 6H–J; bimodality tests: r2 35 °C: $P<0.05$, 37 °C: $P<0.01$; r4 35 °C: $P<0.05$, 37 °C: $P<0.01$; r6 35 °C: $P<0.001$, 37 °C: $P<0.01$). Upon 37 °C HS, the two peaks are displaced toward higher levels of fluorescence than at 35 °C (Figure 6H–J). We interpret the highest fluorescence peak (peak 2 in Figure 6H) as corresponding to the population of cells that have stably engaged into the autoregulatory process. In contrast, peak 1 corresponds to cells that have failed to do so, although mKrox20 has transiently activated *mcherry* expression, explaining why the level of fluorescence is not null. We confirmed these interpretations using numerical simulations. For this purpose, the mathematical model was modified to introduce the *Tg(cA:h2b-mcherry)* transgene, and to take into account the large stability of the mCherry protein (in contrast to Krox20; Nadine Peyriéras, personal communication) and the possible different accessibilities of element A in its transgenic form versus endogenous element A. With these modifications, the model accounts for the experimental cell distributions, including the displacement of the peaks upon increasing initiation levels and the non-zero position of peak 1 (Supplementary Figure S9). The stability of the mCherry also explains its reduced level in even- as compared with odd-numbered rhombomeres, where expression is initiated earlier.

In conclusion, our experimental results confirm that the positive feedback loop transforms a transient pulse of Krox20 into a bimodal cell distribution in the hindbrain. As expression of *Krox20* determines cell fate, this feature is an essential function of the autoregulatory loop.

© 2013 EMBO and Macmillan Publishers Limited

Rhombomere size is buffered against fluctuations in the initiation signal

Simulations show that the fraction of cells that commit to the Krox20-positive fate (probability to fully activate element A) saturates when the initiation rate Φ_I is higher than the value $\Phi_I \approx 0.3\Phi_A$ (Supplementary Figure S7B). In WT embryos, we estimated $\Phi_I \approx 0.7\Phi_A$ (see section 2.5 in Supplementary Information), suggesting that the initiation stimulus is saturating and that even a 50% reduction, as in the heterozygous situation, does not significantly change the cell fate distribution. Further, reductions in the initiation signal almost linearly affect the fraction of positive cells (Supplementary Figure S7B) because of the strong impact of fluctuations at low initiation levels. Hence, there is no threshold-like behaviour and the fraction of positive cells is strongly buffered against variations in the initiation signal. As the sizes of r3 and r5 reflect the fraction of cells that commit to a Krox20-positive fate, they are also protected.

We tested this model prediction experimentally. We have previously shown that Fgf signalling in the hindbrain regulates the level of *Krox20* initiation (Chomette *et al*, 2006; Labalette *et al*, 2011). To investigate the effect of variation in the initiation signal on the probability of element A activation, we quantified r3 + r5 size and *krox20* mRNA initiation levels upon variations in Fgf signalling. For this purpose, we used a drug, SU5402, which can be added to the embryo medium and acts as a specific inhibitor of FGF receptors and, therefore, prevents FGF signalling. We compared embryos treated with SU5402 with those that were mock-treated. We used *krox20*$^{fh227/fh227}$ embryos to measure the level of *krox20* mRNA at 5s by RT–qPCR, corresponding only to initiation, and WT siblings to estimate the area r3 + r5 at 15s (steady-state condition). Short and long treatments with SU5402 were performed, affecting FGF signalling from different initial stages during development (Figure 4E–G). Treatments led to reductions of the initiation stimulus to 29 and 23% of the control (Figure 4H) and to reductions of the r3 + r5 area to 65 and 57% of the control, respectively (Figure 4I). Simulations with reductions in Φ_I to 29 and 23% of the control value $\Phi_I = 0.13$ mRNA/min predict that the probability to fully activate element A is reduced to 74 and 62%, respectively (Figure 4J), consistent with the experimental data.

In conclusion, this analysis reveals that autoregulation turns an initiation signal into a fraction of Krox20-positive cells that determines rhombomere size and provides robustness by dampening fluctuations in the initiation.

Destabilization of the autoregulatory loop by repressor molecules

Experimentally, we observe that *krox20* expression starts to decline with linear kinetics from 8s and is extinguished at around 25s, whereas the model predicts that the autoregulatory loop is stable once established (Figure 4B). We hypothesized that this progressive loss of expression originates from a modification in one of the parameters of the loop around 8s. We, therefore, systematically altered parameters in the model and compared simulations with experimental data: we tested modifications in Φ_A, the translation rate ϕ, the mRNA and protein degradation rates Ψ and ψ, and the effect of masking one of the binding sites. None of these modifications reproduced the experimental data (Supplementary Figure S10). However, we find that modifying the ratio $\beta = \lambda/\mu$ induced a linear decrease in *Krox20* expression compatible with experimental data (Figure 7A). The parameter β controls the amount of Krox20 protein necessary to maintain element A fully activated (Supplementary Figure S5B–D) and reflects the stability of the interaction of Krox20 with element A *in vivo*.

Such a change in Krox20 binding may result from either a chromatin modification or the appearance of a repressor that affects the interaction between Krox20 and element A. We have recently shown that the transcriptional repressors Nlz1 and Nlz2 can antagonize the activity of element A in zebrafish embryos (Labalette *et al*, in preparation). As *nlz1* and *nlz2* expression becomes reinforced in r3 and r5 during somitogenesis (Labalette *et al*, in preparation), these factors may modify the *Krox20* autoregulatory system to trigger the decrease observed beyond 8s. To address this issue, zebrafish embryos were co-injected with morpholinos against *nlz1* and *nlz2* mRNAs and a time course of *krox20* expression was performed by semi-quantitative fluorescent *in-situ* hybridization. This analysis shows that the knockdown of Nlz leads to a reduction in the slope of declining *krox20* expression of ~2.1- and 6-fold in r3 and r5, respectively (Figure 7B and C). This indicates that the Nlz factors have a major role in the destabilization of the autoregulatory loop and the loss of *Krox20* expression.

Discussion

In this study, we demonstrate that a direct, positive autoregulatory loop is required for amplification and maintenance of the expression of Krox20. The loop relies on a *cis*-acting element containing four Krox20-binding sites, whose function was established by a mouse knockout. We develop a mathematical model that integrates the molecular characteristics of the activation process, accounts for their stochastic nature and is constrained by quantitative data obtained *in vivo* and *in vitro*. Combination of computer simulations and experimental analyses allowed us to reach a number of major conclusions: (i) the positive feedback loop underpins a bistable switch that turns a transient input into cell fate commitment; (ii) the cell distribution between the two fates is controlled by the duration and strength of the input signal, and is regulated by the expression of only a few mRNA molecules; (iii) the transient input splits one initially homogeneous population into two different homogeneous populations; (iv) binding cooperativity and synergistic activation of transcription generate a system that is reliably controlled by low input level of activator; (v) r3 and r5 size reflects the fraction of cell that commit to a Krox20-positive fate and is strongly buffered against variations in input level size; (vi) the progressive extinction of *Krox20* expression in the hindbrain involves a destabilization of the loop by repressor molecules.

This work allows to understand at a molecular level how Krox20 activation leads to an unambiguous cell fate decision and controls the sizes of the rhombomeres. These processes are critical to vertebrate hindbrain segmentation.

© 2013 EMBO and Macmillan Publishers Limited

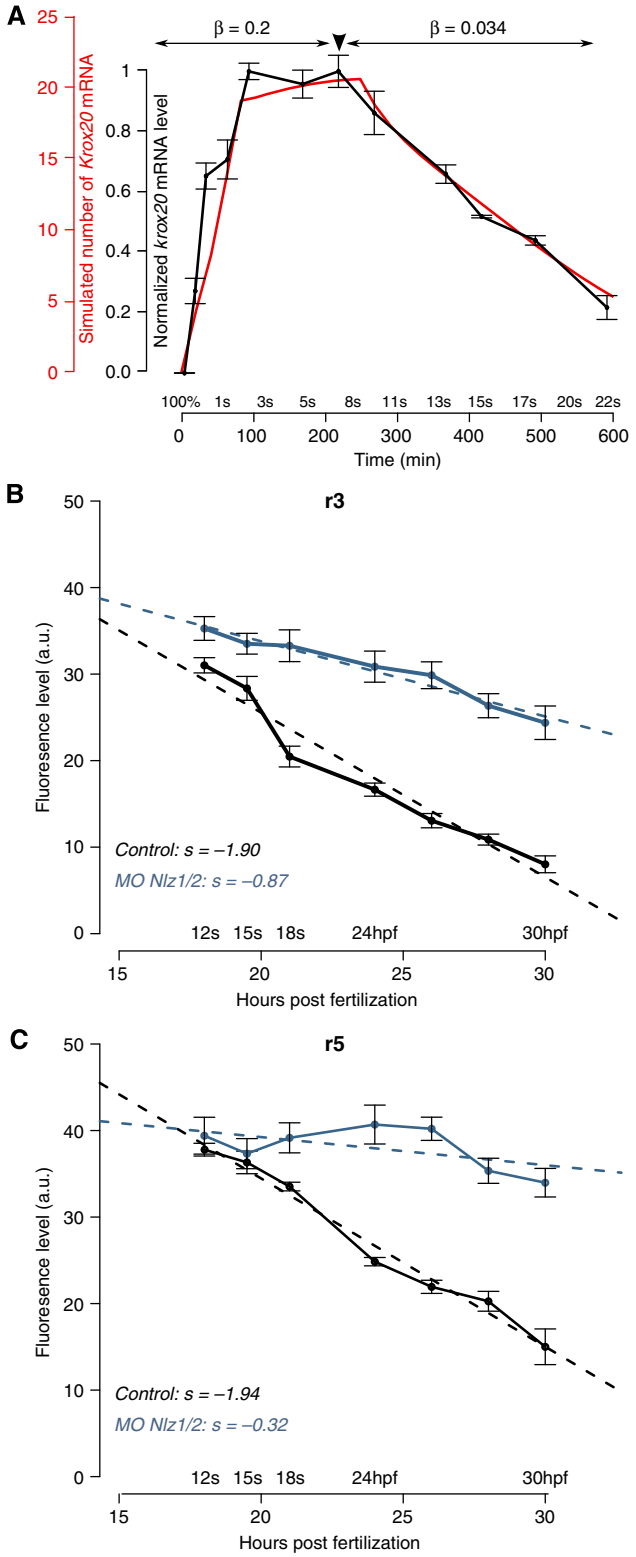


Figure 7 Role of Nlz factors in the destabilization of the autoregulatory loop. (**A**) The experimental time course of *krox20* expression (grey) can be fully recapitulated by the model if a 6.2-fold reduction in the β-factor is introduced at t = 250 min (red). (**B**, **C**) Time course of *krox20* expression in r3 (**B**) and r5 (**C**) measured by fluorescent *in-situ* hybridization in embryos injected with control (grey) or *nlz1/2* morpholinos (blue). hpf, hours post fertilization; s, slope. Source data for this figure is available on the online supplementary information page.

Autoregulation turns a transient input into a fate commitment

In mouse *Krox20*$^{\Delta A/\Delta A}$ and zebrafish *krox20*$^{fh227/fh227}$ embryos, where the autoregulatory loop is non-functional, *Krox20* expression reaches a lower plateau level and is more transient than in WT embryos, leading to an approximately fivefold reduction of the dose. This reduction results in a 2.8 ± 0.2-fold decrease in the number of cells positive for the marker EphA4 at 16s (Figure 1C). Furthermore, these EphA4-positive cells do not all maintain an odd-numbered identity, as their relative number, compared with WT, decreases between 10s and 16s (unpublished observations). Therefore, the absence of autoregulation leads to a significant reduction in the size of odd-numbered rhombomeres. We conclude that the positive feedback loop generates an amplification that is essential for the conversion of a transient input into a stable fate commitment.

A bistable switch underlies hindbrain patterning

Hindbrain segmentation is characterized by homogeneous levels of *Krox20* expression within alternating Krox20-positive and -negative rhombomeres (Schneider-Maunoury *et al*, 1993; Giudicelli *et al*, 2001; Cooke and Moens, 2002). Bistability has been proposed as a means to establish such a feature (Meinhardt, 1978; Ferrell, 2002; Lopes *et al*, 2008; Zhang *et al*, 2012). Our study demonstrates that the Krox20 autoregulatory loop provides a bistable switch generating a bimodal cell distribution with homogeneous levels of *Krox20* expression. To reach this conclusion, we developed a stochastic mathematical model based on molecular dynamics that depends on 14 parameters listed in Supplementary Table S1. We determined most of the parameter values either directly using our experimental results, or by comparing simulations with our experimental results. For two parameters, the *Krox20* mRNA production rate and the protein translation rate, we relied on the data published by Schwanhäusser *et al* (2011; 2013); see section 2.4 and 2.6 in the Supplementary Information).

Our analysis shows that the critical region where bistability occurs involves only a few mRNA molecules and that around eight molecules are already sufficient to fully activate the autoregulatory loop with a high probability and to commit cells to a Krox20-positive fate (Supplementary Figure S7I). The fully activated loop itself sustains about 23 mRNA molecules, showing that element A is strong enough to stably maintain an expression level and cell fate. The distribution of the Krox20-positive cells according to the number of Krox20 molecules depends only on the properties of element A and is independent of the characteristics of the initiation phase. This ensures that these cells form a homogeneous population. Surprisingly, for an autoregulatory system our FGF loss-of-function experiments showed that rhombomere size and, therefore, the number of Krox20-positive cells is gradually modulated by the level of the initiation signal. The model analysis revealed that this feature is entirely due to the stochasticity of the activation process at non-saturating input levels. This is in contrast with conclusions derived from a deterministic analysis of autoregulatory systems, which

© 2013 EMBO and Macmillan Publishers Limited

predicts that all cells acquire either positive or negative fate, when the input level exceeds or falls below a threshold value, respectively. Hence, autoregulation does not only allow the specification of precisely defined developmental stages, but also its control in a graded manner.

Morphogen gradients activate target genes in a concentration-dependent manner, resulting in distinct spatial domains of expression in developing tissues (Wolpert, 1969; Meinhardt, 1978). Recently, Zhang *et al* (2012) presented a model of hindbrain patterning involving Krox20, Hoxb1a, another essential transcription factor in the segmentation process, and retinoic acid (RA), which acts as a morphogen and controls the transcription of the two transcription factors. They use coarse-grained equations to model the dynamics of *krox20* and *hoxb1a* in a single cell with cross inhibition and autoactivation. Their model predicts that the system resolves into a stripped pattern of gene expression. By adding artificial white noise to the system, they investigate the effect of fluctuations on the precision of boundary formation. By varying the noise amplitudes, they show that fluctuations in RA concentration alone induce a rough boundary, whereas additional noise in *hoxb1a/krox20* expression sharpens the boundary (see also Holcman *et al*, 2007). In the present work, we focused on Krox20 and, with our stochastic model, we dissected the molecular dynamics of Krox20 activation and derived the intrinsic fluctuations. We demonstrate that the Krox20 autoregulatory loop establishes a bistable switch and generates Krox20-positive and -negative territories. However, we find that the stochasticity of Krox20 activation precludes the formation of sharp boundaries, and additional mechanisms are therefore required for generating the sharp boundaries observed *in vivo*. The work of Zhang *et al* (2012) indicates that the cross inhibition between *krox20* and *hoxb1a* is likely to constitute a key component in the refinement of the boundaries. In a subsequent step, it would be important to develop and investigate a detailed molecular model that additionally includes *hoxb1a*. A synergetic approach that combines the methods developed in this work with the complementary model of Zhang *et al* (2012) would provide an accurate description of hindbrain patterning fully based on the molecular details.

Robustness of rhombomere size

In the r2–r6 region, the different rhombomeres are approximately of the same AP extent and affecting their relative sizes can have deleterious consequences. Modifications in Fgf signalling have been shown to alter the relative size of r3, r4 and r5 (Marín and Charnay, 2000; Maves *et al*, 2002; Walshe *et al*, 2002; Labalette *et al*, 2011). As indicated above, endogenous *Krox20* expression occurs in saturating conditions, with an initiation level above the range required to activate the loop with high probability. This situation, combined with the graded input–output relation that we have found at lower levels of initiation, is at the origin of robustness of hindbrain patterning, as it prevents a drastic reduction in odd-numbered rhombomere size upon decreasing input signal. Indeed, we have found that rhombomere sizes are buffered from variations in the FGF-tuned initiation signal: e.g., a fourfold reduction in the input signal results in a less than twofold reduction in the number of cells that activate the autoregulatory loop, and therefore in the size of odd-numbered rhombomeres (Figure 4E–J). This buffering effect is likely to provide robustness and have a protective role in hindbrain patterning against other genetic, environmental or stochastic sources of input variation. In this respect, we have recently shown that *Krox20* initiation is subject to a repression mechanisms distributed as a caudo-rostral gradient over the r3–r4 region, which can dramatically affect the relative size of r3 and r4 (Labalette *et al*, in preparation). The graded input–output relationship is also likely to protect the hindbrain against fluctuations of this gradient.

Materials and methods

Mouse and zebrafish lines

All experiments involving animals were performed in accordance with French and European regulations. The mouse *Krox20*^{NA*AK} line was generated at the Institut Clinique de la Souris (Illkirch, France) by homologous recombination in ES cells. The *Krox20*^{NA}, *Krox20*^{A*} and *Krox20*^{AA} alleles were obtained as described in Figure 1A. The *Tg(cA:Krox20*$*_{HA}$) line was generated by transgenesis as described previously (Sham *et al*, 1993). The r2-HPAP transgenic line contains the human placental alkaline phosphatase gene under the control of an r2-specific, *Hoxa2* enhancer element (Studer *et al*, 1996). All primers used for genotyping are presented in Supplementary Table S2A. Zebrafish (*Danio rerio*) were raised and staged as described (Kimmel *et al*, 1995). The WT lines were TL and TU. The *krox20*^{fh227} mutant line was previously described (Monk and Talbot, 2009). The *Tg(cA:h2bm-cherry)* (gift from N Peyrieras, Gif-sur-Yvette, France) and *Tg(hsp:mKrox20*$_{HA}$) lines were obtained by Tol2-mediated transgenesis (Labalette *et al*, 2011).

In-situ hybridization and immunohistochemistry

Mouse and zebrafish *in-situ* hybridizations were performed on whole embryos as described previously (Giudicelli *et al*, 2001) with the following digoxigenin-labelled riboprobes: *mKrox20* (Wilkinson *et al*, 1989), *mEphA4* (Gilardi-Hebenstreit *et al*, 1992) and *zkrox20* (Oxtoby and Jowett, 1993). For immunochemistry in zebrafish and chick embryos, we used rabbit anti-DsRed (1:200, Clontech) and rat anti-GFP (1:500, NacalaiTesque) as primary antibodies and Alexa594 anti-rabbit and Dy488 anti-rat (Jackson ImmunoResearch) as secondary antibodies. Nuclei were counterstained with Hoechst 33342 (Sigma).

Morpholino injection and semi-quantitative *in-situ* hybridization

Nlz knockdown was performed by injecting four morpholinos targeting *nlz1* and *nlz2* at the one-cell stage as previously described (Hoyle *et al*, 2004). Embryos were then subjected to fluorescent *in-situ* hybridization for *krox20* using the Fastred substrate (Roche). Flat-mounted embryos were imaged on a Leica TCS sp5 confocal microscope. The level of fluorescence was then measured within a region of interest (ROI) using ImageJ on stacks of 10 sections, and was normalized by the corresponding area. The fluorescence level corresponds to the mean intensity of all pixels within the ROI (16-bit images). All embryos analysed were processed and imaged in parallel.

Bimodality assay

Five somite *Tg(hsp:mKrox20*$_{HA}$*);Tg(cA:h2b-mcherry)* zebrafish embryos were HS treated for 10 min at 35 ($n = 7$) or 37 °C ($n = 8$), and allowed to develop for 4 h, until ~15s, at 28 °C in embryo medium. Anti-Cherry immunostaining was performed with rabbit anti-DsRed primary

© 2013 EMBO and Macmillan Publishers Limited

antibodies (1:200, Clontech) and Alexa594 anti-rabbit secondary antibodies (Jackson ImmunoResearch). Nuclei were counterstained with Hoechst 33342 (Sigma). Stacks made of eleven 16-bit images of the r2–r6 region were taken with a Leica TCS sp5 confocal microscope. The *z*-step was set 1 μm, such that the total depth corresponded approximately to the depth of one nucleus. Using Fiji software, all images were first smoothened and the projections of the 11 sections were computed by summing the intensity of corresponding pixels. No background substraction, modification of contrast, luminosity or exposure was performed. Fluorescence levels were quantified in single nuclei after manual segmentation: a circular ROI of 12 pixels was placed on each nucleus identified using the Hoechst staining, and fluorescence intensity was measured and averaged over the 12 pixels. Fifty to 80 nuclei were typically quantified in each rhombomere. The plots shown in Figure 6H–L were obtained using the plot density function of the R software with the Sheather and Jones bandwidth selector. The bimodality of these plots was finally assessed statistically using the 'bimodalitytest' package in R (Holzmann and Vollmer, 2008).

Protein extracts and gel retardation assay

Bacterial protein extracts containing Krox20 were prepared as described previously (Nardelli *et al*, 1992). For gel retardation, the DNA probes consisted of a *Hind*III–*Xho*I restriction fragment containing the chick orthologue of element A. The fragment was labelled with [γ-^{32}P]-labelled ATP using polynucleotide kinase (New England Biolabs). The competitors consisted of double-stranded oligonucleotides carrying Krox20-binding sites (Supplementary Table S2B). The EMSA assays were performed as described (Labalette *et al*, 2011). The proportion of DNA fragments retarded owing to complex formation was quantified using a FLA3000 PhosphoImager.

mRNA quantitative analysis

Embryos were collected at the appropriate stage and individually dissected in embryo medium to remove yolk material and to save a piece of embryonic tissue for genotyping. They were then placed in 30 μl of the preservative solution RNA later (Ambion, Life Applied). Total RNA was isolated using the Ambion RNAqueous Micro-kit (Life Applied). Primers for *mKrox20*, *zkrox20* and the housekeeping genes *β-actin* and *eif1a* were designed in single exons to yield amplicons of ~150 bp. Reverse transcription was performed using the Superscript III enzyme (Life Invitrogen) and quantitative PCR analysis was carried out with the SYBR Green Master Mix (Roche Applied Biosystems) using a LightCycler 480 device (Roche Applied Biosystems). In each experiment, a standard curve was established by measuring the threshold-crossing cycle number (Ct) for a series of dilutions of purified genomic DNA. This allowed normalization of the assay- and primer-dependent amplification efficiency. In a single qPCR experiment, the Ct values for all genes were obtained in two duplicate reactions. Means and s.d. were calculated from a minimum of three independent experiments.

Plasmid constructs and *in-ovo* electroporation

Various reporters were cloned into a modified pTol2 plasmid (Stedman *et al*, 2009) under the control of chick element A. The mutant versions were obtained by PCR-mediated directed mutagenesis using the high-fidelity Phusion Taq polymerase (Finnzyme). DNA sequences were verified by sequencing. Electroporation was performed as previously described (Giudicelli *et al*, 2001). Eighteen hours later, the embryos were collected in phosphate-buffered saline, fixed in 4% paraformaldehyde for 3 h and processed for immunostaining.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank François Giudicelli, Patricia Gongal and Denis Thieffry for critical reading of the manuscript, and Nadine Peyriéras for personal communications and for providing us with the *Tg(cA:h2b-mcherry)* zebrafish line. This work was funded by grants from MESR, INSERM, CNRS and ANR to PC and from MESR, INSERM, CNRS and ERC to DH.

Authors Contributions: YXB conceived and performed a large part of the experimental work, carried out simulations and participated in the writing of the manuscript. JR developed and analysed the mathematical model, developed the simulation tools, carried out simulations and participated in the writing of the manuscript. CL participated in the experiments involving the repression of the autoregulatory loop. MAW generated zebrafish lines and initiated the project. ET participated in the analysis of the mouse mutants. CD-TD generated a mouse transgenic line. DH participated in the development of the mathematical model and in the writing of the manuscript. PG-H co-led the experimental part and participated in the writing of the manuscript. PC led the project and participated in the writing of the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Aday AW, Zhu LJ, Lakshmanan A, Wang J, Lawson ND (2011) Identification of cis regulatory features in the embryonic zebrafish genome through large-scale profiling of H3K4me1 and H3K4me3 binding sites. *Dev Biol* **357:** 450–462

Briscoe J (2009) Making a grade: Sonic Hedgehog signalling and the control of neural cell fate. *EMBO J* **28:** 457–465

Burz DS, Rivera-Pomar R, Jäckle H, Hanes SD (1998) Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the *Drosophila* embryo. *EMBO J* **17:** 5998–6009

Chomette D, Frain M, Cereghini S, Charnay P, Ghislain J (2006) Krox20 hindbrain cis-regulatory landscape: interplay between multiple long-range initiation and autoregulatory elements. *Development* **133:** 1253–1262

Cooke JE, Moens CB (2002) Boundary formation in the hindbrain: Eph only it were simple. *Trends Neurosci* **25:** 260–267

Ferrell JE (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* **14:** 140–148

Georges AB, Benayoun BA, Caburet S, Veitia RA (2010) Generic binding sites, generic DNA-binding domains: where does specific promoter recognition come from? *FASEB J* **24:** 346–356

Gilardi-Hebenstreit P, Nieto MA, Frain M, Mattéi MG, Chestier A, Wilkinson DG, Charnay P (1992) An Eph-related receptor protein tyrosine kinase gene segmentally expressed in the developing mouse hindbrain. *Oncogene* **7:** 2499–2506

Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* **22:** 403–434

Giudicelli F, Taillebourg E, Charnay P, Gilardi-Hebenstreit P (2001) Krox-20 patterns the hindbrain through both cell-autonomous and non cell-autonomous mechanisms. *Genes Dev* **15:** 567–580

Graham TGW, Tabei SMA, Dinner AR, Rebay I (2010) Modeling bistable cell-fate choices in the *Drosophila* eye: qualitative and quantitative perspectives. *Development* **137:** 2265–2278

Holcman D, Kasatkin V, Prochiantz A (2007) Modeling homeoprotein intercellular transfer unveils a parsimonious mechanism for gradient and boundary formation in early brain development. *J Theor Biol* **249:** 503–517

Holzmann H, Vollmer S (2008) A likelihood ratio test for bimodality in two-component mixtures with application to regional income distribution in the EU. *Adv Stat Anal* **92:** 57–69

© 2013 EMBO and Macmillan Publishers Limited

Hoyle J, Tang YP, Wiellette EL, Wardle FC, Sive H (2004) nlz gene family is required for hindbrain patterning in the zebrafish. *Dev Dyn* **229:** 835–846

Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF (1995) Stages of embryonic development of the zebrafish. *Dev Dyn* **203:** 253–310

Labalette C, Bouchoucha YX, Wassef MA, Gongal PA, Le Men J, Becker T, Gilardi-Hebenstreit P, Charnay P (2011) Hindbrain patterning requires fine-tuning of early krox20 transcription by Sprouty 4. *Development* **138:** 317–326

Laslo P, Spooner CJ, Warmflash A, Lancki DW, Lee H-J, Sciammas R, Gantner BN, Dinner AR, Singh H (2006) Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell* **126:** 755–766

Lopes FJP, Vieira FMC, Holloway DM, Bisch PM, Spirov AV (2008) Spatial bistability generates hunchback expression sharpness in the *Drosophila* embryo. *PLoS Comput Biol* **4:** e1000184

Lumsden A, Keynes R (1989) Segmental patterns of neuronal development in the chick hindbrain. *Nature* **337:** 424–428

Lumsden A, Krumlauf R (1996) Patterning the vertebrate neuraxis. *Science* **274:** 1109–1115

Ma X, Yuan D, Diepold K, Scarborough T, Ma J (1996) The *Drosophila* morphogenetic protein Bicoid binds DNA cooperatively. *Development* **122:** 1195–1206

Marín F, Charnay P (2000) Hindbrain patterning: FGFs regulate Krox20 and mafB/kr expression in the otic/preotic region. *Development* **127:** 4925–4935

Maves L, Jackman W, Kimmel CB (2002) FGF3 and FGF8 mediate a rhombomere 4 signaling activity in the zebrafish hindbrain. *Development* **129:** 3825–3837

Meinhardt H (1978) Space-dependent cell determination under the control of morphogen gradient. *J Theor Biol* **74:** 307–321

Meinhardt H (1982) Theory of regulatory functions of the genes in the bithorax complex. *Prog Clin Biol Res* **85**(Pt A)**:** 337–348

Monk KR, Naylor SG, Glenn TD, Mercurio S, Perlin JR, Dominguez C, Moens CB, Talbot WS (2009) A G protein-coupled receptor is essential for Schwann cells to initiate myelination. *Science* **325:** 1402–1405

Monk KR, Talbot WS (2009) Genetic dissection of myelinated axons in zebrafish. *Curr Opin Neurobiol* **19:** 486–490

Nardelli J, Gibson T, Charnay P (1992) Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis. *Nucleic Acids Res* **20:** 4137–4144

Nardelli J, Gibson TJ, Vesque C, Charnay P (1991) Base sequence discrimination by zinc-finger DNA-binding domains. *Nature* **349:** 175–178

Oxtoby E, Jowett T (1993) Cloning of the zebrafish krox-20 gene (krx-20) and its expression during hindbrain development. *Nucleic Acids Res* **21:** 1087–1095

Ptashne M (1986) Gene regulation by proteins acting nearby and at a distance. *Nature* **322:** 697–701

Schneider-Maunoury S, Seitanidou T, Charnay P, Lumsden A (1997) Segmental and neuronal architecture of the hindbrain of Krox-20 mouse mutants. *Development* **124:** 1215–1226

Schneider-Maunoury S, Topilko P, Seitandou T, Levi G, Cohen-Tannoudji M, Pournin S, Babinet C, Charnay P (1993) Disruption of Krox-20 results in alteration of rhombomeres 3 and 5 in the developing hindbrain. *Cell* **75:** 1199–1214

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. *Nature* **473:** 337–342

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2013) Corrigendum: Global quantification of mammalian gene expression control. *Nature* **495:** 126–127

Sham MH, Vesque C, Nonchev S, Marshall H, Frain M, Gupta RD, Whiting J, Wilkinson D, Charnay P, Krumlauf R (1993) The zinc finger gene Krox20 regulates HoxB2 (Hox2.8) during hindbrain segmentation. *Cell* **72:** 183–196

Stedman A, Lecaudey V, Havis E, Anselme I, Wassef M, Gilardi-Hebenstreit P, Schneider-Maunoury S (2009) A functional interaction between Irx and Meis patterns the anterior hindbrain and activates krox20 expression in rhombomere 3. *Dev Biol* **327:** 566–577

Studer M, Lumsden A, Ariza-McNaughton L, Bradley A, Krumlauf R (1996) Altered segmental identity and abnormal migration of motor neurons in mice lacking Hoxb-1. *Nature* **384:** 630–634

Swiatek PJ, Gridley T (1993) Perinatal lethality and defects in hindbrain development in mice homozygous for a targeted mutation of the zinc finger gene Krox20. *Genes Dev* **7:** 2071–2084

Tajbakhsh S, Rocheteau P, Le Roux I (2009) Asymmetric cell divisions and asymmetric cell fates. *Annu Rev Cell Dev Biol* **25:** 671–699

Theil T, Frain M, Gilardi-Hebenstreit P, Flenniken A, Charnay P, Wilkinson DG (1998) Segmental expression of the EphA4 (Sek-1) receptor tyrosine kinase in the hindbrain is under direct transcriptional control of Krox-20. *Development* **125:** 443–452

Voiculescu O, Taillebourg E, Pujades C, Kress C, Buart S, Charnay P, Schneider-Maunoury S (2001) Hindbrain patterning: Krox20 couples segmentation and specification of regional identity. *Development* **128:** 4967–4978

Walshe J, Maroon H, McGonnell IM, Dickson C, Mason I (2002) Establishment of hindbrain segmental identity requires signaling by FGF3 and FGF8. *Curr Biol* **12:** 1117–1123

Wassef MA, Chomette D, Pouilhe M, Stedman A, Havis E, Desmarquet-Trin Dinh C, Schneider-Maunoury S, Gilardi-Hebenstreit P, Charnay P, Ghislain J (2008) Rostral hindbrain patterning involves the direct activation of a Krox20 transcriptional enhancer by Hox/Pbx and Meis factors. *Development* **135:** 3369–3378

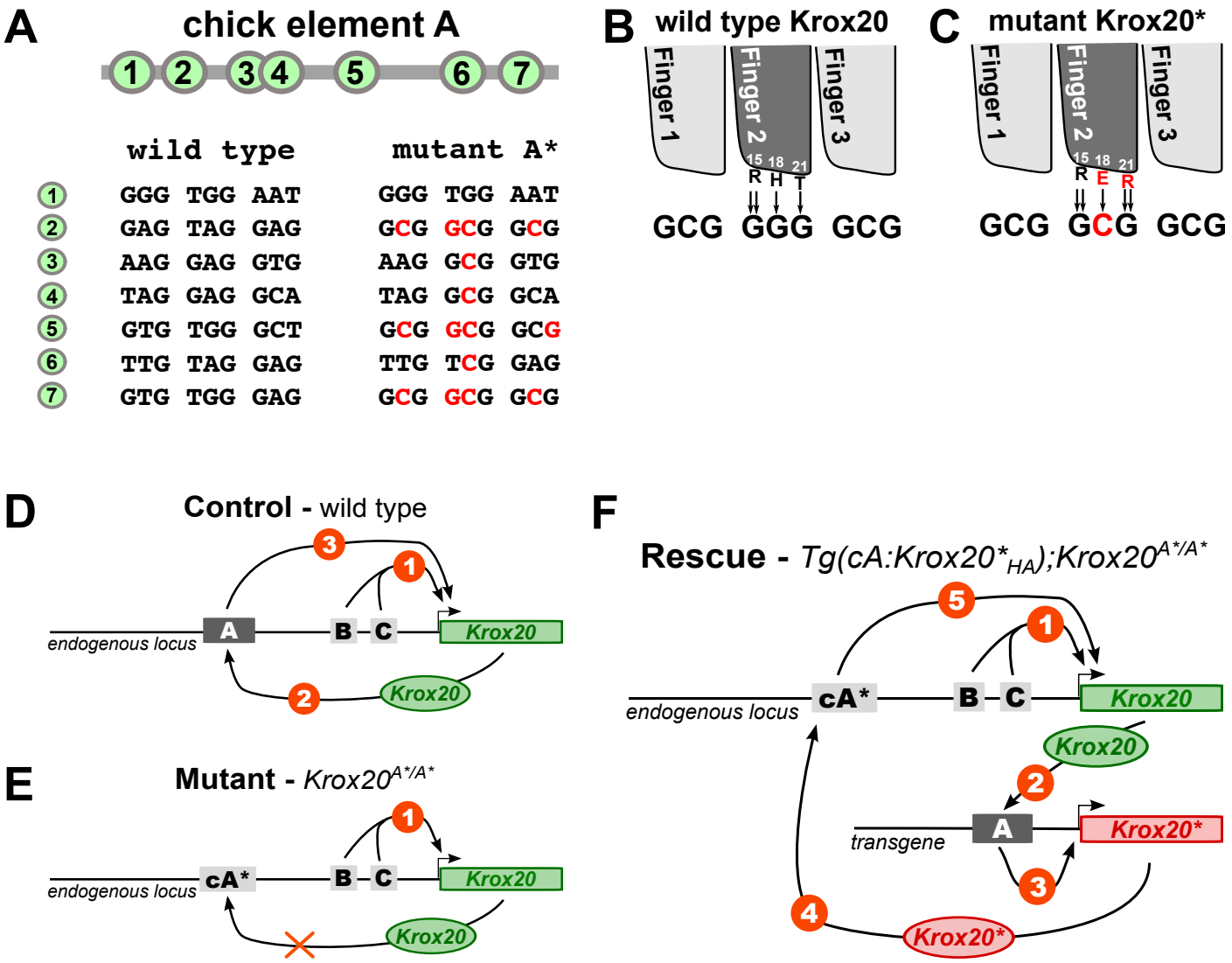
Wilkinson DG, Bhatt S, Chavrier P, Bravo R, Charnay P (1989) Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. *Nature* **337:** 461–464

Wolpert L (1969) Positional information and the spatial pattern of cellular differentiation. *J Theor Biol* **25:** 1–47

Zhang L, Radtke K, Zheng L, Cai AQ, Schilling TF, Nie Q (2012) Noise drives sharpening of gene expression boundaries in the zebrafish hindbrain. *Mol Syst Biol* **8:** 613

***Molecular Systems Biology* is an open-access journal published by the *European Molecular Biology Organization* and *Nature Publishing Group*. This work is licensed under a Creative Commons Attribution 3.0 Unported Licence. To view a copy of this licence visit http://creativecommons.org/licenses/by/3.0/.**

Supplementary Figure S1



Supplementary Figure S1 (related to Figure 1). Krox20 DNA binding specificity and design of the autoregulatory rescue.

(A) Positions of the seven putative Krox20 binding sites along element A and their nucleotide sequences in either their wild type element A or mutant element A*. Modified nucleotides in the mutant version are indicated in red.

(B) Schematic representation of the interactions of wild type Krox20 and mutant Krox20* with their respective consensus binding sites. Modified amino acids and nucleotides in the mutant versions are indicated in red.

(D)Schematic representation of the autoregulatory loop in the wild type situation:*Krox20* expression involves two phases, initiation (1) and autoregulation (2,3).

(E) In the *Krox20*$^{A^*/A^*}$ mutant, the autoregulation phase is abolished since element A is replaced by the mutantelement A*.

(F) The*Krox20*$^{A^*/A^*}$mutant can theoretically be rescued by providing a transgene that expresses*Krox20** in r3 and r5. In this situation, *Krox20* is normally initiated (1) and activates the exogenous element A driving expression of *Krox20**(2,3); in turn Krox20* activates A* on the endogenous locus (4), leading to further production of Krox20 (5). The rescue is based on a novel, indirect autoregulatory loop.

- Complément -

L'autorégulation du gène *Krox20* est un processus décisif pour la spécification cellulaire du rhombencéphale

1. Introduction

L'autorégulation du gène *Krox20* permet le maintien et l'amplification du gène par la fixation directe de la protéine KROX20 sur l'élément A. Nous avons montré que ce processus permet de contrôler la taille des rhombomères r3 et r5 et d'activer correctement une des cibles transcriptionnelles de *Krox20* : le gène *EphA4*. L'expression d'*EphA4* chez le mutant de délétion de l'élément A est initiée correctement mais la taille du domaine d'expression à 8,5 jpc est très diminuée et une partie des cellules ne maintient pas son expression suggérant que ces cellules n'acquièrent pas une identité de rhombomères impairs. Il semblerait donc que l'autorégulation du gène *Krox20* soit déterminante pour la spécification des rhombomères impairs et l'organisation neuronale qui en découle. Ce complément d'article vise à élucider à la fois le devenir des cellules n'ayant pas maintenu *Krox20* et les conséquences sur le patterning du rhombencéphale. En utilisant différentes approches génétiques, nous montrons que certaines cellules acquièrent une identité de rhombomères impairs et sont en partie re-spécifiées en identité de rhombomère pair (sans toutefois exclure une apoptose précoce notamment pour la région correspondant à r3), ce qui a pour conséquence d'affecter la différenciation neuronale et l'organisation des nerfs crâniens plus tardivement au niveau de ce même domaine.

2. Matériels et méthodes

Lignées de souris

En plus de l'allèle $Krox20^{\Delta A}$, plusieurs allèles ou transgènes ont été utilisés dans ce complément d'étude : $Krox20^{LacZ}$ (Schneider-Maunoury et al. 1993), $Krox20^{Cre}$ (Voiculescu et al. 2000), $Krox20^{GFP(DT)}$ (Vermeren et al. 2003) et $Rosa26^{tdTom}$ (Madisen et al. 2010).

Immunofluorescence sur embryons de souris à 9,5 jpc et 10,5 jpc

Les embryons sont fixés 2h en 4 % PFA dans la glace puis déshydratés/réhydratés très progressivement en méthanol. Ils sont bloqués en PBS contenant 0,5 % de blocking reagent

(Roche), 0,3 % Triton et 0,2 % BSA toute la nuit avant d'être incubés avec les anticorps primaires anti-GFPde lapin (Invitrogen, 1/500), anti-2H3 de souris (DSHBD, 1/1000), anti-EphA4 de lapin (anticorps maison, 1/1000) et anti-HoxB1 de lapin (Covance, 1/200) toute la nuit à 4°C. Après lavage en PBS, les anticorps secondaires sont ajoutés : anti-lapin Dy488 de chèvre (Molecular Probes, 1/500) et anti-souris Cy3 d'âne (Jackson ImmunoResearch, 1/500) pendant 2h à l'abri de la lumière. Plusieurs lavages sont nécessaires pour limiter le bruit de fond. Les tubes neuraux sont ensuite montés à plat rapidement en fluoromount.

3. Résultats

L'autorégulation de *Krox20* et le destin cellulaire des rhombomères

A 9 jpc, le domaine r3 est réduit dans le mutant *Krox20*$^{\Delta A/\Delta A}$ par rapport au contrôle observé avec le double marquage *r2-HPAP* pour r2 et *Hoxb1* pour r4 alors que le domaine r4 est quant à lui légèrement plus large (Bouchoucha et al. 2013). A 9,5 jpc, une hybridation *in situ* contre *EphA4* révèle de façon normale un faible marquage dans r2, mais un domaine très réduit correspondant à r3 et un faible marquage dans celui de r5 dans des embryons mutants *Krox20*$^{\Delta A/LacZ}$ par rapport aux contrôles *Krox20*$^{+/LacZ}$ (fig. 13A). Ceci suggère que l'absence de l'autorégulation de *Krox20* peut entraîner une re-spécification des cellules en domaines pairs, une mort cellulaire ou un problème de prolifération. Un traçage cellulaire à 9,5 jpc a été effectué pour tester le devenir des cellules ayant initié l'expression de *Krox20*. Pour cela nous avons combiné l'allèle *Krox20*^{Cre} avec le transgène *Rosa26*^{tdTom} permettant, lorsque le locus *Krox20* est activé, l'excision par la Cre recombinase du transgène et l'activation permanente du rapporteur exprimant la dtTomato en présence de l'allèle sauvage, *Krox20*$^{\Delta A}$ ou *Krox20*^{LacZ}. Par immunofluorescence on observe chez le contrôle *Krox20*$^{+/Cre}$ *; R26*^{tdTom} que les cellules tracées coïncident parfaitement avec les cellules *EphA4* positives dans les rhombomères 3 et 5 en accord avec le fait que toutes les cellules exprimant *Krox20* dans r3 et r5 activent l'expression de *EphA4*. Dans le mutant *Krox20*$^{\Delta A/Cre}$ *; R26*^{tdTom}, les cellules *EphA4* positives représentent une fine bande dans r3 et sont difficiles à distinguer dans r5 comme observé en hybridation *in situ* (fig. 13A, flèches). Les cellules tracées quant à elles forment un petit domaine aux frontières floues au niveau de r3 dont une partie est négative pour *EphA4* au niveau de r2 et r4 (fig. 13B, flèches) et un grand domaine aux frontières floues au niveau de r5 qui est plus large que chez le contrôle. Chez le mutant nul *Krox20*$^{LacZ/Cre}$ *; R26*^{tdTom}, le nombre de cellules tracées correspondant à r3 et r5 est plus faible encore que pour le mutant d'autorégulation, mais la forme des domaines est similaire (fig. 13B). Ceci suggère que chez

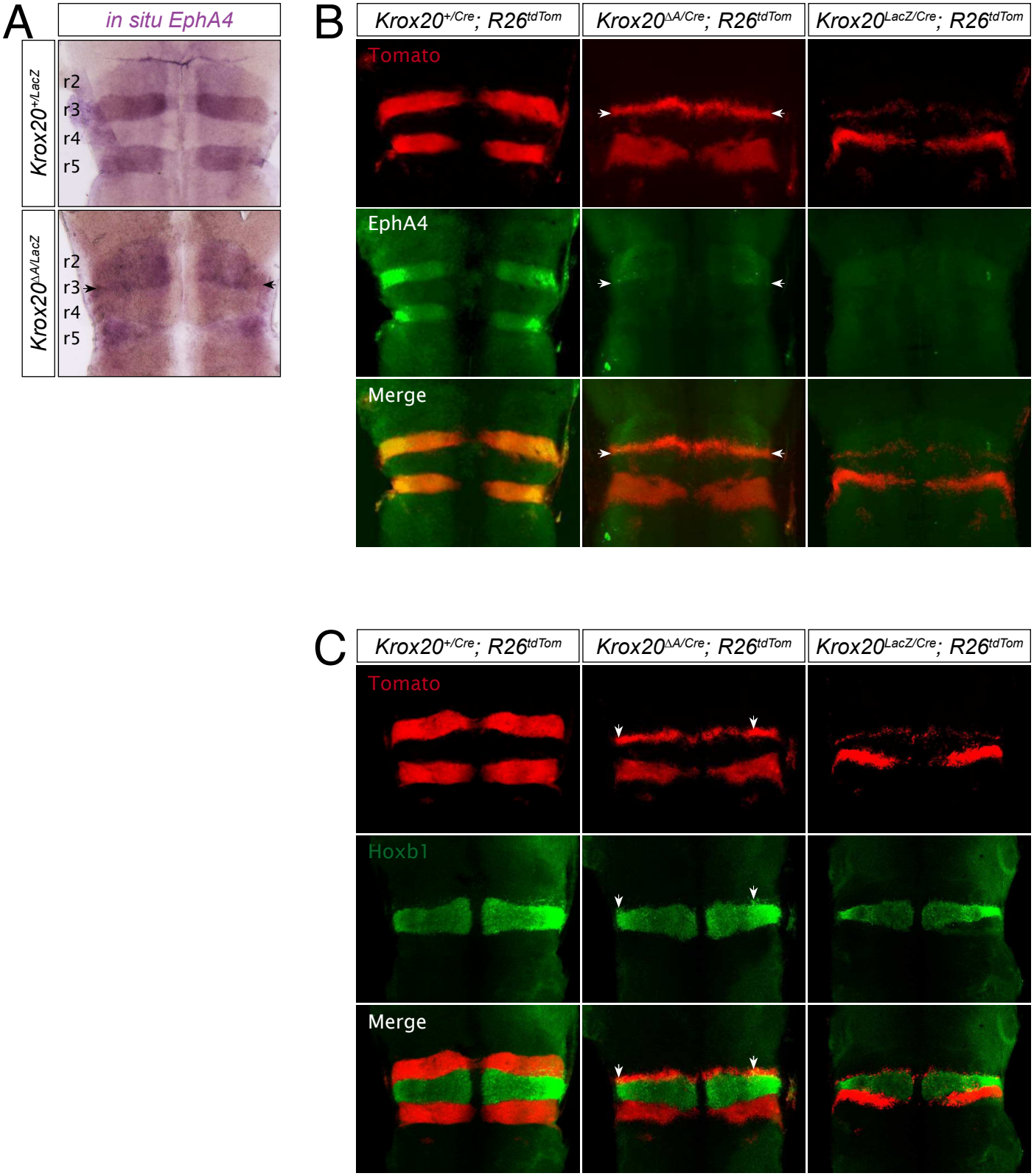


Figure 13 : L'autorégulation de *Krox20* dans la spécification des rhombomères

A. Hybridation *in situ EphA4* sur des embryons à 9,5 jpc contrôles *Krox20*$^{+/LacZ}$ et mutants *Krox20*$^{\Delta A/LacZ}$ montés à plat entre lames et lamelles. Les flèches noires indiquent le marquage *EphA4* dans r3.

B. Immunofluorescence sur embryons entiers à 9,5 jpc contrôles *Krox20*$^{+/Cre}$*; R26*^{tdTom} et mutants *Krox20*$^{\Delta A/Cre}$*; R26*^{tdTom} et *Krox20*$^{LacZ/Cre}$*; R26*^{tdTom} montés à plat entre lames et lamelles marqués par l'anticorps anti-EphA4 et visualisation directe de la tdTomato (tdTom). Les flèches horizontales indiquent un chevauchement des marquages EphA4 et tdTomato.

C. Immunofluorescence sur embryons entiers à 9,5 jpc contrôles *Krox20*$^{+/Cre}$*; R26*^{tdTom} et mutants *Krox20*$^{\Delta A/Cre}$*; R26*^{tdTom} et *Krox20*$^{LacZ/Cre}$*; R26*^{tdTom} montés à plat entre lames et lamelles marqués par l'anticorps anti-HoxB1 et visualisation directe de la tdTomato (tdTom). Les flèches verticales indiquent un chevauchement des marquages Hoxb1 et tdTomato.

le mutant *Krox20*$^{\Delta A/Cre}$; *R26*^{tdTom} une partie des cellules tracées est re-spécifiée en rhombomères r2 et r4. Ceci est appuyé par la même expérience avec le marquage Hoxb1 dans r4 où on peut observer un chevauchement des cellules tracées avec les cellules Hoxb1 positives (fig. 13C, flèches). En revanche, il n'y a pas de mélange entre les cellules positives Hoxb1 et le domaine r5 des cellules tracées, ce qui laisse penser que le grand nombre de cellules tracées dans le domaine r5 sont re-spécifiées en r6. Comme pour le mutant nul, une partie des cellules ayant activé le gène *Krox20* mais dont l'autorégulation est abolie avec la délétion de l'élément A semble re-spécifiée en rhombomères pairs. Cependant, contrairement au mutant nul où un vague d'apoptose réduit fortement le nombre de cellule tracées à la fois dans les domaines correspondant à r3 et r5 à 9,5 jpc après la re-spécification, il semblerait que ce phénomène soit plus limité chez le mutant de l'autorégulation dans r3 et faible dans r5 ainsi que plus précoce (Voiculescu et al. 2001, fig. 13 et données non présentées).

L'autorégulation de *Krox20* et l'organisation neuronale du rhombencéphale

La segmentation du rhombencéphale préfigure l'organisation neuronale du cerveau postérieur, notamment grâce au rôle décisif du gène *Krox20* dans cette organisation. Cette dernière a donc été observée par le marquage des neurofilaments (2H3) sur des embryons *Krox20*$^{\Delta A/GFP}$ à 9,5 jpc. Chez l'embryon contrôle *Krox20*$^{+/GFP}$ (fig. 14A) les zones correspondant à r3 et r5-r6 sont de façon attendue moins denses en neurofilaments ce qui est dû à un décalage temporel de différenciation neuronale. On ne retrouve pas au niveau de r3 cette zone déplétée en neurones chez le mutant (fig. 14A), mais une région continue, dense en neurofilaments, correspondant à r2 et r4. En revanche, la zone r5-r6, quant à elle ,est peu dense en neurofilaments et ne semble pas avoir subi de réduction de taille. La perte d'autorégulation chez le mutant perturbe la segmentation du rhombencéphale et l'organisation neuronale qui en découle au niveau de la région r2 à r4 principalement. Ce phénotype est proche de celui observé pour le mutant *Krox20*$^{LacZ/LacZ}$ (Schneider-Maunoury et al. 1997).

Afin de mieux caractériser l'effet de l'absence d'autorégulation du gène *Krox20* grâce à la délétion de l'élément A sur l'organisation neuronale, un marquage 2H3 a aussi été réalisé sur des embryons entiers à 10,5 jpc. À ce stade, les nerfs crâniens prennent racine dans le rhombencéphale et projettent dans les arcs branchiaux sous-jacents. Chez le contrôle *Krox20*$^{+/\Delta A}$, les racines du nerf V (trijumeau) et des nerfs VII (facial) et VIII (auditif) sont bien séparées, ce qui n'est pas le cas chez le mutant *Krox20*$^{\Delta A/\Delta A}$ (fig. 14B, flèches). Les autres racines des nerfs ne semblent pas affectées mais la taille globale du rhombencéphale est réduite chez le mutant comparée au contrôle à ces stades tardifs. Ce rapprochement est en

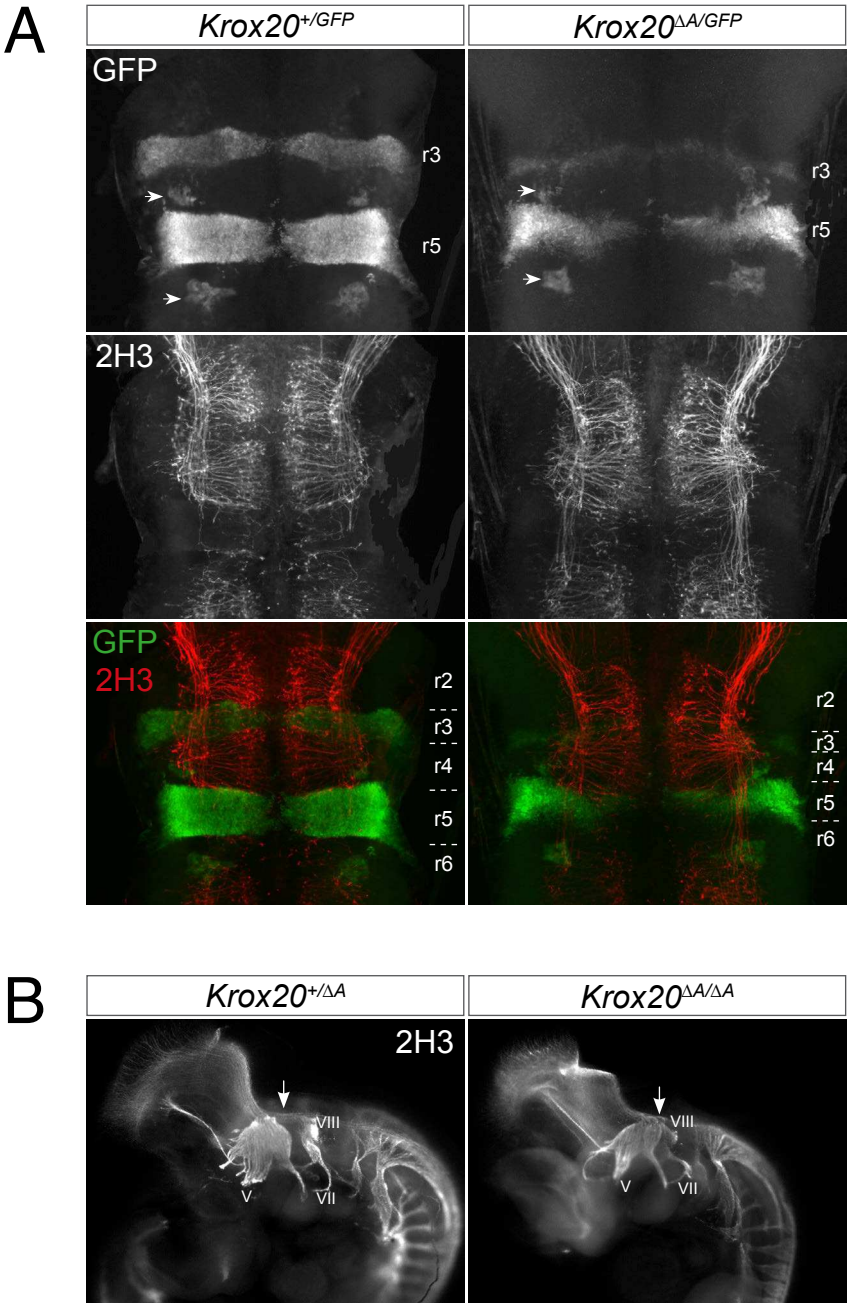


Figure 14 : L'autorégulation de *Krox20* et l'organisation neuronale du rhombencéphale

A. Immunofluorescence sur embryons entiers à 9,5 jpc contrôles *Krox20*$^{+/GFP}$ et mutants *Krox20*$^{\Delta A/GFP}$ montés à plat entre lames et lamelles marqués par les anticorps anti-GFP et anti-neurofilaments 2H3. Les flèches indiquent les cellules des capsules frontières.

B. Immunofluorescence sur embryons entiers à 10,5 jpc contrôles *Krox20*$^{+/\Delta A}$ et mutants *Krox20*$^{\Delta A/\Delta A}$ marqués par l'anticorps anti-neurofilaments 2H3. Les flèches indiquent la zone de rapprochement du nerf V (trijumeau) et des nerfs VII (facial) et VIII (auditif).

accord avec l'organisation neuronale continue de la zone r2-r4 observée précédemment. Ce phénotype a été observé chez le mutant *Krox20*$^{LacZ/LacZ}$ qui présentait aussi une forte proximité des nerfs IX et X (Schneider-Maunoury et al. 1993). Le maintien du gène *Krox20* est donc indispensable lors de la segmentation du rhombencéphale puisqu'il permet la mise en place des neurones à l'origine des nerfs crâniens et par la suite à leur différenciation.

4. Discussion

En plus de définir la taille des rhombomères impairs, l'autorégulation du gène *Krox20* permet la mise en place de frontières nettes en contrôlant l'expression de différentes molécules d'adhésion dont EphA4, participant ainsi à la ségrégation cellulaire grâce au système de récepteur Eph et ligand Ephrins exprimés dans les rhombomères complémentaires (Xu et al. 1999). La perte d'expression d'*EphA4* dans une partie des cellules du mutant de l'autorégulation conduit à une re-spécification de ces cellules en rhombomères pairs et probablement une vague d'apoptose précoce plus importante dans r3 que r5. Ceci est en accord avec la perturbation de la différenciation neuronale et le rapprochement des nerfs crâniens chez le mutant dans r3 alors que la région r5 est peu affectée. Ce phénotype pour la région r3 est similaire à ce qui est observé chez le mutant nul mais il semblerait que les processus mis en jeu soient temporellement différents. En effet, chez le mutant nul, la re-spécification précède la vague d'apoptose alors que ce serait l'inverse pour le mutant de l'autorégulation. On peut envisager que lorsque les cellules reçoivent une petite dose de Krox20, la re-spécification précoce ne soit plus possible. Il n'est pas exclu que de problèmes de prolifération soit aussi à l'origine de la réduction des domaines r3 et r5, mais cette hypothèse est difficile à investiguer au vu du faible nombre de cellules dans le rhombencéphale. Enfin, en contrôlant la taille et les frontières des rhombomères, l'autorégulation de *Krox20* est un processus qui assure la robustesse de l'expression du gène, en tamponnant les modifications génétiques ou environnementales et qui permet d'assurer l'intégrité de cette étape clef du développement.

- Chapitre 2 -

Manuscript in preparation

Different modes of cooperation between *cis*-acting elements govern the expression of the *Krox20/Egr2* hindbrain developmental gene

Elodie Thierion[1,2], Johan Le Men[1], Fanny Coulpier[1], Céline Hernandez[1], Morgane Thomas-Chollier[1], Samuel Collombet[1], Daan Noordermeer[3], Pascale Gilardi-Hebenstreit[1], Patrick Charnay[1]

[1] Ecole Normale Supérieure, Institut de Biologie de l'ENS (IBENS), and Inserm U1024, and CNRS UMR 8197, 46 rue d'Ulm, 75005 Paris, France

[2] Sorbonne Universités, UPMC Univ Paris 06, IFD, 4 Place Jussieu, 75252 Paris cedex 05

[3] Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris Sud, 91198 Gif-sur-Yvette, France

Author contributions:

ET performed most of the experiments.

JLM participated in the analysis of the mouse knockout of element C.

FC generated the libraries and performed the sequencing.

CH, MTC and SC performed the computational analysis.

DN participated in the 4C-seq experiments and interpreted the results.

ET, PGH and PC conceived the experiments, interpreted the results and wrote the manuscript.

PRÉAMBULE

Ce chapitre, rédigé en anglais, regroupe les résultats majeurs obtenus au cours de ma thèse, sous la forme d'un manuscrit préliminaire qui sera envoyé pour publication prochainement car certaines expériences doivent être finalisées. Nous avons examiné la contribution de l'élément *cis*-régulateur C à l'expression du gène *Krox20*, qui est le seul élément identifié ayant une activité précoce dans r3 indépendante *de Krox20*, grâce à la délétion de cet élément dans le génome murin. De façon surprenante, l'élément C n'est pas indispensable à l'initiation de l'expression dans r3 mais il est requis pour l'autorégulation du gène puisqu'en son absence l'expression de *Krox20* n'est ni amplifiée ni maintenue dans r3. Grâce à des approches génétiques, nous avons montré que l'élément C coopère en *cis* avec l'élément A au cours de la phase d'autorégulation. Par des approches de conformation et d'accessibilité de la chromatine, nous avons montré que l'élément A interagit de façon constitutive avec le promoteur et que l'élément C affecte à distance l'accessibilité de l'élément A pendant l'autorégulation. Cette étude met en évidence plusieurs modes de coopération entre éléments *cis*-régulateurs pour assurer un patron d'expression robuste du gène *Krox20* au cours du développement du rhombencéphale chez la souris.

Different modes of cooperation between *cis*-acting elements govern the expression of the *Krox20/Egr2* hindbrain developmental gene

INTRODUCTION

Transcriptional *cis*-regulatory elements play key roles in establishing spatial-temporal gene expression patterns throughout the development of an organism. A major class of *cis*-regulatory elements are the transcriptional enhancers that are able to recruit combinations of transcription factors (TFs) and to activate transcription from (a) gene promoter(s) relatively independently of their location, distance or orientation in respect to this promoter (Banerji et al. 1981; Maniatis et al. 1987). Multiple enhancers can act sequentially or concomitantly for the control of the expression of a gene. They can be located very far from their target gene, up to 1 Mb (Lettice et al. 2003), raising serous issues about the mechanisms that allow them to communicate with the promoters. So far, most enhancers have been functionally characterized by assay of their transcriptional activity either in cell culture or in the animal after transgenesis with a construct in which the enhancer is linked to a minimal promoter driving the reporter gene. Such approaches provide useful data on the spatial and temporal activity of the enhancer, but do not establish whether the suspected target gene is under the control of the enhancer and in which conditions, in other words whether the enhancer is really an element controlling the expression of the genome. The only way to definitively address this issue is to delete or mutate the putative enhancer at its endogenous location. Unfortunately, this essential step has not been carried out for most suspected enhancers in vertebrates and, in particular, in the mouse. An additional difficulty arises from the existence of redundancy between enhancers controlling the expression of the same gene. Such redundancy is thought to ensure transcriptional robustness in conditions of genetic or environmental variations (Lagha et al. 2012; Barolo 2012). But it can result in difficulties in demonstrating the existence of a phenotype after knock-out of one of the redundant enhancers, as observed in the cases of the *Otx2* and *Pomc* genes (Kurokawa et al. 2004; Lam et al. 2015).

Transcriptional activation by enhancers is intimately linked to the organization of the chromatin. Hence, pioneer TFs can decompact chromatin at enhancers and facilitate the binding of other TFs and the recruitment of various epigenetic machineries (Zaret & Carroll 2011). Reversely, spatial chromatin conformation can impact gene expression, for instance when enhancers are engaged in long-range interactions with their target promoters (Gibcus & Dekker 2013). In this respect, architectural proteins CTCF and cohesin are involved in

enhancer-promoter looping (Hadjur et al. 2009; Dowen et al. 2014). More generally, the genome is partitioned into domains called topologically associated domains (TADs), 1 Mb-long on average, which are demarcated by boundaries enriched in CTCF binding sites. TADs are thought to constitute the chromatin backbone of the genome, because they display coherent chromatin signatures and gene expression patterns (Dixon et al. 2012; Nora et al. 2012; Sexton et al. 2012). Furthermore, TADs are providing specificity to enhancer-promoter interactions by preventing aberrant gene activation by *cis*-regulatory elements located in adjacent TADs (Nora et al. 2012; Symmons et al. 2014). Among the epigenetic modifications, H3K4me3 and H3K27ac at promoters and H3K4me1 and H3K27ac at enhancers mark active chromatin state (Heintzman et al. 2007; Heintzman et al. 2009).

To provide insights in the mechanisms involved in the regulation of a gene by multiple enhancers, we investigated the case of the *Krox20/Egr2* gene for which the *cis*-regulatory sequences involved in its control during hindbrain development have been studied in detail. The hindbrain is an attractive model to investigate the genetic control of morphogenesis in vertebrates. It is subject to a transient segmentation process leading to the formation of 7 to 8 segments called rhombomeres (r) (Lumsden & Krumlauf 1996). Rhombomeres constitute cell compartments and developmental units for neuronal differentiation and branchiomotor nerve organization (Lumsden & Keynes 1989). The specification and formation of segments r3 and r5 are governed by the expression of the transcription factor KROX20/EGR2 (Schneider-Maunoury et al. 1993; Schneider-Maunoury et al. 1997; Voiculescu et al. 2001). Three evolutionarily conserved sequences exhibiting enhancer activity in the hindbrain have been so far identified and are termed A, B and C (Chomette et al. 2006). Element B, located 164 kb upstream of the promoter, drives the expression of a reporter construct specifically in r5. Element C, located 144 kb upstream of the promoter, is active in the r3-r5 region (Chomette et al. 2006). Finally, element A, located 217 kb upstream of the promoter, is active in r3 and r5. This element requires direct binding of the KROX20 protein for its activity, suggesting that it is an autoregulatory element (Chomette et al. 2006). Indeed, deletion of element A in the mouse led to a phenotype in which early *Krox20* expression occurs normally in r3 and r5, but is not amplified nor maintained at later stages (Bouchoucha et al. 2013). Additional studies indicated that element A is actually required for a positive feedback loop that acts as a binary switch for specification of odd- versus even–numbered rhombomere identity (Bouchoucha et al. 2013). In contrast, several observations suggested that elements B and C are involved in the initial steps of *Krox20* expression: i) they are transcriptionally active at early stages of *Krox20* transcription (Chomette et al. 2006); ii) they are activated by transcription factors known to act upstream of *Krox20* activation. Element B is under the

direct transcriptional control of the Vhnf11 and MafB TFs (Chomette et al. 2006; Labalette et al. 2011) that are known genetically to be upstream of *Krox20* activation in r5 (Wiellette & Sive 2003; Sabine P. Cordes & Barsh 1994). Element C is under the direct transcriptional control of Meis, Hox/Pbx and Sp5 factors (Wassef et al. 2008; Labalette, M. A. Wassef, et al. 2015). Pbx and Meis have been shown to be required for the establishment of the anterior domain of *Krox20* expression (Waskiewicz et al. 2002; Choe et al. 2002). Evidence for the implication of Hox factors in *Krox20* expression has been obtained in Xenopus, where combined HoxA1, B1, D1 loss of function leads to a total absence of *Krox20* expression (McNulty et al. 2005). iii) We have also shown that elements B and C are responsive to FGF signalling (Labalette et al. 2011; Labalette, M. A. Wassef, et al. 2015), which has been shown to be required for the initiation of *Krox20* and to determine the size of r3 and r5 (Walshe et al. 2002; Aragon & Pujades 2009). iv) Finally, elements B and C do not require the presence of the KROX20 protein for their activity (Chomette et al. 2006).

In the present study, we have investigated the contribution of element C to *Krox20* expression, as it is the only characterized enhancer element with an activity in r3 independent of Krox20 itself. Using a conditional knock-out mutation of element C generated in the mouse, we show that, unexpectedly, this element is dispensable for the *Krox20* initial expression in r3. At this stage, element C is likely to act redundantly with another *cis*-regulatory element. In contrast, it is absolutely required for the maintenance of *Krox20* expression in this rhombomere. This activity involves a cooperation in *cis* between elements A and C for establishing the autoregulatory loop. Investigation of chromatin conformation and accessibility in the *Krox20* locus indicated that element C is able to affect the accessibility of element A, distant from 73 kb. These data reveal that a single *cis*-acting element can cooperate with other enhancers in the same locus according to different modalities (redundant or synergistic) to establish the spatio-temporal pattern of gene expression and provide important clues for understanding the molecular mechanisms governing such interactions.

MATERIALS & METHODS

Mouse lines

All mouse lines were maintained in a mixed C57BL6/DBA2 background. We used the following alleles: *Krox20*^{Cre} (Voiculescu et al. 2000) and *Krox20*$^{\Delta A}$ (Bouchoucha et al. 2013). The mouse *Krox20*^{CFlox} line was generated at the Institut Clinique de la Souris (Illkirch, France) by homologous recombination in ES cells. The *Krox20*$^{\Delta C}$ allele was obtained as described in Figure 1. All animal manipulations were approved by a French Ethical Committee (Project Ce5/ 2012/120) and were performed according to French and European Union regulations.

In situ hybridization

Mouse *in-situ* hybridizations were performed on whole embryos as previously described (Giudicelli et al. 2001) with the following digoxigenin-labelled riboprobe: *mKrox20* (Wilkinson et al. 1989).

ChIP-seq experiments

ChIP experiments were performed as previously described (Vitobello et al. 2011). Briefly, 10 embryos at E9.5 and 20 embryos at E8.5 were dissected in cold PBS. Cell suspensions were obtained by passing them through a 21G needle fitted onto a 5mL syringe. The cells were crosslinked at 1% formaldehyde for 10 min and washed twice in PBS, 1mM PMSF, 1X PIC. The sonication was performed on a Covaris S220 using the following programme at E9.5/E8.5: duty factor = 10/5, peak incident power = 140, cycles per burst = 200 during 600/480 seconds. 5-10μg of chromatin was used for each IP using 3μg of the following antibodies: anti H3K4me1 (C15410037, Diagenode) and anti H3K27ac (ab4729, Abcam) in RIPA buffer. The libraries were prepared with the NEXTflex ChIP-Seq Kit (Bioo Scientific). Sequencing was done on multiplexed samples using 50 bp Single-end reads on the Illumina HiSeq system according to the manufacturer's specifications at the genomics platform of IBENS. Chip-seq data were analysed using Eoulsan2 in its version alpha 5 with extended support for ChIP-seq workflows. Firstly, reads were filtered using their Illumina quality score (reads with a quality inferior to were discarded). If needed, files corresponding to technical replicates were merged. The mapping was then performed using Bowtie version 2.2.4 (using the --sensitive option). Further filters were applied to the mapped reads using Picard's MarkDuplicates version 1.118 (removed.marked=true) and the filtersam module of Eoulsan (removedunmapped=true). SAM files corresponding to input data were merged using the

mergeinput module, and all SAM files were consecutively converted into BAM (module sam2bam). Peak calling was performed with MACS version 2.1.0.20140616 (p.value=0.001, genome.size=mm, make.bedgraph=true, is.broadpeak=false and additional parameters: --SPMR --call-summits).

4C-seq experiments

4C-seq libraries were constructed as previously described (Noordermeer et al. 2014) with small modifications. 35 embryos at E9.5 and 250 embryos at E8.5 were dissected in cold PBS, followed by dissociation in collagenase type I (Gibco). DpnII (New England Biolabs, Ipswich, MA) was used as the primary restriction enzyme and NlaIII (New England Biolabs) was used as the secondary restriction enzyme. For each viewpoint, up to 600ng of each 4C-seq library was amplified using 16 individual PCR reactions with inverse primers including Illumina adapter sequences (Table S1). Illumina sequencing was done on multiplexed samples, containing PCR amplified material of up to 7 viewpoints, using 100 bp Single-end reads on the Illumina HiSeq system according to the manufacturer's specifications at the iGE3 Genomics Platform of the University of Geneva (Switzerland). 4C-seq reads were sorted, aligned, and translated to restriction fragments using the 4C-seq pipeline of the BBCF HTSstation (David et al. 2014) according to ENSEMBL Mouse assembly NCBIM37 (mm9).
Position of mouse ES cells TADs and associated 40 kb normalized Hi-C matrices (Dixon et al. 2012) were obtained from http://promoter.bx.psu.edu/hi-c/download.html. Interaction matrices are visualized using standard cut-offs.

ATAC-seq experiments

ATAC experiments were performed according to Buenrostros et al., (2013)'s protocol and using a homemade transposome from Picelli et al., (2014). 7-8 embryos at 10-12s were dissected in cold PBS for each genotype and cells were mechanically dissociated. Cells were lysed before transposition using 1μL of transposome and purified using a Qiagen MinElute Kit with 10μL of Elution Buffer. Transposed DNA was amplified by PCR as previously described and quantified by qPCR using 5μL of PCR products. The number of additionnal cycles is determined by plotting linear Rn versus cycle and correspond to the one-third of the maximum fluorescence intensity. The remaining 45μL of PCR products are runned with the additional number of cycles. The final product is purified with Qiagen PCR Cleanup Kit and eluted in 20μL Elution Buffer. Sequencing was done on multiplexed samples using 42 bp Paired-end reads on the Illumina NextSeq system according to the manufacturer's specifications at the genomics platform of IBENS. For computational analysis, paired-end

reads were mapped onto the mouse genome assembly mm9 using STAR (outFilterMultimapNmax 1; outFilterMismatchNmax 999; outFilterMismatchNoverLmax 0.06 ; alignIntronMax 1; alignEndsType EndToEnd; alignMatesGapMax 2000). Duplicates reads were removed using Picard (http://picard.sourceforge.net) (MarkDuplicates, REMOVE_DUPLICATES = true). Bigwig tracks were made using DeepTools BamCoverage (1.5.9.1).

RESULTS

In r3, element C is required for late, but not initial *Krox20* expression

To assess the contribution of element C to the regulation of *Krox20* expression, we generated a mouse line carrying a deletion of this element. The details of the knock-out strategy are presented in Figure 1A. Two alleles were generated: $Krox20^{Cflox}$, in which element C is present but flanked with two loxP sites, and $Krox20^{\Delta C}$, in which element C is deleted. The impact of element C deletion on *Krox20* transcription was analysed by mRNA *in situ* hybridization on 4 to 14 somite stage (s) embryos, comparing homozygous deleted mutants ($Krox20^{\Delta C/\Delta C}$) with heterozygous ($Krox20^{+/\Delta C}$) and wild-type embryos (Fig. 1B and data not shown). Unexpectedly, up until 6s the level of *Krox20* mRNA in r3 and r5 is similar in $Krox20^{\Delta C/\Delta C}$ embryos and control littermates (Fig. 1B and data not shown). However, at 8s, *Krox20* r3 expression is severely reduced in $Krox20^{\Delta C/\Delta C}$ embryos as compared to controls and, at later stages, it is completely lost (Fig. 1B). During the considered period, although the level of *Krox20* mRNA does not appear to be grossly affected in r5, the width of the corresponding domain of expression appears to be slightly reduced after 10s (Fig. 1B). In conclusion, these data indicate that i) element C is not necessary for the initiation of *Krox20* expression in r3 nor in r5; ii) in contrast, it appears to be absolutely required for r3 expression beyond 6s, the period corresponding to the maintenance of *Krox20* expression by the autoregulatory loop (Chomette et al. 2006; Bouchoucha et al. 2013); iii) in r5, the contribution of element C to *Krox20* expression appears to be rather limited; iv) another element drives the expression of *Krox20* during the initiation phase in r3.

Elements A and C cooperate in *cis* for establishing the autoregulatory loop

In $Krox20^{\Delta C/\Delta C}$ mutants, the dynamics of the *Krox20* expression pattern in r3 is very similar to the one observed in $Krox20^{\Delta A/\Delta A}$ embryos, which carry a homozygous deletion of the autoregulatory element A (Bouchoucha et al. 2013). This similarity raises the possibility of an involvement of both elements in a common regulatory process. To investigate this issue, we analyzed the expression of *Krox20* in composite heterozygous embryos, $Krox20^{\Delta A/\Delta C}$, carrying deletions of element A on one chromosome and of element C on the other (Fig. 1C). Although this combination does not affect *Krox20* expression in r3 and r5 at early stages (data not shown), at 12s, it leads to a complete absence of *Krox20* mRNA in r3, as observed in $Krox20^{\Delta C/\Delta C}$ embryos (Fig. 1B,C). In r5, *Krox20* expression is only slightly affected, similarly

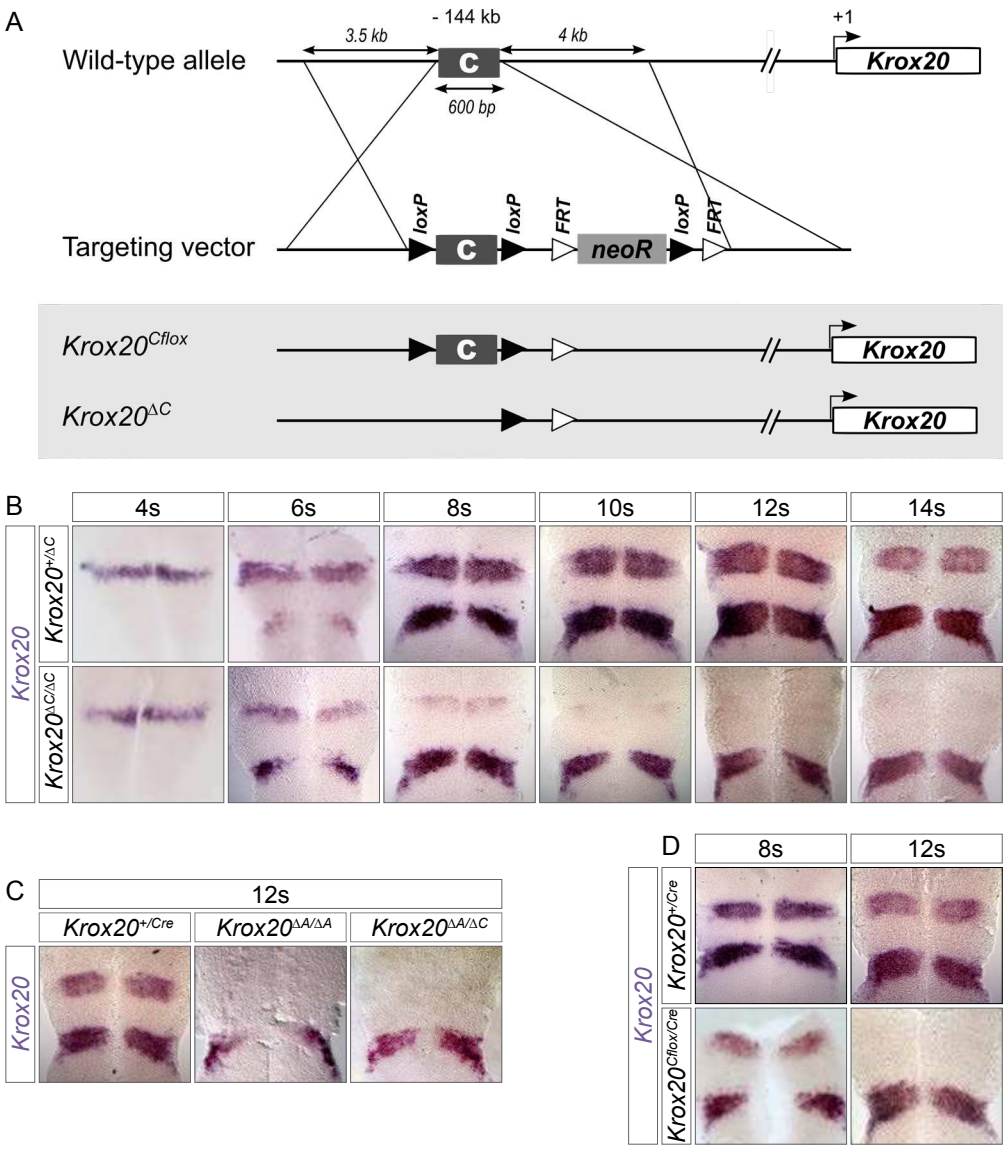


Figure 1: Genetic analysis of element C function.

(A) Strategy for the construction of conditional and null alleles of element C. The targeting vector was introduced into the locus in ES cells by homologous recombination and one of the ES clones subsequently allowed germ line transmission in the mouse. The floxed allele, *Krox20*^{Cflox}, was obtained by crossing the founder mouse line with a *Flp* (targeting FRT sites) deletor line. The null allele, *Krox20*$^{\Delta C}$, was obtained by crossing the *Krox20*^{Cflox} line with a *Cre* (targeting loxP sites) deletor line. **(B)** *In situ* hybridization was performed with a *Krox20* probe on *Krox20*$^{+/\Delta C}$ and *Krox20*$^{\Delta C/\Delta C}$ embryos at the indicated somite (s) stages. **(C)** *In situ* hybridization was performed with a *Krox20* probe on *Krox20*$^{+/Cre}$, *Krox20*$^{\Delta A/\Delta A}$ and composite heterozygous *Krox20*$^{\Delta C/\Delta A}$ embryos at 12s. **(D)** *In situ* hybridization was performed with a *Krox20* probe on *Krox20*$^{+/Cre}$ and *Krox20*$^{Cflox/Cre}$ embryos at the indicated somite stages.

to *Krox20*$^{\Delta C/\Delta C}$ embryos (Fig. 1B,C). This observation contrasts with the fact that, in heterozygous embryos, in which there is a single wild-type allele, *Krox20* is normally maintained. Moreover, we have previously shown that even with initial levels of KROX20 protein reduced to less than 50%, a single allele is sufficient to fully activate and maintain the autoregulatory loop (Bouchoucha et al. 2013). As in *Krox20*$^{\Delta C/\Delta C}$ embryos, the levels of *Krox20* mRNA, and presumably of KROX20 protein, are normal up to 6s in *Krox20*$^{\Delta A/\Delta C}$ mutants (data not shown). It suggests that the KROX20 protein should not be a limiting factor in the activation of the wild-type allele of element A in *Krox20*$^{\Delta A/\Delta C}$ embryos. Therefore, the most likely explanation of our results is that, in r3, the deletion of element C impairs the activity of the element A located on the same chromosome and that the two elements cooperate in *cis* for the establishment and/or maintenance of the autoregulatory loop. Such a possibility does not preclude an early involvement of element C, for instance to poise element A for the subsequent autoregulation phase. To investigate whether the presence of a functional element C is sufficient during the early phase of *Krox20* expression or whether it is required during the autoregulation phase, we generated a genetic situation in which element C is initially active, but is later deleted. This was obtained by combining the *Krox20*^{Cflox} allele with a knock-in allele, *Krox20*^{Cre}, in which the *Krox20* coding sequence has been replaced by the coding sequence of the Cre recombinase (Voiculescu et al. 2000). In such embryos, *Krox20* and *Cre* are expected to be synthetized at early somatic stages. Subsequently, the recombinase leads to deletion of element C in r3 and r5. Analysis of *Krox20*$^{Cflox/Cre}$ embryos revealed that, at 8s, the level of *Krox20* mRNA is reduced in r3 in *Krox20*$^{Cflox/Cre}$ embryos as compared to *Krox20*$^{+/Cre}$ controls (Fig. 1D), although less dramatically than in *Krox20*$^{\Delta C/\Delta C}$ mutants (Fig. 1B). At 12s, *Krox20* expression is completely abolished in r3 in *Krox20*$^{Cflox/Cre}$ embryos (Fig. 1D). These data indicate that the presence of a functional element C during the early phase is not sufficient to rescue the maintenance of the autoregulatory loop. The higher level of *Krox20* in r3 as compared to *Krox20*$^{\Delta C/\Delta C}$ mutants is likely to correspond to abortive activation of this loop following Cre excision of element C.

In conclusion, these results show that element C is required during the autoregulation phase for the establishment of the *Krox20* feedback loop. They further sustain the idea that elements A and C cooperate in *cis* for establishing this loop in r3. This cooperation has to occur during the autoregulatory phase and element C is not simply required to poise element A for activity at the initiation stage.

Multiple physical interactions within the *Krox20* locus and correlation with gene activity

The existence of a cooperation in *cis* between elements A and C during the autoregulatory phase of *Krox20* expression led us to investigate the existence of possible physical interactions between the different *Krox20 cis*-acting elements, which are separated by large distances on the mouse chromosome (for instance 155 kb between elements A and C). In a first series of experiments, we used circular chromosome conformation capture (4C, Simonis et al. 2006) to investigate physical interactions within the *Krox20* locus. To determine whether such interactions might be dependent on the activity of the autoregulatory loop, we compared samples prepared from total embryos at embryonic day (E) 9.5, when *Krox20* is not transcribed anymore (Irving et al. 1996), and from embryo heads at E8.5, when the autoregulatory loop is fully active (Bouchoucha et al. 2013). In this later sample *Krox20*-positive cells represent approximately one tenth of the cell population. In both types of samples, we observed a strong interaction between the *Krox20* promoter and element A regions, using the promoter as a point of view, although the reversed point of view did not give such a clear answer (Fig. 2B). We also observed significant interactions between the promoter or element A and the element B (Fig. 2B). In contrast, no interaction was detected between element C and the promoter. Nevertheless, a possible interaction between element C and element A was suggested when taking element C as a point of view (Fig. 2B). No obvious qualitative difference was observed between the two types of samples.

In conclusion, this analysis suggests that element A, element B and the promoter physically interact, without regard to the activity of the locus. In the case of element C, an interaction with element A is suspected. As the experiment was carried out on samples containing only a small minority of cells belonging to r3 for technical reasons, negative results have to be considered with caution.

Analysis of embryonic stem cells by Hi-C (Dixon et al. 2012) allowed the identification of a TAD covering 720 kb and including the *Krox20* locus, its identified upstream hindbrain regulatory elements (Chomette et al. 2006) and a long downstream region (Fig. 2A, upper part). We examined this region in more details, revealing the existence of two sub-TADs whose common frontier precisely corresponds to the position of the gene (Fig. 2A, upper part). Intrigued by the absence of qualitative difference between the two embryonic stages in the 4C experiment, we decided to perform a quantitative analysis. Taking the *Krox20* promoter viewpoint, we quantified the relative cumulative levels of interactions along the upstream and downstream sub-TADs between E8.5 and E9.5 (Fig. 2B). We observed significant differences between the upstream and downstream sub-TADs, with higher relative

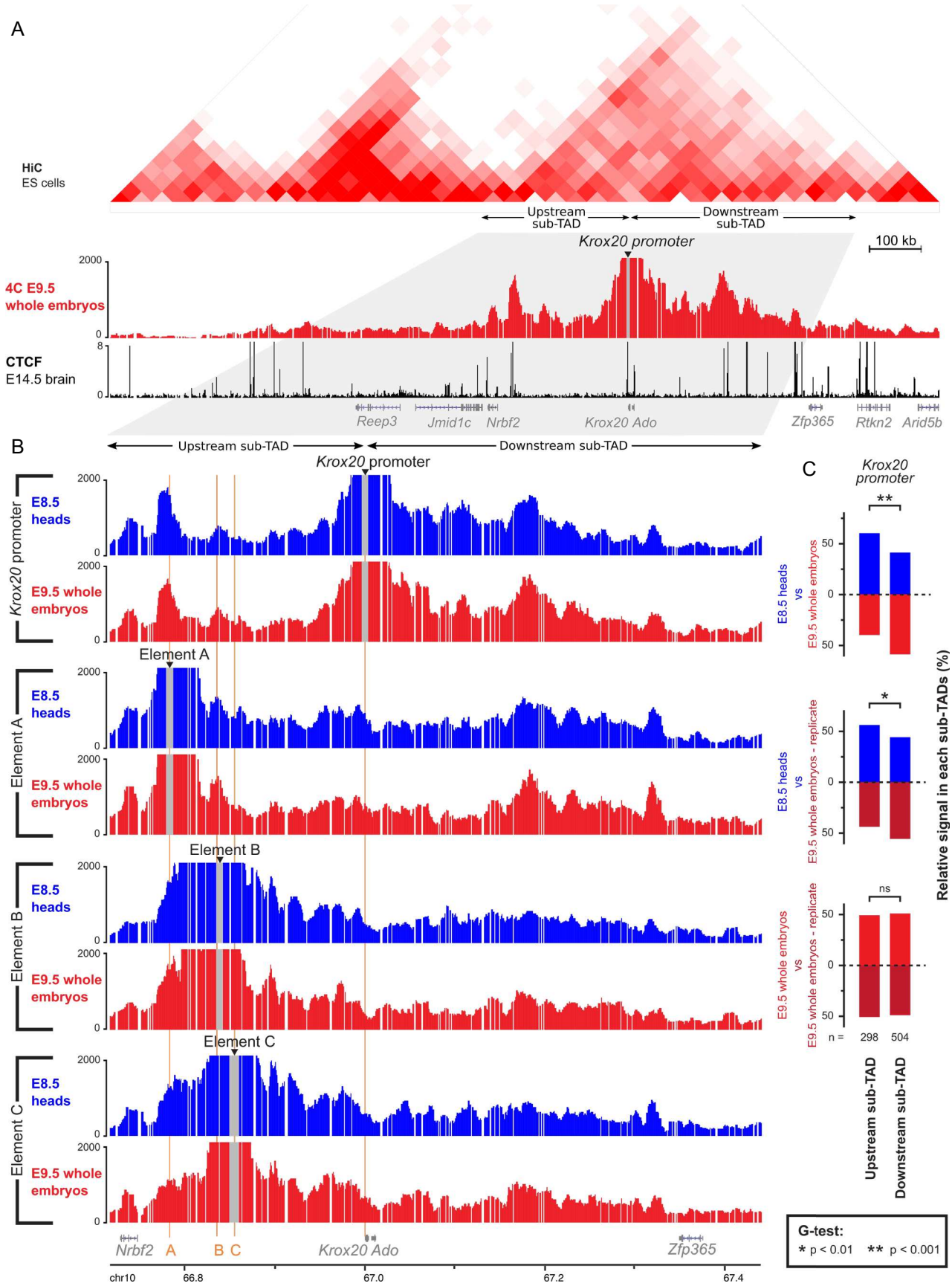


Figure 2: Physical interactions within the *Krox20* locus

(A) Alignment for the *Krox20* and adjacent loci of data obtained by HiC on ES cells (Dixon et al. 2012), by 4C performed on E9.5 whole mouse embryos using the *Krox20* promoter as viewpoint (this work) and by CTCF ChIP-seq on E14.5 mouse brain (ENCODE). Below, a TAD including the *Krox20* locus has been subdivided in upstream and downstream sub-TADs separated by the promoter. **(B)** Enlargement of the *Krox20* locus, showing 4C analyses using as viewpoints the *Krox20* promoter, element A, element B and element C, as indicated on the left. In blue, are shown signals obtained with E8.5 embryo heads, and in red signal obtained with E9.5 whole embryos. Genes, *cis*-regulatory elements and genomic coordinates are indicated below. **(C)** Quantitation of the relative cumulative levels of smoothed 4C-data along the upstream and downstream sub-TADs between E8.5 and E9.5 or E9.5 replicates from the *Krox20* promoter viewpoint. The relative levels (in %) are represented by the vertical bars for each pairwise comparison. Statistical significance of the differences (G-test) between upstream and downstream sub-TADs is indicated above each graph.

levels of interactions in E8.5 embryos in the upstream sub-TAD, whereas the reverse was observed in the downstream sub-TAD (Fig. 2C). In contrast, comparison of the signals in the two sub-TADs in two independent experiments performed at E9.5 did not reveal any significant difference, suggesting that the experiment was not biased (Fig. 2B). These observations are consistent with a modification of the chromatin organisation around the *Krox20* locus between E8.5 and E9.5, with the promoter favouring contacts within the upstream sub-TAD at E8.5 rather than with the downstream sub-TAD. This quantitative evolution is consistent with the period of activity of elements A, B and C, located in the upstream sub-TAD, and suggests that the contacts observed between the promoter and elements A and B (Fig. 2A) are also likely to reach a peak during this period.

Accessibility of element A is correlated with transcriptional activity and dependent on element C

To further investigate interactions between *Krox20 cis*-regulatory elements, we compared histones marks and chromatin accessibility between wild-type embryos and embryos deleted for element C. We first performed ChIP-seq experiments (Barski et al. 2007) on whole E9.5 embryos to analyze two histone modifications, H3K4me1 and H3K27ac, that usually mark enhancers (Zhou et al. 2011). A number of peaks were observed for the epigenetic mark H3K4me1, including ones corresponding to the promoter and elements A, B and C (Fig. 3A). Among the additional peaks, two major ones (M2 and M11) were shown in transgenic studies to correspond to enhancers active in the hindbrain (P. Torbey, personal communication). The signals observed for the H3K27ac mark was low across the *Krox20* locus apart from the promoter region, as it is expected since this mark is preferentially associated with promoters or active enhancers (Zhou et al. 2011) and *Krox20* enhancers should not be active anymore at E9.5. No significant difference in ChIP-seq signal was observed between wild-type and $Krox20^{\Delta C/\Delta C}$ embryos, apart from the deleted region C itself (Fig. 3A). In conclusion, all known *Krox20* hindbrain enhancers are labeled with the H3K4me1 mark at E9.5, but these signals do not appear to be dependent on the presence of element C.

In the above experiment, we could not analyze active enhancers at the appropriate territory and stage of development, because of the difficulty in accumulating sufficient biological material in which *Krox20* is expressed. To overcome these limitations, we assessed chromatin compaction by ATAC-seq (Buenrostro et al. 2013), as enhancer activity is usually associated with increased local chromatin accessibility and this technique can be conducted

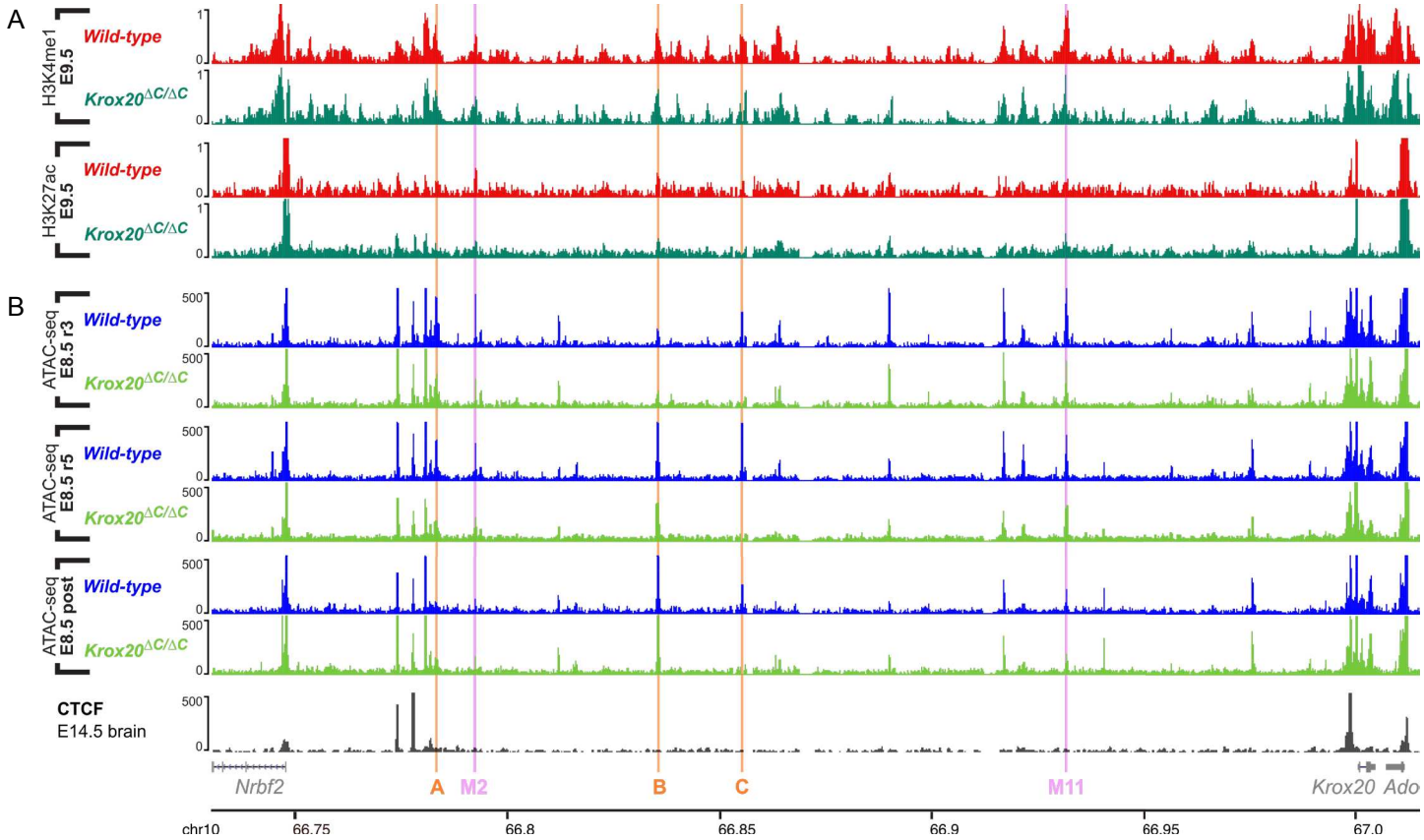


Figure 3: Chromatin accessibility within the *Krox20* locus.

(A) ChIP-seq was performed for the H3K4me1 and H3K27ac marks on wild-type (red) and *Krox20*$^{\Delta C/\Delta C}$ (dark green) E9.5 whole embryos. **(B)** ATAC-seq was performed on dissected parts ("r3", "r5" and "post", see text) wild-type (blue) and *Krox20*$^{\Delta C/\Delta C}$ (light green). Below, results for ChIP-seq for CTCF from E14.5 mouse brain (ENCODE) are indicated (see Fig. 2A). Genes, established (orange) or putative (pink) *cis*-regulatory elements and genomic coordinates are indicated at the bottom.

with much lower cell numbers (a few thousands). 8-10s stage embryos were micro-dissected to isolate samples corresponding to three regions of the embryo: an anterior region, not containing r5 and in which r3 cells constitute approximately one third of the total cell number ("r3" sample), an intermediate region, not containing r3 and in which r5 cells constitute approximately one third of the material ("r5" sample), and a posterior region, not containing r3 nor r5 (posterior sample). In wild-type embryos, we observed peaks of accessibility at the levels of the promoter as well as of each of the characterized hindbrain enhancers, A, B, C (Fig. 3B). The promoter was accessible in all three samples. Element A was specifically accessible in the "r3" and "r5" samples but not in the posterior sample (Fig. 3B). This absence of signal in the posterior part was expected as element A is only active in r3 and r5 (Chomette et al. 2006). Element B was poorly accessible in the "r3" sample, but accessible in the "r5" and posterior samples (Fig. 3B). This very limited accessibility of element B in r3 was also expected as this element is never active in r3 (Chomette et al. 2006). Element C was accessible in all three samples (Fig. 3B), consistent with the enhancer activity of element C which spans r3, r4, r5 and more posterior regions (Chomette et al. 2006). The new M2 and M11 elements presented patterns of accessibility similar to those of element A (for M2) and of element C (for M11) (Fig. 3B). In conclusion, this analysis revealed that the patterns of accessibility of the characterized enhancers are highly correlated with their activities and therefore constitute pertinent markers to assess the consequences of genetic modifications on their function.

We therefore went on to investigate the effects of the deletion of element C on the accessibility of the other elements. The accessibilities of elements B and M11 were not significantly affected by the deletion of element C (Fig. 3B). In contrast, the deletion specifically reduced the signals associated with elements A and M2 in r3 and r5 (Fig. 3B). These data establish that element C has the capacity to modulate the accessibility of element A and therefore probably its activity as well. They are consistent with above analyses indicating that *Krox20* autoregulation is dependent on the late activity of element C (Fig. 1).

In conclusion, element A is highly accessible in r3 and r5 and not in more posterior regions, consistently with the specific activity of this element that governs the *Krox20* autoregulatory loop in these rhombomeres. This accessibility appears modulated by element C, in agreement with the existence of a *cis*-interaction between A and C. These data therefore provide a first hint on the mechanisms that could be involved in this interaction in *cis*.

DISCUSSION

In the present study, we have further investigated the molecular mechanisms governing the expression of the *Krox20* gene in the developing mouse hindbrain, using both *cis*-acting element knock-out approaches and investigations of chromatin topology, modification and accessibility at the level of the locus. This reveals a complex interplay between at least four *cis*-acting elements and the promoter dispersed over a region of 220 kb, involving physical and functional interactions between different enhancers, with both redundant and synergistic modes of cooperation.

Redundant enhancers for the early phase of *Krox20* expression in r3

Previous analyses had suggested a rather simple mode of regulation of the *Krox20* gene, with element C responsible for the initiation of the expression in r3, element B, possibly together with element C, in charge of the initiation in r5, and element A governing the maintenance of the expression and a definitive switch toward an odd-numbered rhombomere fate through a positive feedback loop (Chomette et al. 2006; Labalette et al. 2011; Bouchoucha et al. 2013). Knock-out analysis of element C now leads to major revisions of this scenario. First, despite the early enhancer activity of element C observed in transgenesis experiments in the mouse and fish (Chomette et al. 2006; Labalette, Michel Adam Wassef, et al. 2015), we found that deletion of element C does not affect early *Krox20* expression in r3 (Fig. 1B). This suggests that another *cis*-acting element is responsible for promoting this expression. It is not likely to be element B, which is only active in r5 in transgenesis experiments (Chomette et al. 2006), nor element A, which is absolutely dependent on the presence of the KROX20 protein (Chomette et al. 2006; Bouchoucha et al. 2013). Therefore another enhancer, active in r3 and not dependent on KROX20, must be in charge. Possible candidates for this unknown element have been identified. Chromatin accessibility assessed by ATAC-Seq appears to be a particularly interesting indicator in this matter as i) a limited number of high accessibility peaks (in the order of 10) are observed within the sub-TAD region going from element A to the promoter (Fig. 3B); ii) the background level is very low with this type of data; iii) each of the known *Krox20* enhancer elements active during early embryogenesis (A, B and C, as well as a neural crest enhancer (Ghislain et al. 2003 and data not shown)) coincides with one of these peaks; iv) there is a good correlation between accessibility and enhancer activity: the elements A, B and C appear in open configuration precisely in the anterior-posterior regions where the TFs that activate them are present: Krox20 in r3 and r5 for element A (Chomette et al. 2006; Bouchoucha et al. 2013), Hnf1B from r5 to r8 for element B (Chomette et al. 2006;

Labalette et al. 2011) and Hox/Pbx/Meis factors in the hindbrain and posteriorly for element C (Wassef et al. 2008). These conclusions led us to further investigate the two regions corresponding to accessibility peaks named M2 and M11. Both carry H3K4me1 marks (Fig. 3A) and are conserved during evolution (P. Torbey, personal communication). Furthermore, zebrafish transgenesis experiments have shown that they possess enhancer activities in the hindbrain (P. Torbey, personal communication). Element M11 is particularly interesting as it is active early in r3, with a pattern similar to that of element C (P. Torbey, personal communication). Further analysis will be required, in particular its knock-out, to determine whether element M11 is indeed involved in governing the initial phase of *Krox20* expression in r3.

The absence of phenotype during the early phase of *Krox20* expression in the *Krox20*$^{\Delta C/\Delta C}$ mutant embryos does not preclude a role for element C during this phase. Several examples of redundancy have been observed among enhancers governing the expression of developmental genes (Hong et al. 2008; Kurokawa et al. 2004; Lam et al. 2015). In addition to the early activity of element C in r3, we have shown previously that its enhancer function is dependent on the binding of Meis and Hox/Pbx factors (Wassef et al. 2008), as well as on FGF signaling mediated by the Sp5 factor (Labalette et al. 2011; Labalette, M. A. Wassef, et al. 2015). These different factors are known to be required for *Krox20* expression in r3 (Wiellette & Sive 2003; S P Cordes & Barsh 1994; Waskiewicz et al. 2002; Choe et al. 2002; McNulty et al. 2005; Walshe et al. 2002; Aragon & Pujades 2009), supporting an implication of element C. Redundant enhancers often share the same regulatory logic, since their activities have to be, at least in part, concomitant (Lagha et al. 2012). The analysis of the characteristics of the M11 element will help to determine whether it depends of the same TFs as element C. In the few cases of redundancy between *cis*-acting elements that have been investigated in detail so far, it has been shown that redundancy provides robustness to the system, although, in specific genetic or environmental conditions, phenotypes can be revealed in absence of one of the elements (Ludwig et al. 2011).

Synergistic interactions between enhancers during the autoregulatory phase

Our study revealed a novel and unexpected function of element C: an interaction with element A that is required for the autoregulatory phase of the expression of *Krox20*. Using genetic approaches, we showed that this interaction must occur in *cis* and that it has to take place during the period of activity of element A, refuting the idea based on the early enhancer activity of element C, that it could simply be involved in poising element A at early stages,

before its actual activation. ATAC-seq experiments revealed that element C is likely to act by increasing the accessibility of element A, nevertheless the precise mechanism involved will require further studies. In this respect, it is interesting to note that we have previously found that Nlz factors are likely to repress *Krox20* expression by reducing the accessibility of the KROX20 protein to element A (Bouchoucha et al. 2013). Furthermore, we have shown that this repressive function may be mediated by element C (Labalette, Michel Adam Wassef, et al. 2015). Finally, several recent studies have shown that enhancers can be transcribed and that the products of this transcription, called eRNAs, can act locally in *cis* by promoting the expression of the target gene (Li et al. 2016). It can be proposed that element C might similarly be transcribed into eRNAs that act on element A in *cis* to promote *Krox20* expression.

In r5, *Krox20* is almost normally expressed in absence of element C, unlike what is observed in absence of element A (Bouchoucha et al. 2013), suggesting that a possible similar cooperation between elements A and C (suggested by the reduction of accessibility of element A in absence of element C) is not essential in this rhombomere to support element A activation. Element B, which is likely to constitute an initiator element in r5 and which we have shown to physically interact with element A (Fig. 2B), could also perform a dual function and play the key role in the cooperation with element A in this rhombomere.

Possible security locks on autoregulatory elements

The analysis of the organization of the *Krox20* locus by Hi-C in embryonic stem cells revealed the existence of a TAD including the locus and covering 720 kb (Dixon et al., 2012). Furthermore, 4C experiments performed in the present study indicated that a sub-domain within this TAD, including the *Krox20* promoter and all known hindbrain *cis*-acting elements, showed increased physical interactions upon the period of activity of the gene (E8.5), in particular between the element A and the promoter (Fig. 2C). ATAC-seq analysis indicated that these two latter regions present an open chromatin configuration at E8.5 (Fig. 3B) and ChIP experiment revealed the presence of strong CTCF binding sites in the vicinity of these elements (ENCODE). Together these data are consistent with the existence of a chromatin loop that would structure the domain between element A and the promoter and would play a decisive role in the capacity of element A, upon presence of the KROX20 protein, to stimulate transcription from the *Krox20* promoter and therefore in the establishment of the positive feedback loop necessary for rhombomere fate selection. This chromatin loop would

therefore establish a permissive, relatively stable, environment that would confer robustness in gene activation during development and differentiation (Montavon & Duboule 2012).
If they would only be dependent on their cognate TF, positive feedback loops may however constitute a potential danger for a cell or an organism, as their inappropriate activation could occur upon low level, illegitimate expression of the TF. Furthermore, developmental genes may possess several positive autoregulatory *cis*-elements that have to function at specific stages or in different tissues (for instance *Krox20* possesses at least three, r3/r5-, neural crest- and bone-specific, autoregulatory elements (Chomette et al. 2006; Ghislain et al. 2003; Monique Frain, personnal communication). Mechanisms must exist to prevent their spurious activation in some of the territories where the cognate protein is present. At least two types of barriers can be imagined to control the activity of positive *cis*-regulatory elements. First, we have previously shown in the case of the autoregulatory loop governed by element A that its activity relies on stochastic processes. As a result, to reach a reasonable probability of activating or maintaining it, a sufficient amount of KROX20 protein must be present (Bouchoucha et al. 2013). In these conditions, illegitimate transcription is not likely to lead to levels sufficient enough to activate the feedback loop. A second mechanism of control of the activity of the loop could reside in the accessibility of the autoregulatory *cis*-acting element for its cognate TF. If this property is controlled by another *cis*-acting element with no autoregulatory potential itself, like element C in the case element A, this will provide a powerful security lock that will place the autoregulatory loop under the control of the TFs that control the second *cis*-acting element (for instance, Hox, Pbx and Meis factors for element C), preventing its activation in inappropriate domains. In this way the cooperation between the two *cis*-acting elements becomes essential for establishing the specificity of the feedback loop. Such a security lock can nevertheless be transgressed, as we have shown that high levels of ectopic Krox20 in zebrafish embryos can led to activation of the autoregulatory loop in hindbrain regions where it is normally not functional (Bouchoucha et al., 2013), suggesting that high concentrations of the KROX20 protein can probably counter-balance the reduced accessibility of element A.

Conclusion

Apart from providing further insights into the complex regulation of *Krox20* in the developing hindbrain, this work has led to two observations of more general significance: i) a *cis*-acting element can control in *cis* the activity of a distant transcriptional enhancer, presumably by modulating the accessibility of the chromatin at the level of the enhancer; ii) a *cis*-acting

element can have dual functions at different stages of development and possibly in different tissues. Hence, the element C appears to first act as a classical transcriptional enhancer at early somite stages (although its actual implication in the regulation of *Krox20* expression at this stage is not yet established). Later on, as discussed above, the element C appears to modulate the *Krox20* autoregulatory loop by controlling the activity of element A. An interesting issue for the future is whether the two functions are separate or intricate. In other words, is the enhancer activity, which persists during the autoregulatory phase, required for *cis*-modulation of the other enhancer and are the known TFs that directly modulate its enhancer activity (Wassef et al. 2008; Labalette, Michel Adam Wassef, et al. 2015) also involved in the indirect modulation of element A activity? The introduction of specific point mutations into element C should provide cues to resolve this interesting issue.

This manuscript is not yet in its final stage as some experiments are still in progress. First, we want to analyze the chromatin marks H3K4me1 and H3K27ac at E8.5 in wild-type embryo heads by ChIP-seq. The number of embryos required to get enough material at this stage is important and we are currently collecting them. Second, we need replicates in the ATAC-seq experiment for control and Krox20$^{\Delta C/\Delta C}$ embryos and we want to compare the chromatin accessibility in Krox20$^{\Delta A/\Delta A}$ embryos and at later stages, when Krox20 in not expressed. These samples are currently being sequenced. Third, we are investigating the activity of element M2 and M11 by transgenesis in zebrafish. These different elements will be incorporated in the final manuscript that will also be completed (for instance introduction of the abstract). A more complete version of the paper will be provided to the referees before the thesis defense. We apologize for this situation.

REFERENCES

Aragon, F. & Pujades, C., 2009. FGF signaling controls caudal hindbrain specification through Ras-ERK1/2 pathway. *BMC developmental biology*, 9, p.61.

Banerji, J., Rusconi, S. & Schaffner, W., 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27(2 Pt 1), pp.299–308.

Barolo, S., 2012. Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *BioEssays*, 34(2), pp.135–141.

Barski, A. et al., 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, 129(4), pp.823–837.

Bouchoucha, Y.X. et al., 2013. Dissection of a Krox20 positive feedback loop driving cell fate choices in hindbrain patterning. *Molecular Systems Biology*, 9(690).

Buenrostro, J.D. et al., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods*, 10(12), pp.1213–8.

Choe, S.-K., Vlachakis, N. & Sagerström, C.G., 2002. Meis family proteins are required for hindbrain development in the zebrafish. *Development*, 129(3), pp.585–595.

Chomette, D. et al., 2006. Krox20 hindbrain cis-regulatory landscape: interplay between multiple long-range initiation and autoregulatory elements. *Development (Cambridge, England)*, 133(7), pp.1253–62.

Cordes, S.P. & Barsh, G.S., 1994. The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transcription factor. *Cell*, 79(6), pp.1025–1034.

Cordes, S.P. & Barsh, G.S., 1994. The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transcription factor. *Cell*, 79(6), pp.1025–34.

David, F.P.A. et al., 2014. HTSstation: A web application and open-access libraries for high-throughput sequencing data analysis. *PLoS ONE*, 9(1).

Dixon, J.R. et al., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), pp.376–80.

Dowen, J.M. et al., 2014. Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. *Cell*, 159(2), pp.374–387.

Ghislain, J. et al., 2003. Neural crest patterning: autoregulatory and crest-specific elements co-operate for Krox20 transcriptional control. *Development (Cambridge, England)*, 130(5), pp.941–953.

Gibcus, J.H. & Dekker, J., 2013. The hierarchy of the 3D genome. *Molecular cell*, 49(5), pp.773–82.

Giudicelli, F. et al., 2001. Krox-20 patterns the hindbrain through both cell-autonomous and non cell-autonomous mechanisms. *Genes & development*, 15(5), pp.567–80.

Hadjur, S. et al., 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*, 460(7253), pp.410–413.

Heintzman, N.D. et al., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*, 39(3), pp.311–8.

Heintzman, N.D. et al., 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459(7243), pp.108–112.

Hong, J.-W., Hendrix, D. a & Levine, M.S., 2008. Shadow enhancers as a source of evolutionary novelty. *Science (New York, N.Y.)*, 321(5894), p.1314.

Irving, C. et al., 1996. Progressive spatial restriction of Sek-1 and Krox-20 gene expression during hindbrain segmentation. *Developmental biology*, 173(1), pp.26–38.

Kurokawa, D. et al., 2004. Regulation of Otx2 expression and its functions in mouse forebrain and midbrain. *Development (Cambridge, England)*, 131(14), pp.3319–31.

Labalette, C. et al., 2011. Hindbrain patterning requires fine-tuning of early krox20 transcription by Sprouty 4. *Development (Cambridge, England)*, 138(2), pp.317–26.

Labalette, C., Wassef, M.A., et al., 2015. Molecular dissection of segment formation in the

developing hindbrain. *Development*, 142(1), pp.185–195.
Labalette, C., Wassef, M.A., et al., 2015. Molecular dissection of segment formation in the developing hindbrain. *Development (Cambridge, England)*, 142(1), pp.185–95.
Lagha, M., Bothma, J.P. & Levine, M., 2012. Mechanisms of transcriptional precision in animal development. *Trends in genetics : TIG*, 28(8), pp.409–16.
Lam, D.D. et al., 2015. Partially Redundant Enhancers Cooperatively Maintain Mammalian Pomc Expression Above a Critical Functional Threshold. *PLOS Genetics*, 11(2), p.e1004935.
Lettice, L.A. et al., 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, 12(14), pp.1725–1735.
Li, W., Notani, D. & Rosenfeld, M.G., 2016. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nature Reviews Genetics*, 17(4), pp.207–223.
Ludwig, M.Z. et al., 2011. Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. *PLoS Genetics*, 7(11).
Lumsden, A. & Keynes, R., 1989. Segmental patterns of neuronal development in the chick hindbrain. *Nature*, 337(6206), pp.424–8.
Lumsden, A. & Krumlauf, R., 1996. Patterning the vertebrate neuraxis. *Science (New York, N.Y.)*, 274(5290), pp.1109–15.
Maniatis, T.O.M., Goodbourn, S. & Fischer, J. a, 1987. Inducible and.
McNulty, C.L. et al., 2005. Knockdown of the complete Hox paralogous group 1 leads to dramatic hindbrain and neural crest defects. *Development (Cambridge, England)*, 132(12), pp.2861–71.
Montavon, T. & Duboule, D., 2012. Landscapes and archipelagos: spatial organization of gene regulation in vertebrates. *Trends in cell biology*, 22(7), pp.347–54.
Noordermeer, D. et al., 2014. Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. *eLife*, 2014(3), pp.1–21.
Nora, E.P. et al., 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398), pp.381–5.
Schneider-Maunoury, S. et al., 1993. Disruption of Krox-20 results in alteration of rhombomeres 3 and 5 in the developing hindbrain. *Cell*, 75(6), pp.1199–214.
Schneider-Maunoury, S. et al., 1997. Segmental and neuronal architecture of the hindbrain of Krox-20 mouse mutants. *Development (Cambridge, England)*, 124(6), pp.1215–26.
Sexton, T. et al., 2012. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148(3), pp.458–472.
Simonis, M. et al., 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature genetics*, 38(11), pp.1348–54.
Symmons, O. et al., 2014. Functional and topological characteristics of mammalian regulatory domains. *Genome Research*, 24(3), pp.390–400.
Vitobello, A. et al., 2011. Hox and Pbx Factors Control Retinoic Acid Synthesis during Hindbrain Segmentation. *Developmental Cell*, 20(4), pp.469–482.
Voiculescu, O. et al., 2001. Hindbrain patterning: Krox20 couples segmentation and specification of regional identity. *Development (Cambridge, England)*, 128(24), pp.4967–78.
Voiculescu, O., Charnay, P. & Schneider-Maunoury, S., 2000. Expression pattern of a Krox-20/Cre knock-in allele in the developing hindbrain, bones, and peripheral nervous system. *Genesis (New York, N.Y. : 2000)*, 26(2), pp.123–6.
Walshe, J. et al., 2002. Establishment of hindbrain segmental identity requires signaling by FGF3 and FGF8. *Current Biology*, 12(13), pp.1117–1123.
Waskiewicz, A.J., Rikhof, H. a & Moens, C.B., 2002. Reveals a Hindbrain Ground State.

Developmental Cell, 3, pp.723–733.
Wassef, M. a et al., 2008. Rostral hindbrain patterning involves the direct activation of a Krox20 transcriptional enhancer by Hox/Pbx and Meis factors. *Development (Cambridge, England)*, 135(20), pp.3369–78.
Wiellette, E.L. & Sive, H., 2003. vhnf1 and Fgf signals synergize to specify rhombomere identity in the zebrafish hindbrain. *Development (Cambridge, England)*, 130(16), pp.3821–3829.
Wilkinson, D.G. et al., 1989. Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. *Nature*, 337(6206), pp.461–4.
Zaret, K.S. & Carroll, J.S., 2011. Pioneer transcription factors: Establishing competence for gene expression. *Genes and Development*, 25(21), pp.2227–2241.
Zhou, V.W., Goren, A. & Bernstein, B.E., 2011. Charting histone modifications and the functional organization of mammalian genomes. *Nature reviews. Genetics*, 12(1), pp.7–18.

Table S1: Sequences of 4C primers with Illumina daptors

Primer names	Sequences
Krox20_ promoter	F_AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT TCCGTCACTGACTCTCTC R_CAAGCAGAAGACGGCATACGAAGATGCACCTGGTCACCA
Element A	F_AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT GAGGTAGGGAGAAACAGTAG R_CAAGCAGAAGACGGCATACGATGAGTCTTGGCAGCTAATG
Element B	F_AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT CTTTGTTCCTTTTGCTCAGT R_CAAGCAGAAGACGGCATACGAGATGTGACAGAGGTTGAGAG
Element C	F_AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTCCGATCT ATTTGAAAAATGATGTATCACTGAG R_CAAGCAGAAGACGGCATACGACTGACAGCCACAGGAAGT

- Chapitre 3 -

Développement de deux systèmes pour l'étude de la régulation du gène *Krox20* chez les Vertébrés

Ce chapitre résume deux approches que j'ai développées au laboratoire afin de pallier aux limitations du système embryonnaire murin *in vivo*. En effet, les questions posées sur le rôle des éléments *cis*-régulateurs du gène *Krox20* dans le rhombencéphale nécessiteraient pour y répondre, dans l'idéal, un grand nombre de cellules d'identité homogène. C'est pourquoi nous avons entrepris de développer dans un premier temps un système *in vitro* de différenciation de cellules souches embryonnaires murines en cellules d'identité rhombencéphalique exprimant le gène *Krox20*. Dans un second temps, nous avons tiré avantage de l'extrême conservation du patron d'expression du gène *Krox20* chez les Vertébrés en utilisant le modèle du poisson-zèbre. L'idée était de reproduire les délétions des éléments *cis*-régulateurs identifiés chez le poisson-zèbre pour analyser, d'une part, dans quelle mesure les mécanismes de régulation de *Krox20* sont conservés. D'autre part si ce modèle permet d'obtenir facilement un grand nombre d'embryons, il pourrait être utile pour décrypter les mécanismes moléculaires mis en œuvre dans la régulation de l'expression de *Krox20*.

1. Dérivation de cellules souches embryonnaires murines

1.1 Introduction

Les cellules souches embryonnaires (ES) sont des cellules pluripotentes issues de la masse interne des blastocystes. Les cellules ES sont capables de former, par différenciation, les trois couches primitives (endoderme, mésoderme et ectoderme) de l'organisme, de même que la lignée germinale. Afin d'étudier les premiers stades du développement, les cellules ES sont différenciées en structures tri-dimensionnelles appelées corps embryonnaires (EB) du fait de leur capacité à mimer les tissus embryonnairse post-implantatoires. Il est ensuite possible de dériver ces EBs en différents types de neurones. L'induction vers la voie neurale suit le 'modèle par défaut' par inhibition de la voie BMP (Bone Morphogenetic Protein). La différentiation antéro-postérieure du tube neural est sous le contrôle de plusieurs facteurs de signalisation majeurs tel l'acide rétinoïque (AR), les FGFs et les Wnts. En collaboration avec Stéphane Nédélec (i-Stem, Evry), j'ai développé un protocole de différenciation de cellules ES en progéniteurs neuronaux Krox20^{+}. Ce protocole est basé sur la double inhibition des SMADs qui sont des effecteurs de la voie BMP (Chambers et al. 2009) et sur l'action combinée des facteurs de différenciation cités précédemment.

1.2 Matériels et Méthodes

1.2.1 Préparation des fibroblastes embryonnaires de souris et culture des cellules ES

Les fibroblastes embryonnaires de souris (13,5 jpc) sont cultivés en DMEM contenant 10 % de sérum de veau fœtal (SVF), 1 % de L-glutamine, 1 % de pyruvate de sodium, 1% d'acides aminés non-essentiels, 1% de pénicilline/streptamycine et 0,2 % de β-mercaptoéthanol. Après un premier passage, les fibroblastes à confluence sont traités avec 5μL de mitomycine C (1mg/mL) dans 5mL de milieu pendant 2 à 3h à 37°C. Les cellules sont ensuite lavées 3 fois en PBS 1X puis rapidement trypsinées et étalées à une concentration de 2,5.10^{6} cellules/mL sur boîtes B60 gélatinées à 0,1 %. Le milieu est changé tous les deux jours.
Les cellules ES *Krox20*$^{GFPDT/+}$ (5.10^{6} cellules/mL) sont étalées sur les fibroblastes mitomycinés dans le milieu suivant : Knockout DMEM contenant 15% de SVF, 1% de L-glutamine, 1% d'acides aminés non-essentiels, 1% de pénicilline/streptamycine, 1% de nucléosides, 0,2% de β-mercaptoéthanol et 50u de LIF (Leukemia inhibitory factor). Le lendemain, le milieu est changé et, 48h plus tard, les cellules sont passées et étalées à une concentration de 2. 10^{6} cellules/mL.

1.2.2 Différenciation des cellules ES en EBs et induction en progéniteurs neuronaux

Les cellules ES à confluence sont trypsinées et diluées en milieu de différentiation N2B27 : contenant pour moitié du ADMEM/F12 et du neurobasal medium ainsi que 1% de pénicilline/streptamycine, 1% de L-glutamine, 0,2% de β-mercaptoéthanol, 2% de supplément B27 sans vitamine A et 1% de supplément N2 avant d'être centrifugées à 1200 rpm pendant 5 min. Les cellules sont resuspendues dans 10mL de milieu N2B27 et comptées afin d'être cultivées en boîte de Pétri non-adhérentes à une concentration comprise entre 5.10^4 et 1.10^5 cellules/mL en milieu N2B27 supplémenté en acide ascorbique (10μM), SB431542 (20μM) et LDN193189 (0,1μM), des inhibiteurs des voies TGF-β et BMP, respectivement. Le jour suivant, les EBs sont transférées dans un Falcon de 15mL et centrifugées à 500 rpm pendant 3min. Du milieu frais N2B27 supplémenté en SB431542 (20μM), LDN193189 (0,1μM) et acide ascorbique (10μM) est ajouté aux EBs qui sont réétalées sur de nouvelles boîtes non adhérentes.

Deux jours après le début de la différentiation les EBs sont transférées en Falcon et centrifugées à 500 rpm pendant 3min puis rincées en PBS 1X avant d'être divisés en 4 boîtes de Pétri non-adhérentes en milieu N2B27 supplémenté en SB431542 (20μM), LDN193189 (0,1μM) et acide ascorbique (10μM) et contenant des concentrations à définir de l'acide rétinoïque, du CHIR99021 et du FGF2. Le milieu est changé le lendemain si le temps d'induction est supérieur à 24h.

1.2.3 RT-PCR et RT-qPCR

A la fin de l'induction, 1/3 des cellules sont utilisées pour l'extraction d'ARN réalisée avec le kit RNAqueous micro (Ambion, Life Applied) en suivant les recommandations du fabricant avec un traitement à l'ADNase I, 20 min à 37°C. La transcription inverse est effectuée avec l'enzyme superscript III (Invitrogen, Life Technologies) à partir de 500ng d'ARN. Après un traitement à la RNase H de 20 min à 37°C, les ADNc sont purifiés sur colonnes QIAquick (QIAGEN). La PCR a été réalisée avec une polymérase classique pendant 35 cycles à une température d'hybridation de 60°C. La PCR quantitative a été réalisée avec le SYBER Green Master Mix (Roche Applied Biosystems) en utilisant un LightCycler 480 (Roche Applied Biosystems). Pour les deux types de PCR, les amorces suivantes ont été utilisées : mKrox20, eGFP, EphA4, Hoxb1, Otx2, Hoxb4, Hoxb9, β-actine.

1.2.4 Immunohistochimie

Les 2/3 des cellules restantes sont fixées en 4 % paraformaldéhyde pendant 10 min, lavées trois fois en PBS 1X. Les sphères sont ensuite incluses en OCT et congelées à -80°C jusqu'à

la découpe des blocs au cryostat d'une épaisseur 12μm. Les lames sont bloquées en PBS avec 4% de BSA et 0,1% de Triton avant ajout de l'anticorps de rat anti-GFP (1/500ème, Nacalaï Tesque) toute la nuit. Après rinçage, l'anticorps secondaire de chèvre anti-rat couplé à l'Alexa 488 (1/200ème, Jackson ImmunoResearch) est ajouté pendant 2h à température ambiante et lavé plusieurs fois avant l'ajout du Hoeschst. Les lames sont ensuite montées en fluoromount et observées au microscope à épifluorescence.

1.3 Résultats

1.3.1 Définition des meilleures conditions d'activation du locus

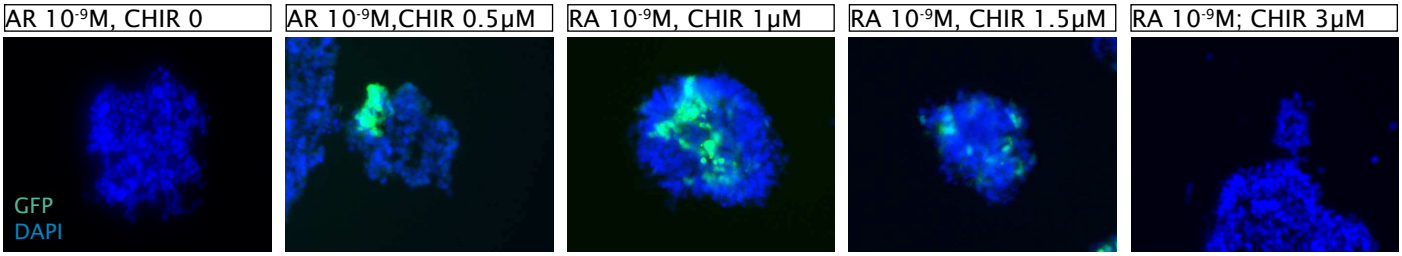
Afin de suivre l'expression du gène *Krox20*, j'ai utilisé les cellules ES *Krox20*$^{+/GFP}$ qui ont permis de générer la lignée de souris Knock-In *Krox20*$^{GFP(DT)}$ (Vermeren et al. 2003). Le protocole de Stéphane Nédélec a été établi pour différencier des cellules ES en motoneurones spinaux en utilisant une forte concentration de CHIR99021, un activateur de la voie Wnt, qui permet une identité neurale postérieure au rhombencéphale en présence d'acide rétinoïque (Maury et al. 2014). Dans un premier temps, nous avons réalisé un large criblage de 96 conditions pour le CHIR99021, le FGF2 et l'AR (fig. 15A) en plaque 384 puits. On observe très nettement une anti-corrélation pour l'activation du locus entre le FGF2 et l'AR. Une forte concentration de FGF2 couplée à une faible concentration d'AR entraîne une plus grande quantité de cellules GFP$^+$ dans une gamme allant de CHIR99021 de 0,5μM à 1μM, et inversement. En parallèle de ce large criblage, une expérience en plaque 6 puits a été réalisée (fig. 15B) et montre une activation du locus cette fois à faible concentration d'AR sans FGF2 pour une gamme de CHIR99021 s'étendant de 0.5μM à 1,5μM. Il est possible d'expliquer ce résultat par une plus forte proportion de cellules dans la plaques 6 puits qui peuvent elles aussi sécréter du FGF dans le milieu. J'ai donc décidé d'optimiser ces conditions en plaque 6 puits (fig. 15C) et de vérifier la différenciation après seulement 8h, 24h et 32h puis à 48h. La différenciation débute à 32h, mais l'activation du locus est plus importante à 48h (fig. 15C). L'ajout de FGF2 à des concentrations de CHIR99021 0,75μM et 1μM et d'AR à 10^{-9}M semblent être les meilleures conditions pour obtenir des cellules GFP$^+$. Une estimation grossière fait état de 10 % de cellules exprimant *Krox20*. De plus, il semblerait que les cellules ayant activé le locus sont préférentiellement localisées à la surface de la sphère, suggérant que les facteurs de différenciation n'agissent pas de façon homogène sur les cellules dans cette configuration.

A

	CHIR 3µM		ChIR 1,5µM		CHIR 1µM		CHIR 0,75µM		CHIR 0,5µM		CHIR 0	
	AR 0	AR 10^{-9}M	AR 0	AR 10^{-9}M	AR 0	AR 10^{-9}M	AR 0	AR 10^{-9}M	AR 0	AR 10^{-9}M	AR 0	AR 10^{-9}M
FGF2 0	-	-	-	-	-	-	-	-	-	-	-	-
FGF2 5ng/mL	-	-	-	-	-	++	-	-	-	-	-	-
FGF2 15ng/mL	-	-	-	-	-	++	-	++	-	+	-	-
FGF2 50ng/mL	-	-	-	-	-	+++	-	++	-	++	-	-
	AR 10^{-8}M	AR 10^{-7}M	AR 10^{-8}M	AR 10^{-7}M	AR 10^{-8}M	AR 10^{-7}M	AR 10^{-8}M	AR 10^{-7}M	AR 10^{-8}M	AR 10^{-7}M	AR 10^{-8}M	AR 10^{-7}M
FGF2 0	-	-	-	-	++	-	+++	-	+++	-	-	-
FGF2 5ng/mL	-	-	-	-	++	-	++	-	++	-	-	-
FGF2 15ng/mL	-	-	-	-	+	-	++	-	++	-	-	-
FGF2 50ng/mL	-	-	-	-	++	-	++	-	-	-	-	-

- : aucune cellule GFP+
+ : quelques cellules GFP+
++ : petit nombre de cellules GFP+
+++ : grand nombre de cellules GFP+

B



C

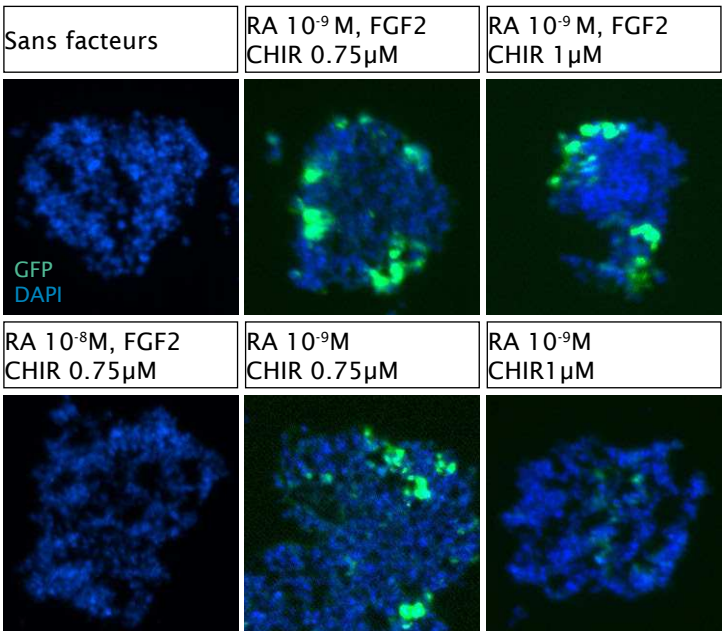


Figure 15 : Criblage des conditions de culture pour l'obtention de cellules GFP+

A. Tableau présentant les différentes conditions de différenciation des EBs utilisés pour un criblage en plaque 384 : 4 concentrations de FGF2, 4 concentrations d'acide rétinoïque (AR) et 6 concentrations de CHIR99021 (CHIR) après 48h.

B & C. Immunohistochimie sur des coupes d'EBs après 48h de différenciation en plaque 6 puits contre la GFP. En B. seule la concentration de CHIR99021 varie alors que la concentration d'acide rétinoïque est identique. En C. différentes conditions de facteurs sont testées dont l'ajout de FGF2 (20ng/mL) en comparaison avec l'induction seule.

1.3.2 Identité des cellules différenciées

Pour connaître l'identité moléculaire des cellules différenciées, l'expression des gènes clefs de l'identité du tube neural en développement a été quantifiée. Le gène *Otx2* à une expression antérieure dans le mésencéphale et le prosencéphale. *Krox20* est exprimé dans r3 et r5 et active le gène cible *EphA4* dans ces mêmes rhombomères. Ce dernier est en plus faiblement exprimé dans r2. Pour les *Hox*, nous nous sommes intéressé à *HoxB1* exprimé dans r4, *Hoxb4* est exprimé à partir de r7 et dans la moelle et *Hoxb9* uniquement dans la moelle (fig. 16A). Dans un premier temps, l'expression de ces gènes a été quantifiée par RT-PCR sur la population entière de cellules différenciées à l'acide rétinoïque (10^{-9}M) et un gradient de concentration de CHIR99021 (fig. 16B). On n'observe pas de différence d'expression pour le gène *Krox20* et la *GFP* en fonction des concentrations de CHIR99021, contrairement aux résultats précédents d'immunohistochimie (fig. 15B et 16B). Ceci peut s'expliquer par le fait que le RT-PCR n'est pas une technique quantitative et qu'une très faible activation du locus est donc détectable. De plus, *Krox20* ou *Egr2* (Early Growth Response 2) a été identifié par sa réponse précoce au sérum (Chavrier et al. 1988) et on peut supposer que les cellules en culture activent le locus en réponse aux facteurs de croissance du milieu. Le gène *EphA4* est exprimé de façon plus importante à une concentration de CHIR99021 entre 0,5 et 1μM. L'expression du gène *Otx2* dans les EBs est plus importante à faible concentration de CHIR99021 comme attendu. Pour les gènes *Hox*, *Hoxb1* semble avoir une expression constante dans les EBs quelle que soit la concentration de CHIR99021, contrairement aux *Hox* postérieurs *Hoxb4* et *Hoxb9* dont l'expression est restreinte à une concentration de CHIR99021 de 3μM en accord avec ce qu'observe Stéphane Nédélec pour la différentiation en motoneurones spinaux. Afin d'avoir une quantification plus fine des niveaux d'expression de ces gènes, j'ai réalisé une RT-qPCR sur des EBs différenciées à l'acide rétinoïque (10^{-9}M), au FGF2 (20ng/mL) et à deux concentrations de CHIR99021 à 0,75 ou 1μM où l'activation du locus était la meilleure (fig. 16C). Par comparaison avec des EBs non induites par ces facteurs, on observe une diminution de l'expression du gène antérieur *Otx2*, et une augmentation pour les gènes du rhombencéphale *Krox20* et *EphA4* ainsi que *Hoxb4* postérieur pour les deux concentrations. Cependant, l'augmentation est plus faible à 0,75μM de CHIR99021 pour les trois gènes, démontrant une balance entre une identité rhombencéphalique et spinale qui demande donc une optimisation des conditions de différenciation.

J'ai ensuite essayé de caractériser l'identité moléculaire des EBs au niveau cellulaire par immunohistochimie et hybridation *in situ* contre *Krox20* et *Hoxb1* car leurs expressions sont mutuellement exclusives dans l'embryon (Giudicelli et al. 2001). L'immunohistochimie

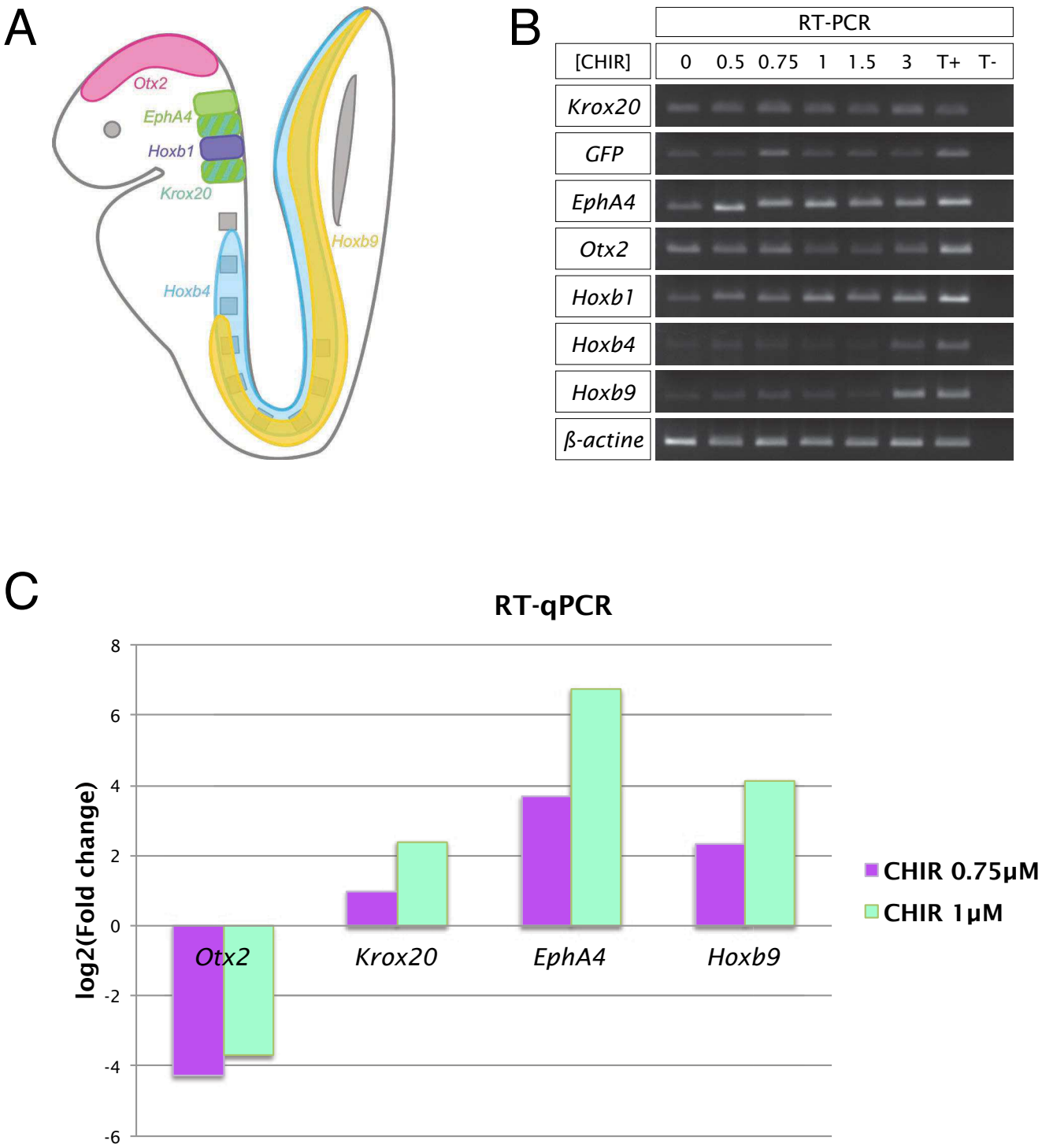


Figure 16 : Identité des cellules après induction

A. Embryon de souris schématique à 8,5 jpc montrant le patron d'expression des gènes suivants : *Otx2* dans le mésencéphale et prosencéphale ; *EphA4* dans r2, r3 et r5, *Krox20* dans r3 et r5, *Hoxb1* dans r4, *Hoxb4* à partir de r7 et dans la moelle épinière et *Hoxb9* uniquement dans la moelle épinière. Les carrés représentent les somites.

B. RT-PCR sur les EBs différenciés en plaque 6 puits et différenciés à l'acide rétinoïque (10^{-9}M) et à des concentrations croissantes de CHIR (en μM) utilisant des primers contre des gènes exprimés chez l'embryon de souris à 8,5 jpc. La ß-actine est le témoin de charge. Des embryons sauvages ou *Krox20*$^{+/GFP(DT)}$ à 8,5 jpc ont été utilisés en témoin positif.

C. RT-qPCR sur les EBs différenciés en plaque 6 puits à l'acide rétinoïque (10^{-9}M), au FGF2 (20 ng/mL) et deux concentrations croissantes de CHIR (en μM) utilisant des amorces contre des gènes exprimés chez l'embryon de souris à 8,5 jpc et rapporté aux EBs non induites par ces facteurs.

contre ces deux protéines s'est révélée infructueuse. L'hybridation *in situ* contre ces deux ARNm n'a quant à elle pas permis de discriminer les cellules des EBs.

1.4 Conclusions et perspectives

En collaboration avec Stéphane Nédélec, j'ai développé un système cellulaire permettant de dériver des cellules ES murines en cellules d'identité rhombencéphalique exprimant le gène *Krox20*. Cependant, ce système ne comporte qu'une faible proportion de cellules Krox20/GFP^{+} et mériterait d'être peaufiné pour obtenir une population plus homogène et dont l'identité est clairement définie. De plus, nous aurions aimé pouvoir utiliser des cellules ES homozygotes pour les délétions des éléments A et C, ce qui n'est pas compatible avec un tri cellulaire en utilisant l'allèle *Krox20*$^{GFP(DT)}$. Enfin, il faudrait tester l'activité des éléments *cis*-régulateurs de *Krox20* caractérisés dans le rhombencéphale chez la souris par transfection de plasmides rapporteurs. En effet, bien que les cellules Krox20+ semblent posséder une identité rhombencéphalique, il n'est pas garanti que l'activation du locus soit dépendante des éléments *cis*-régulateurs actifs dans le rhombencéphale car d'autres éléments contrôlent l'expression pléiotropique du gène *Krox20*. Stéphane Nédélec s'intéresse à présent à la différentiation des motoneurones issus du rhombencéphale, ce système lui servira de base pour ces futurs travaux. Il a notamment mis en évidence un effet antagoniste entre le CHIR99021 et l'acide rétinoïque dont l'ajout doit être découplé dans le temps. Dans le même temps, un nouveau protocole de différenciation de cellules souches humaines en cellules neuroépithéliales en monocouche a été mis au point pour s'affranchir à la fois du passage en EBs et de l'utilisation d'inhibiteurs (Lippmann et al. 2014). Il a été appliqué pour la dérivation d'ES en neuroectoderme grâce à l'activation maîtrisée et séquentielle des gènes *HOX* (Lippmann et al. 2015). Ce protocole constitue une alternative intéressante pour l'obtention de progéniteurs neuronaux d'identité rhombencéphalique à partir de cellules souches murines et permettra peut-être une accessibilité plus homogène des cellules aux différents facteurs de différenciation par rapport aux EBs.

2. Délétion d'un enhancer chez le poisson-zèbre

2.1 Introduction

Le poisson-zèbre est un modèle idéal pour générer un grand nombre d'embryons rapidement. De plus, le développement de la technique de CRISPR/Cas9 permet de créer facilement des mutations chez cet organisme par co-injection au stade 1 cellule d'ARNs guides (ARNg) qui ciblent la région à muter et d'ARNm de la nucléase Cas9. Ceci permet une coupure double brin ciblée de l'ADN pris en charge ensuite par la machinerie de réparation de l'ADN et d'introduire soit des petites mutations, soit des délétions si on couple deux ARN guides (Hwang et al. 2013). J'ai donc entrepris de mettre au point cette technique chez le poisson-zèbre afin de déléter l'élément C dont la partie cœur est bien conservée en séquence (fig. 17A) et en activité par transgénèse chez le poisson de l'élément C de poisson (données non montrées) ou de poulet (Labalette et al. 2011).

2.2 Matériels et méthodes

2.2.1 Génération des ARNs guides et de l'ARNm de la Cas9 et injection

Les ARNs guides sont générés par hybridation de 2 oligonucléotides (15μg chacun), dont les séquences sont les suivantes : 5'-AAA CAA ACT CAA TAC CTA TGT A-3' et 5'-TAG GTA CAT AGG TAT TGA GTT T-3' (ARNg 1), 5'-AAA CGG GGT GAA GAC ACA AAG G-3' et 5'-TAG GCC TTT GTG TCT TCA CCC C-3' (ARNg 2), 5'-AAA CCA TTT ATA CAG CTA AAC TTT-3' et 5'-TAG GAA AGT TTA GCT GTA TAA ATG-3' (ARNg3), 5'-AAA CTC TAC CTA TTC TTT TAT T-3' et 5'-TAG GAA TAA AAG AAT AGG TAG A-3' (ARNg 4). Ils sont ensuite clonés dans le vecteur pDR274 (Addgene, Hwang et al. 2013), préalablement linéarisé par BsaI. Le plasmide recombinant est amplifié en culture bactérienne et extrait par maxiprep. Le plasmide est ensuite linéarisé par DraI et purifié par une extraction en phénol/chloroforme. La transcription en ARNs est effectuée avec le MAXIscript kit T7 (Ambion) en suivant les instructions du fabricant et les ARNs sont purifiés sur colonnes (SIGMA).

L'ARNm de la Cas9 est généré à partir du plasmide pT3TS-nCas9n (Addgene, Jao et al. 2013) linéarisé par XbaI en utilisant le mMessage mMachine T3 kit (Ambion) en suivant les instructions du fabricant. L'ARNm est ensuite purifié sur colonnes (SIGMA).

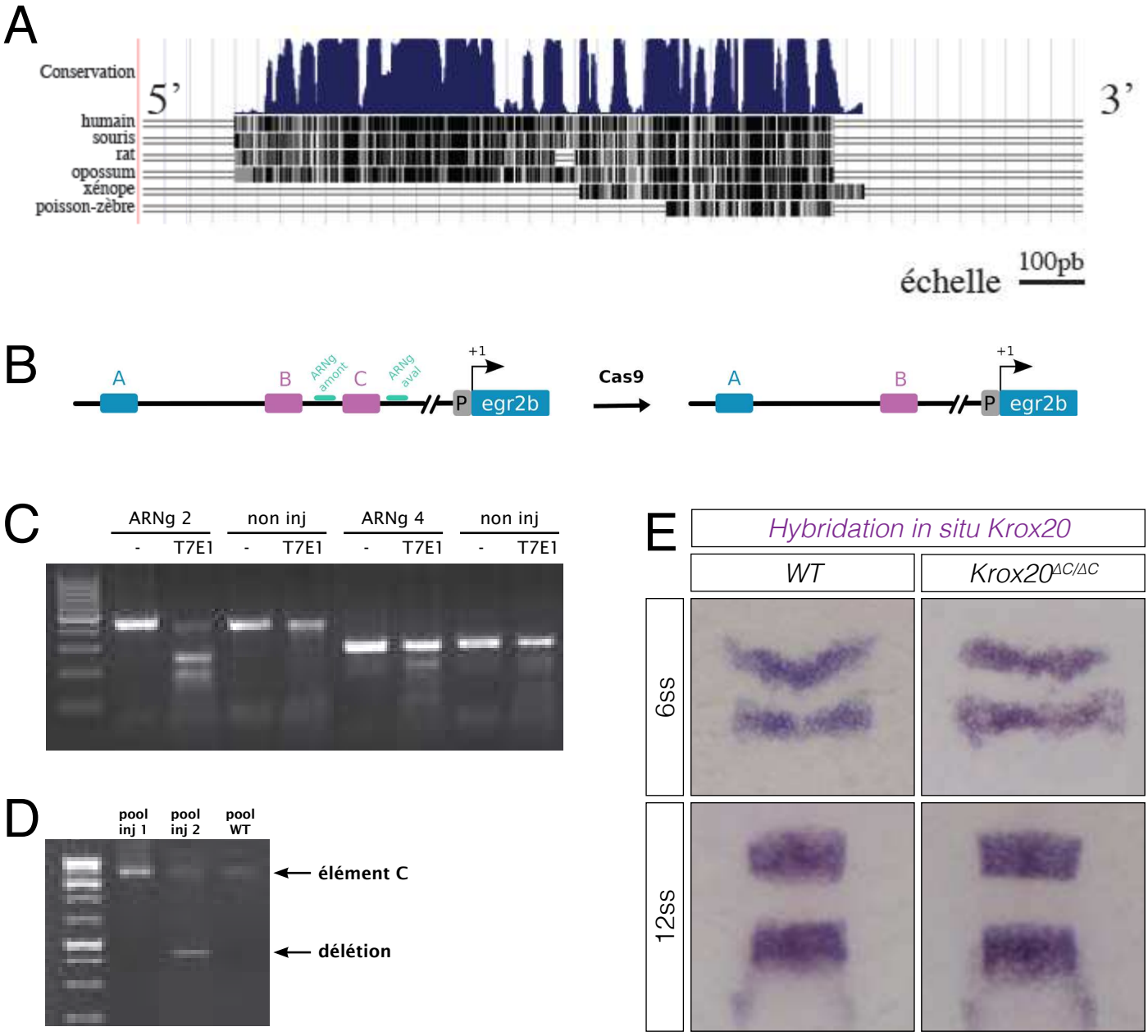


Figure 17 : Délétion de l'élément C chez le poisson-zèbre

A. Conservation de séquence de l'élément C entre le génome de référence du poulet puis l'humain, la souris, le rat, l'opossum, le xénope et le poisson-zèbre tirée des alignements du génome browser UCSC.

B. Stratégie de délétion de l'élément C par la technique de CRISPR/Cas9 en ciblant les séquences amont et aval de l'élément C par des ARNg.

C. Amplification par PCR d'ADN extrait de 10 embryons non injectés ou injectés avec l'ARNg 2 ou 4 et la Cas9 suivi du test de l'endonucléase T7. Un contrôle sans endonucléase a aussi été réalisé (tiret).

D. Amplification par PCR d'ADN extrait de 10 embryons injectés au non avec les 2 ARNg et la Cas9. La délétion de l'élément C engendre une bande plus petite que le sauvage.

E. Hybridation *in situ Krox20* sur des embryons de poisson sauvages ou portant la délétion homozygote de l'élément C (*Krox20*$^{\Delta C/\Delta C}$) à deux stades de développement différents.

non inj : non injecté, ss : stade somitique, T7E1 : endonucléase T7.

D'après le manuscrit de thèse de Johan Le Men (A) et les données de Patrick Torbey (E).

Les embryons de poisson-zèbre sont injectés au stade 1 cellule avec 2nL des différents ARNs : les deux ARNg à une concentration finale de 15ng/μL chacun et l'ARNm Cas9 à 150ng/μL.

2.2.2 Criblage des mutations par le test de l'endonucléase T7E1 et PCR

Un échantillonnage d'embryons à 24hpf est digéré en protéinase K pour extraire l'ADN. Une PCR en utilisant des amorces flanquant le site de reconnaissance de l'ARNg est réalisée avec les amorces suivantes : 5'-TGT GGC TCT TTT TAG ATT CTC GG-3' et 5'-ACC AGC ATT ATG GCT GTG TCA-3' (ARNg 1 et 2), 5'-AGC GGA GCG GGT AAT AAA GG-3' et 5'-CAG ACA GAC AAC AGA CAC TGG T-3' (ARNg 3), 5'-ACC AGT GTC TGT TGT CTG TCT-3' et 5'-AGC TAA ACT TAG ACC ACG TCC AT-3' (ARNg 4). Les produits de PCR sont ensuite purifiés sur colonnes QIAquick (QIAGEN). Une moitié de l'ADN purifié est alors dénaturé dans le tampon de l'enzyme (1X) suivant le programme : 5 min à 95 °C, de 95°C à 85°C à –2 °C/s, de 85°C à 25°C à –0.1 °C/s, puis à 4 °C (Reyon et al. 2012). L'ADN réhybridé est digéré par l'endonuclease T7E1 (New England Biolabs) à 37°C pendant 30min avant d'être déposé sur gel d'agarose 2% avec l'autre moitié du produit de PCR non digéré.
Le criblage des délétions est fait par PCR en utilisant les amorces de chaque côté de l'élément, ici : 5'-TGT GGC TCT TTT TAG ATT CTC GG-3' (amorce sens pour ARNg 1 et 2) et 5'-AGC TAA ACT TAG ACC ACG TCC AT-3' (amorce antisens pour ARNg 4).

2.3 Résultats

J'ai préparé deux ARNg ciblant en amont de l'élément C et deux autres en aval (fig. 17B) pour réaliser la délétion de cet élément *cis*-régulateur du gène *krox20* appelé *egr2b* chez le poisson-zèbre. La capacité de coupure de la protéine Cas9, après injection avec les différents ARNg, a été contrôlée par le test de l'endonuclease T7. L'ADN génomique à proximité du site putatif de ciblage de l'ARNg est amplifié par PCR, puis le produit est déshybridé et réhybridé afin d'obtenir des hétéroduplexes. Ces derniers comportent des mésappariements reconnus par l'endonuclease T7 qui coupe l'ADN simple brin. Ces tests ont permis de déterminer que l'ARNg 2 et l'ARNg 4 sont efficaces pour couper en amont et en aval de l'élément C respectivement (fig. 17C). La co-injection de ces deux ARNg et de l'ARNm de la Cas9 permet d'obtenir la délétion mosaïque de l'élément C dans environ 15% des poissons de première génération ou F0 (fig. 17D). Les F0 sont ensuite croisés avec des poissons sauvages pour obtenir des hétérozygotes pour la délétion de l'élément C en F1. La transmission de la délétion en F1 est plus rare et nous avons obtenu deux fondateurs sur plus

d'une trentaine de poissons. Le croisement des F1 permet l'obtention d'embryons homozygotes pour la délétion et le phénotype sur l'expression de *krox20* est analysé par hybridation *in situ* contre *krox20*. La délétion de l'élément n'engendre pas de modifications de l'expression du gène *krox20* que ce soit pendant l'initiation autour de 6 somites ou lors de l'autorégulation à 12 somites (fig. 17E). L'élément C de poisson n'est donc pas le seul élément *cis*-régulateur actif dans r3 pendant l'initiation du gène *Krox20* comme nous l'avons montré chez la souris (cf Chapitre 2). De plus il ne semble pas jouer de rôle pendant l'autorégulation contrairement à ce qui est observé chez la souris. Cette absence de fonction de C pendant l'autorégulation de *krox20* chez le poisson ne m'a donc pas permis d'utiliser ce modèle pour décrypter les mécanismes mis en œuvre dans la coopération entre élément A et C uniquement observée dans l'embryon de souris.

2.4 Conclusions et perspectives

L'absence de phénotype pour la délétion de l'élément C chez le poisson pourrait s'expliquer par une activité différente de l'élément C poisson par rapport à l'élément C poulet ou souris notamment à cause de la partie non conservé de l'élément. Cependant, les expériences de transgénèse réalisées chez le poisson-zèbre ne permettent pas de mettre en évidence une différence d'activité. Les deux éléments n'ont pas gardé la même fonction *in vivo* au locus *Krox20* malgré une activité similaire par transgénèse. Ceci pourrait s'expliquer par la distance entre le promoteur et l'élément A qui est de 217 kb chez la souris et seulement de 74 kb chez le poisson. Il est possible d'imaginer que cette différence soit à l'origine d'une conformation différente du locus qui ferait de l'élément C une plateforme pour le rapprochement de l'élément A du promoteur pendant la phase d'autorégulation chez la souris. Une distance plus faible ne nécessiterait pas l'aide de l'élément C pour l'autorégulation chez le poisson. Il y aurait ainsi une utilisation différente d'éléments *cis*-régulateurs conservés en séquence pour l'expression d'un gène de développement au cours de l'évolution en réponse à la complexification des génomes. Une autre hypothèse serait que d'autres redondances d'éléments *cis*-régulateurs chez le poisson compenseraient aussi le phénotype d'autorégulation observé chez la souris. Ce travail a été repris par un étudiant en thèse qui mène une analyse systématique d'identification des enhancers du gène *krox20* par transgénèse et des délétions, parfois combinées, de ces éléments *cis*-régulateurs dans le génome du poisson afin de proposer un modèle de régulation d'un gène de développement le plus complet possible.

DISCUSSION

Le gène *Krox20* est un gène clef dans le développement du rhombencéphale. L'expression de ce gène suit deux phases : une phase de démarrage dirigée par des éléments initiateurs fonctionnant indépendamment de la protéine KROX20 et une phase d'autorégulation qui requiert la présence de la protéine. Pour analyser la contribution des éléments *cis*-régulateurs identifiés dans le rhombencéphale dans la régulation du gène *Krox20*, j'ai effectué une analyse fonctionnelle de ces éléments. Dans un premier temps, la délétion de l'élément A a permis de montrer son rôle direct dans l'autorégulation du gène *Krox20*. Dans un deuxième temps, la délétion de l'élément C a révélé son rôle non pas au cours de l'initiation comme on l'attendait, mais pendant l'autorégulation en coopération en *cis* avec l'élément A grâce à des approches de génétiques et génomiques. Dans un dernier temps, j'ai démontré que l'élément C n'a pas le même rôle chez le poisson-zèbre pour la régulation du gène *Krox20*. Je discuterai dans cette partie des conséquences des délétions de ces deux enhancers chez la souris, puis chez le poisson-zèbre ; délétions qui ont révélé que d'autres éléments restaient à identifier pour expliquer la régulation du gène *Krox20* dans le rhombencéphale. Je détaillerai les candidats récemment identifiés notamment grâce aux données d'accessibilité de la chromatine et de conservation de séquence. Je finirai par discuter des limitations de notre système d'étude et des technologies qui permettront prochainement d'analyser finement la régulation des gènes.

1. Délétion d'un élément autorégulateur chez la souris

Afin d'étudier le rôle de l'autorégulation du gène *Krox20 in vivo* dans la segmentation du rhombencéphale, une délétion de l'élément A a été réalisée par recombinaison homologue dans le génome murin. Cette étude a permis de démontrer de façon formelle que l'autorégulation du gène *Krox20 via* l'élément A est absolument requise pour son expression totale et pour le patron de développement du rhombencéphale. Un sauvetage fonctionnel a été réalisé grâce au système original A*/Krox20* qui a permis, à l'aide d'une boucle d'autorégulation indirecte, de montrer que le phénotype observé chez l'homozygote *Krox20*$^{\Delta A/\Delta A}$ est exclusivement dû à l'absence d'autorégulation du gène *Krox20*. Ceci suggère que le niveau d'initiation n'est pas affecté dans le mutant de l'élément A, que seule la boucle d'autorégulation est responsable du phénotype et que, par là même, la boucle est suffisante pour le maintien de l'expression de *Krox20*. Ce sauvetage a été visualisé par l'expression du gène *EphA4*, une cible de *Krox20*, car la sonde *Krox20* utilisée dans cette étude ne permet pas de discriminer les transcrits *Krox20* et *Krox20**.

Dans le mutant de l'autorégulation, la dose initiale de Krox20 est suffisante pour activer au moins un de ses gènes cibles *EphA4*, et probablement aussi les gènes *Hox*. Cependant, l'expression d'*EphA4* n'est pas maintenue dans une partie des cellules lorsque l'élément A est délété. Nous avons montré que ces cellules sont re-spécifiées en identité de rhombomères pairs et qu'une vague d'apoptose précoce plus importante dans r3 que dans r5 réduit probablement le nombre de cellules. Cette dernière hypothèse est cohérente avec la perturbation de la différenciation neuronale et le rapprochement des nerfs crâniens observés chez les embryons $Krox20^{\Delta A/\Delta A}$ dans r3 alors que la région r5 est peu affectée. Chez le mutant nul, la région r3 est affectée de façon similaire au mutant de l'autorégulation mais la re-spécification précède la vague d'apoptose (Voiculescu et al. 2001). Contrairement au mutant nul, les cellules du mutant de l'autorégulation reçoivent une petite dose de Krox20 qui permet l'activation transitoire de gènes cibles et retarderait alors la re-spécification. Il est aussi possible qu'un défaut léger de prolifération affecte le nombre de cellules r3 et r5 dans ce mutant mais le nombre faible de cellules du rhombencéphale à ce stade rend difficile la validation de cette hypothèse.

La moitié des souris homozygotes $Krox20^{\Delta A/\Delta A}$ meurent avant 3 semaines, alors que les autres survivent sans phénotype apparent et sont fertiles (observation sur un échantillon de 100 souris issues de croisement entre hétérozygotes), contrairement au mutant $Krox20^{LacZ/LacZ}$ où 90 % des souriceaux meurent 24h après la naissance et le reliquat avant 3 semaines (Schneider-Maunoury et al. 1993). Plusieurs hypothèses ont été avancées pour expliquer les deux phases de mortalité du mutant $Krox20^{LacZ/LacZ}$. Les défauts dans l'organisation des nerfs crâniens sont à l'origine de déficit du mouvement respiratoire ou de la mastication en lien avec les anomalies du rhombencéphale et pourraient être responsable de la mortalité observée à la naissance. En revanche, la seconde phase serait causée par le rôle majeur de *Krox20* dans la myélinisation des cellules de Schwann (Topilko et al. 1994) entraînant une paralysie totale des animaux. De façon intéressante, le mutant $Krox20^{\Delta A/\Delta A}$ constitue le premier mutant affectant spécifiquement la fonction de *Krox20* dans le rhombencéphale et pas les autres organes où ce gène est exprimé (cœur, cellules de Schwann, os, etc…). La létalité observée chez ce mutant correspondrait à la première phase chez le mutant $Krox20^{LacZ/LacZ}$ avec une pénétrance moins élevée que dans le mutant nul ce qui est sans doute le reflet de la variabilité phénotypique observée au niveau de l'expression d'*EphA4* et de la différenciation neuronale.

2. La boucle entre l'élément A et le promoteur apporterait de la robustesse

Nous avons montré que l'élément A interagit de façon constitutive avec le promoteur au cours de la segmentation du rhombencéphale. Ce type de communication à distance s'inscrit dans un modèle permissif qui s'oppose au modèle instructif. Dans la situation permissive, la conformation tridimensionnelle existante du locus permet l'activation de la transcription dans n'importe quel type cellulaire par le recrutement des facteurs de transcription tissu-spécifiques. Dans le modèle instructif, une conformation chromatinienne *de novo* est établie au locus par des facteurs de transcription spécifiques pour l'activation transcriptionnelle. Les structures préformées offrent plusieurs bénéfices de régulation. Dans un premier temps, elles empêchent l'activation de gènes voisins par l'enhancer. Dans un second temps, elles assurent l'activation robuste et rapide des gènes par le simple recrutement des facteurs (de Laat & Duboule 2013). Le niveau de *Krox20* doit atteindre un certain seuil pour enclencher rapidement la boucle d'autorégulation par la fixation synergique des protéines KROX20 sur l'élément A (Bouchoucha et al. 2013). Les interactions constitutives entre l'élément A et le promoteur faciliteraient l'intégration rapide et robuste de cette dose de protéines KROX20 pour la mise en place de la boucle d'autorégulation. Cependant, cette situation permissive de l'élément A peut être dangereuse, une faible quantité de protéines dans la cellule pourrait activer la boucle d'autorégulation dans un tissu où l'amplification de l'expression de *Krox20* n'est pas requise. Un premier niveau de contrôle est assuré par le seuil de protéines à atteindre dans la cellule (Bouchoucha et al. 2013). Un second niveau de contrôle pourrait être apporté en agissant en *cis* sur l'activité de l'élément A pendant l'autorégulation. L'élément C pourrait remplir ces deux conditions pour assurer l'autorégulation de *Krox20* au moins dans r3 en participant d'une part, au niveau d'initiation de l'expression du gène *Krox20* et d'autre part, en agissant en *cis* sur l'activité de l'élément A tout au long de l'autorégulation.

Les interactions entre le promoteur et l'élément A sont corrélées avec la fixation du facteur CTCF au locus *Krox20* (fig. 2A). Le facteur CTCF fixe une séquence consensus non-palindromique (Ong & Corces 2014, fig.18A) dont l'orientation semblerait importante pour les interactions enhancer-promoteur (Guo et al. 2015; Wit et al. 2015). En comparant les données de ChIP-seq (ENCODE) à plusieurs stades et dans plusieurs tissus, j'ai observé 5 sites de fixation CTCF au locus *Krox20*, nommés BS 0 à BS 4 (fig. 18B). Leurs séquences sont très proches de la matrice consensus et m'ont permis de définir leur orientation (fig. 18B). Les sites BS 0 et BS 1 sont divergents l'un par rapport à l'autre et sont de bons candidats pour la frontière du TAD séparant le locus *Krox20* de *Nrbf2*. Le site BS 2 n'est pas

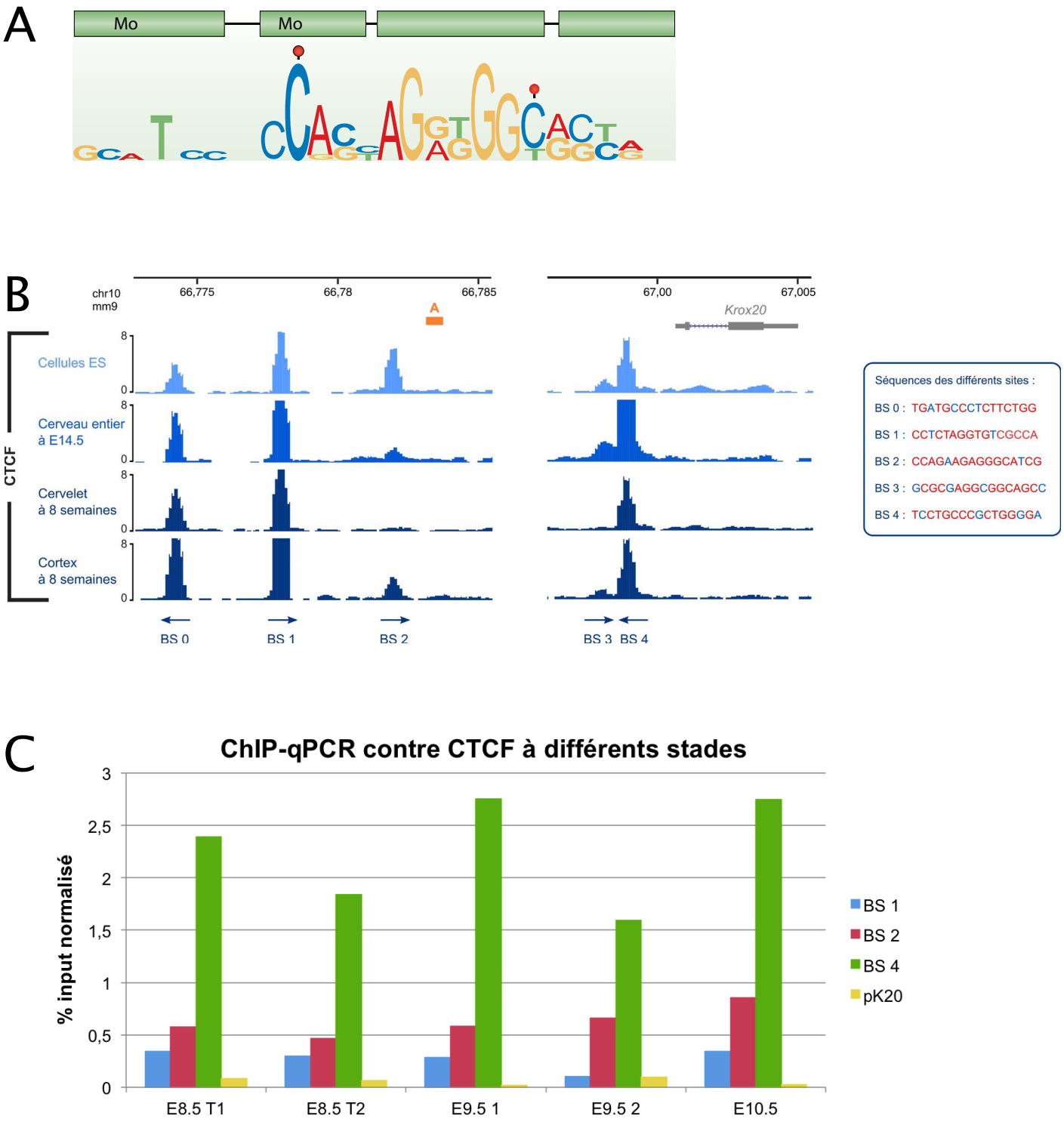


Figure 18 : Sites et fixation du facteur CTCF au locus *Krox20*

A. Séquence consensus de la fixation du facteur CTCF.

B. Détail de la fixation du facteur CTCF à proximité de l'élément A et du promoteur de *Krox20* à partir des données de ChIP-seq CTCF (ENCODE). Du haut vers le bas : cellules ES, cerveau entier à E14.5, cervelet à huit semaines, cortex à huit semaines. Les séquences de chaque site de fixation sont indiquées à droite (en rouge les nucléotides en accord avec la matrice et les autres en bleu) et leur orientation est représentée par les flèches en bas de la figure. Les coordonnées génomiques, l'élément A (orange) et le promoteur (gris) sont représentés en haut de la figure.

C. ChIP-qPCR contre le facteur CTCF sur des embryons de souris disséqués à 8,5 jpc (uniquement la tête), entiers à 9,5 jpc et 10,5 jpc. Deux expériences ont été réalisées à 8,5 jpc et 9,5 jpc (1 & 2).

D'après Ong and Corces, 2014 (A).

constitutivement fixé dans tous les tissus et est le plus proche de l'élément A avec une orientation sens. Du côté du promoteur, les sites BS 3 et BS 4 sont convergents avec BS 3 dans une orientation sens et BS 4 anti-sens et plus proche du promoteur. Etant donné leurs orientations relatives, les sites BS 1 et BS 2 sont avec BS 4 les meilleurs candidats pour faciliter la bouche constitutive entre l'élément A et le promoteur. J'ai réalisé des expériences de ChIP-qPCR contre CTCF sur des embryons de souris à différents stades de développement ciblant ces trois sites et le promoteur comme région où l'on n'attend pas de fixation de CTCF (fig. 18C). Les données sont normalisées par rapport à la fixation de CTCF sur deux autres sites à la frontière du TAD voisin du gène *Krox20* car la fixation du facteur est moins susceptible de varier. Le promoteur n'est pas fixé par CTCF et constitue un bon témoin négatif. La fixation du site BS 2 par CTCF est plus importante que pour le site BS 1 pour tous les échantillons suggérant que le site BS 2 qui n'est pas fixé par CTCF dans tous les tissus (fig. 18B) joue un rôle dans la conformation tridimensionnelle du locus *Krox20* au stade embryonnaire. Le site BS 4 est le plus fortement fixé par CTCF dans les échantillons, pointant l'importance de ce site dans l'interaction du promoteur *Krox20* avec les régions en amont. Les tendances sont les mêmes pour la tête à 8,5 jpc et les embryons entiers à 9,5 jpc et 10,5 jpc et en accord avec une boucle entre les sites convergents BS 2 et BS 4 (fig. 18C). La délétion de l'un, de l'autre ou des deux sites permettrait de confirmer l'importance de ces sites dans cette conformation constitutive.

3. Nouvelles questions sur le rôle de l'élément C dans l'expression en deux phases du gène *Krox20*

Par transgénèse chez la souris, il a été montré que l'élément C présente une activité précoce dans r3 puis r5 et que cette activité se renforce dans ces deux rhombomères et suit donc le patron d'expression du gène *Krox20* (Chomette et al. 2006, fig. 12I). Cette activité est indépendante de la présence de la protéine KROX20 mais dépendante des facteurs Hox/Pbx et Meis, faisant de l'élément C un bon candidat pour l'initiation de l'expression de *Krox20* seul dans r3 et avec l'élément B dans r5 (Chomette et al. 2006; Wassef et al. 2008). De plus il a été montré que l'activité de l'élément C est sensible à la signalisation FGF, par l'intermédiaire de la fixation du facteur SP5 (Labalette et al., 2015), signalisation indispensable à l'initiation de *Krox20* dans r3 et r5 (Walshe et al. 2002; Aragon & Pujades 2009). Cependant, nous avons montré que la délétion de l'élément C dans le génome murin n'impacte pas l'initiation de l'expression du gène *Krox20* mais son maintien dans r3. Ceci indique qu'un autre enhancer conduit l'initiation dans r3 et que l'élément C coopère avec l'élément A. Cette coopération se

fait en *cis* comme le prouve la perte d'expression au cours de l'autorégulation chez l'hétérozygote composite *Krox20*$^{\Delta A/\Delta C}$. Cette expérience suggère aussi que le niveau d'initiation de *Krox20* atteint dans le mutant *Krox20*$^{\Delta C/\Delta C}$ n'est pas un facteur limitant pour enclencher la boucle d'autorégulation. En effet il a été montré que le niveau d'initiation de l'expression de *Krox20* chez l'homozygote *Krox20*$^{\Delta A/\Delta A}$ est normal (Bouchoucha et al. 2013) et qu'un seul allèle contenant le gène *Krox20* suffit à récapituler le patron d'expression du gène *Krox20*. L'hétérozygote composite *Krox20*$^{\Delta A/\Delta C}$ doit donc lui aussi présenter une dose d'initiation normale de *Krox20* mais ne permettant pas l'amplification et le maintien de l'expression car les éléments A et C doivent être présents sur le même chromosome pour que l'expression soit correcte r3. J'ai voulu démontrer formellement que ce niveau d'expression précoce de *Krox20* était identique chez le mutant par rapport au sauvage en réalisant des expériences de transcription inverse suivies d'une PCR quantitative (RT-qPCR) sur des embryons à 4 et 5 somites, pendant la phase d'initiation (fig. 19A). À 4 somites, le niveau d'expression de *Krox20* dans les embryons mutants *Krox20*$^{\Delta C/\Delta C}$ est de plus de 50% du niveau observé dans les embryons sauvages, suggérant que le niveau initial dans le mutant *Krox20*$^{\Delta C/\Delta C}$ est suffisant pour enclencher la boucle d'autorégulation puisque supérieur à ce qui est observé chez un hétérozygote *Krox20*$^{+/-}$. À 5 somites, la variabilité entre les deux embryons sauvages, d'une part, ou entre les deux embryons mutants, d'autre part, est plus importante rendant difficile la comparaison. C'est porbablement du fait que ce stade coïncide avec l'enclenchement de la boucle d'autorégulation comme le suggère le début de la perte d'expression dans r3 chez le mutant *Krox20*$^{\Delta A/\Delta A}$ (Bouchoucha et al. 2013). Cette variabilité peut être le reflet, d'une part, de la variabilité interne de l'expression stochastique du gène *Krox20* pendant l'initiation et, d'autre part, que l'identification du stade par le nombre de somites au cours du développement de l'embryon est légèrement imprécise. En tout état de cause, il faudrait réaliser les RT-qPCR sur un beaucoup plus grand nombre d'embryons pour pouvoir conclure de façon formelle sur le niveau initial de Krox20 dans les mutants *Krox20*$^{\Delta C/\Delta C}$ par rapport au sauvage.

Deux hypothèses non exclusives peuvent expliquer la coopération en *cis* entre les éléments A et C. La première est l'interaction physique entre les deux éléments que nous avons tenté d'observer par la technique de 4C-seq. Il n'a pas été possible de déceler d'interactions entre l'élément C et l'élément A ou le promoteur. Cette technique nécessitant un très grand nombre de cellules, le tissu que nous avons utilisé ne comprend au plus qu'un dixième de cellules exprimant *Krox20*, ce qui peut masquer les interactions possibles si elles ont lieu uniquement dans les cellules exprimant *Krox20*. Il est aussi plausible qu'il n'y ait pas d'interactions physiques à l'origine de cette coopération mais que celle-ci résulte de la

transmission d'une information en *cis*. La deuxième hypothèse permettant d'expliquer une telle coopération en *cis* est la transcription de l'enhancer lui-même en eRNAs. Les différents rôles de la transcription des enhancers et des eRNAs dans la régulation des gènes sont encore peu décrits mais les quelques exemples à ce jour pointent vers une action en *cis* notamment dans les boucles enhancers-promoteurs (cf. Introduction). Il est cependant très difficile d'investiguer le produit de la transcription des enhancers et leur rôle dans notre système embryonnaire car les eRNAs sont transcrits à bas bruit, très peu stables et rapidement dégradés par les exosomes.

Les souris homozygotes *Krox20*$^{\Delta C/\Delta C}$ sont fertiles et aucune mortalité n'a été observée, contrairement aux souris homozygotes *Krox20*$^{\Delta A/\Delta A}$ dont la moitié meurent avant 3 semaines. De façon similaire au mutant de délétion de l'élément A, j'ai analysé l'impact de la délétion de l'élément C sur l'expression du gène cible *EphA4* et sur l'organisation des nerfs crâniens (fig. 19 B et 19C). À 9,0 jpc l'expression du gène *EphA4* par hybridation *in situ* est détectée dans un territoire r3 réduit à peu près de la moitié alors que l'expression dans r5 est normale par rapport au sauvage (fig. 19B). La perte d'expression de *Krox20* dans r3 pendant la phase d'autorégulation chez l'homozygote *Krox20*$^{\Delta C/\Delta C}$ entraîne donc une diminution du nombre de cellules ayant activé *EphA4* dans ce rhombomère. Il est important de noter que le nombre de cellules exprimant *EphA4* n'est pas diminué de plus de la moitié chez le mutant *Krox20*$^{\Delta C/\Delta C}$, contrairement au mutant *Krox20*$^{\Delta A/\Delta A}$ (fig. 13A) à 9,0 jpc dont le phénotype est plus sévère. Il semblerait donc que, là encore, la dose initiale de *Krox20* soit suffisante pour engager un nombre important de cellules en identité du territoire r3. J'ai ensuite regardé l'organisation des nerfs crâniens par immunofluorescence avec le marqueur de neurofilament 2H3 à 10,5 jpc. Comme pour le mutant *Krox20*$^{\Delta A/\Delta A}$, dans les embryons *Krox20*$^{\Delta C/\Delta C}$ les racines du nerf V (trijumeau) et des nerfs VII (facial) et VIII (auditif) ne sont pas bien séparées par rapport au contrôle *Krox20*$^{+/\Delta C}$ (fig. 19C, flèches). Ce phénotype est cohérent avec la perte d'expression du gène *Krox20* chez les embryons *Krox20*$^{\Delta C/\Delta C}$ dans r3. Les conséquences de la délétion de l'élément C sur le développement général du rhombencéphale semblent moins importantes que pour la délétion de l'élément A en accord avec une absence de mortalité évidente des souriceaux.

4. Les éléments *cis*-régulateurs du gène *Krox20* dans le rhombencéphale chez la souris

L'absence de phénotype au cours de l'initiation du gène *Krox20* dans les mutants *Krox20*$^{\Delta C/\Delta C}$ suggère qu'au moins un élément *cis*-régulateur n'a pas encore été identifié pour

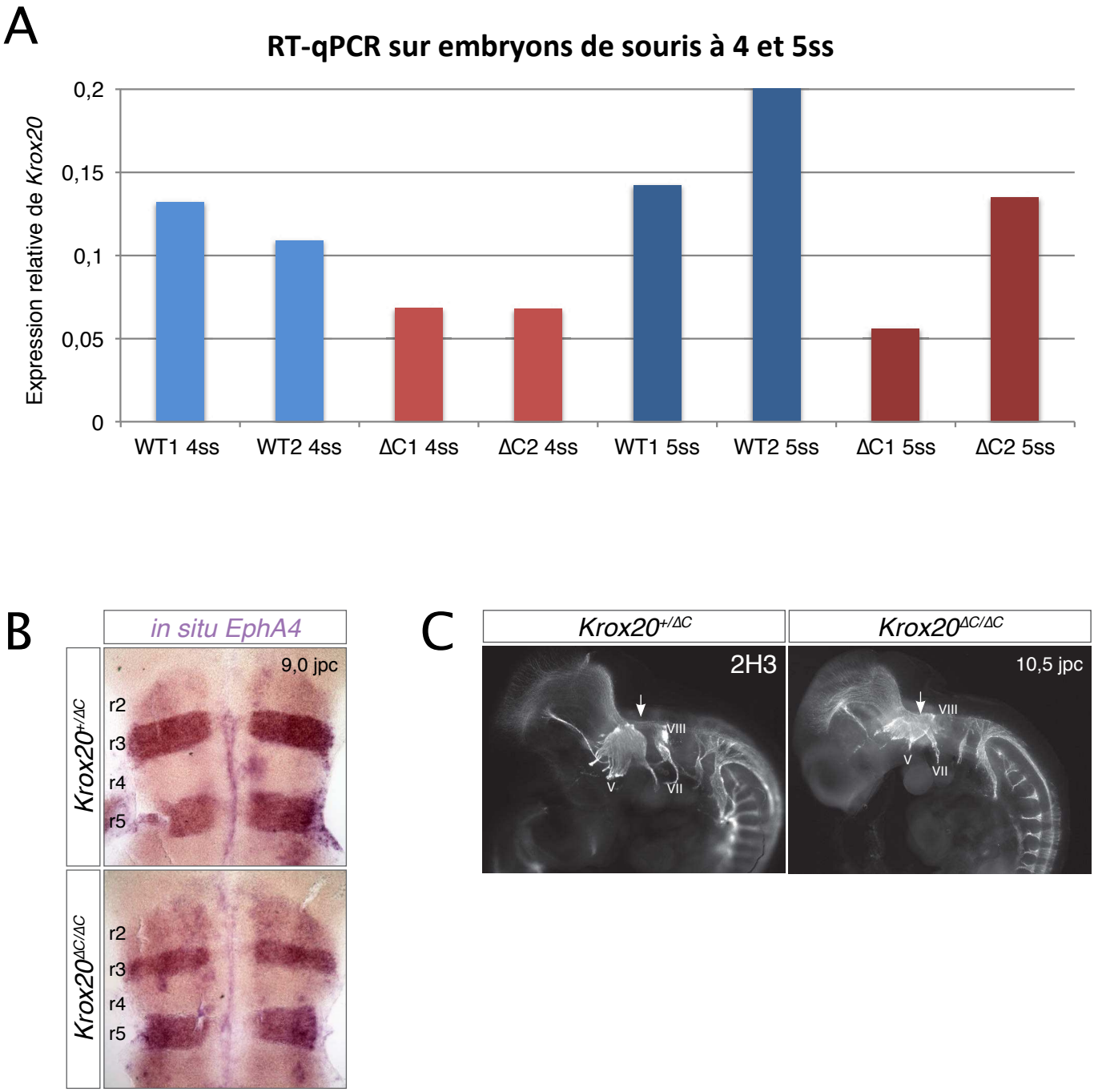


Figure 19 : Les conséquences de la délétion de l'élément C

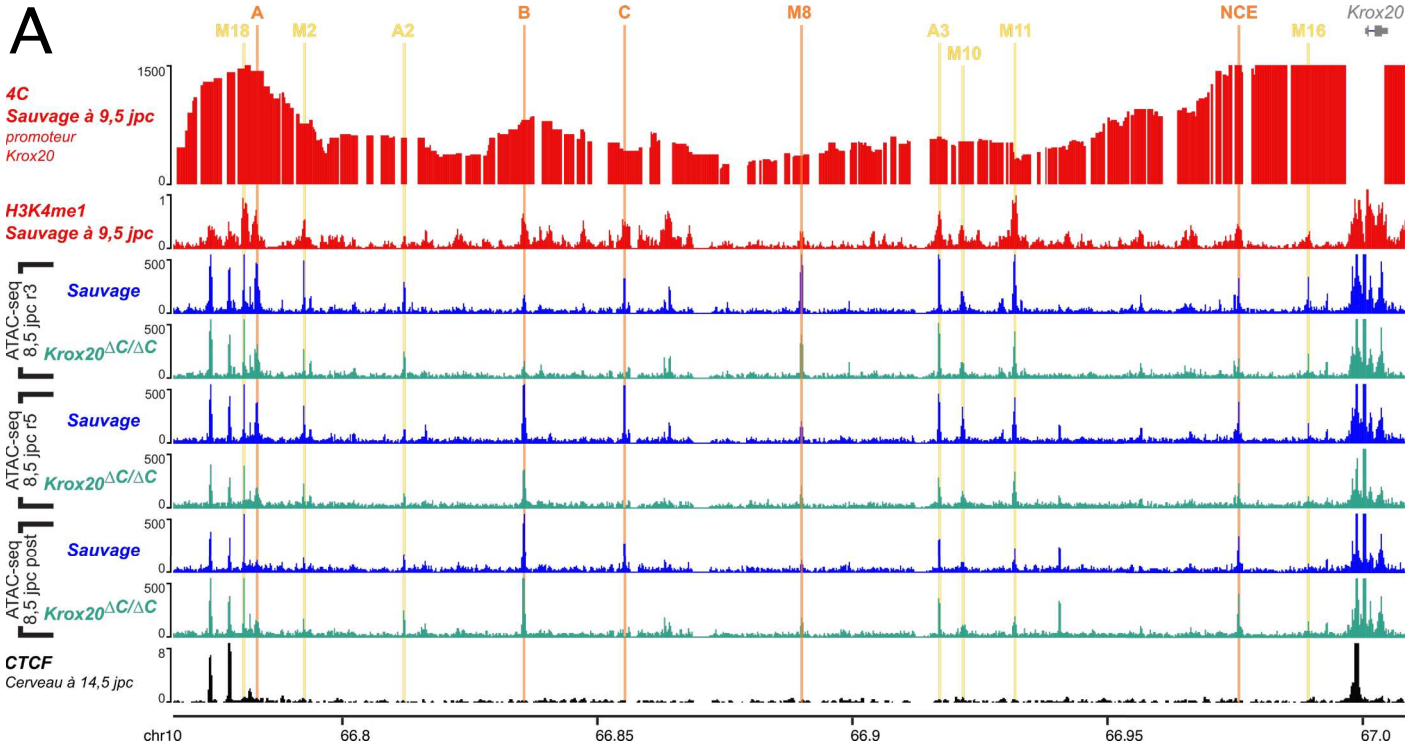
A. RT-qPCR sur embryons de souris à 4ss et 5ss sauvages et *Krox20*$^{\Delta C/\Delta C}$. L'expression de *Krox20* est relative à celle de trois gènes de ménage (*Gapdh*, *Cyclophiline B* et *β-actine*). Deux embryons ont été utilisés par condition (WT1, WT2 et ΔC1, ΔC2).

B. Hybridation *in situ* contre *EphA4* sur des embryons de souris contrôles *Krox20*$^{+/\Delta C}$ et mutants *Krox20*$^{\Delta C/\Delta C}$ à 9,0 jpc.

C. Immunofluorescence sur embryons entiers à 10,5 jpc contrôles *Krox20*$^{+/\Delta C}$ et mutants *Krox20*$^{\Delta C/\Delta C}$ marqués par l'anticorps anti-neurofilaments 2H3. Les flèches indiquent la zone de rapprochement du nerf V (trijumeau) et des nerfs VII (facial) et VIII (auditif).

expliquer l'activation de l'expression du gène *Krox20* dans le r3. En combinant les résultats de 4C-seq, ChIP-seq et ATAC-seq, nous pouvons, d'une part, récapituler les éléments *cis*-régulateurs connus et, d'autre part, identifier les éléments potentiels (fig. 20).
De façon attendue, les éléments A, B et C présentent un enrichissement en H3K4me1 (fig. 20A). Les éléments A et B sont en contact l'un avec l'autre et l'élément A avec le promoteur de façon réciproque à 8,5 jpc et 9,5 jpc (fig. 20A). De plus, les éléments A, B et C présentent une accessibilité qui est corrélée avec leurs domaines d'activité observés par transgénèse chez la souris (Chomette et al. 2006) ainsi que la présence des facteurs qui les fixent : Krox20 dans r3 et r5 pour l'élément A (Chomette et al. 2006; Bouchoucha et al. 2013), Hnf1B dans r5 à r8 et MafB de r5 et r6 pour l'élément B (Chomette et al. 2006; Labalette et al. 2011), les facteurs Hox/Pbx/MeiS dans tout le rhombencéphale et postérieurement pour l'élément C (Wassef et al. 2008) (fig. 20A et 20B). Le gène *Krox20* a une expression pléiotropique, notamment dans les cellules de crête neurale issues de r5 (Schneider-Maunoury et al. 1993) pour lesquelles un élément régulateur appelé NCE a été identifié par transgénèse (Ghislain et al. 2003). Cet élément présente une accessibilité dans les trois régions étudiées. *Krox20* est aussi exprimé dans les cellules des capsules frontières (BC) qui sont à l'interface entre le système nerveux central et périphérique (Schneider-Maunoury et al. 1993; Topilko et al. 1994) et pour lesquelles a été identifié un autre élément enhancer appelé M8 (Pascale Gilardi-Hebenstreit, communication personnelle) qui dirige l'expression de *Krox20* dans les BC et montrent une accessibilité décroissante de r3 à r5 et postérieurement, ce qui pourrait être est en accord avec la maturation progressive des BC selon l'axe antéro-postérieur (fig. 20). De plus, le gène *Krox20* est exprimé dans les cellules de Schwann myélinisantes (Topilko et al. 1994). Un élément appelé MSE, situé en aval du promoteur, dirige l'expression du gène *Krox20* dans ces cellules, ne présente aucune accessibilité dans les trois régions observées, en accord avec une activité de cet élément dans ces cellules du système nerveux périphérique et à un stade plus tardif de développement (Ghislain et al. 2002, données non présentées).

Plusieurs éléments régulateurs potentiels du gène *Krox20* peuvent été identifiés par les données expérimentales de marques de chromatine et d'accessibilité (fig. 20A). Certains éléments n'ont pas d'activité par électroporation chez l'embryon de poulet comme les éléments M18 et M10 (fig. 20B). L'activité des trois éléments, A2, A3 et M16, sera prochainement testée car ils présentent une accessibilité intéressante dans les trois régions étudiées (fig. 20A). Cependant, l'élément A2 ne présente pas de marques H3K4me1 ce qui n'en fait pas un très bon candidat, et l'élément M16 est très proche du promoteur. Deux éléments putatifs nous ont particulièrement intéressés par leur profil d'accessibilité et grâce à leur conservation de séquence chez les mammifères et jusqu'au Lépisosté tacheté (Spotted



B

Échantillons	M18	A	M2	A2	B	C	M8	A3	M10	M11	NCE	M16
H3K4me1 E9.5	+	+	+	-	+	+	+/-	+	+	+	+	+/-
WT r3 E8.5	+	+	+	+	+/-	+	+	+	+	+	+	+
ΔC r3 E8.5	+	+/-	+	+	+/-	-	+	+	+	+	+	+
WT r5 E8.5	+	+	+	+/-	+	+	+/-	+	+	+	+	+/-
ΔC r5 E8.5	+	+/-	+	+/-	+	-	+/-	+	+	+	+	+/-
WT post E8.5	+	-	+/-	+	+	+	+/-	+	+/-	+/-	+	+/-
ΔC post E8.5	+	-	+/-	+	+	-	+/-	+	+/-	+/-	+	+/-
Activité	non	r3/r5	r3/r5	?	r5	r3-r5	BC	?	non	r3/r5/r6	NC	?

+ : pic important
+/- : pic intermédiaire
- : absence de pic

Figure 20 : Les éléments *cis*-régulateurs connus et putatifs du gène *Krox20*

A. Combinaison des résultats du Chapitre 2 pour l'identification de nouveaux éléments. Du haut vers le bas : 4C à 9,5 jpc du point de vue du promoteur, ChIP-seq contre H3K4me1 sur des embryons à 9,5 jpc et ATAC-seq à 8,5 jpc sur une région comprenant r3, une région contenant r5 et une région postérieure dans des embryons sauvages ou mutants *Krox20*$^{\Delta C/\Delta C}$. En haut de la figure, les éléments *cis*-régulateurs connus sont en orange et les éléments potentiels en jaune. Les coordonnées génomiques sont représentées en bas de la figure.

B. Tableau récapitulatif des résultats de ChIP-seq et d'ATAC-seq précédents. L'activité de certains éléments a été testée par transgénèse chez la souris (A, B, C), chez le poisson-zèbre (A, B, C, M11, M2) ou par électroporation chez le poulet (A, B, C, M2, M11, M18, M8, M10, NCE). L'activité des éléments A2, A3 et M16 n'a pas encore été testée.

r3/r5 : activité dans r3 et r5 ; r3-r5 : activité de r3 à r5 ; BC : activité dans les cellules des capsules frontières ; NC : activité dans les cellules des crêtes neurales.

gar) dont je reparlerai dans la partie suivante. Ces éléments ont une activité par électroporation chez le poulet dans r3 et r5 pour l'élément M2 et dans r3, r5-r7 pour M11. L'élément M2 présente principalement une accessibilité dans les régions r3 et r5 de façon similaire à l'élément à l'élément A. Cet élément contient deux sites de fixation forts pour Krox20, qui, quand ils sont mutés abolissent l'activité de l'élément dans r3 et r5 par électroporation (Pascale Gilardi-Hebensttreit, communication personnelle). Ces résultats suggèrent que l'élément M2 est un élément de réponse à la protéine KROX20 dans le rhombencéphale. Il est possible que l'élément M2 soit un autre élément nécessaire, avec l'élément A, à l'autorégulation.

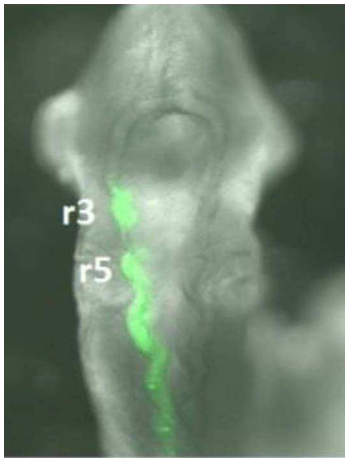
Intéressons-nous maintenant à l'élément M11 qui est le meilleur candidat jusqu'à présent pour être l'élément manquant de l'activation du gène *Krox20* dans r3. Les résultats qui suivent ont été obtenus par Patrick Torbey, étudiant en thèse au laboratoire. Dans un premier temps, une recherche *in silico* a permis d'identifier que la séquence minimale de l'élément M11 de 639 pb contient plusieurs sites de fixation putatifs pour MafB et Meis, deux sites putatifs pour les facteurs Hox/Pbx et trois sites pour Krox20 (fig. 21A). Les facteurs Hox, Pbx et Meis sont connus pour être nécessaires à l'expression de *Krox20* dans r3 et se fixent sur l'élément C (Wassef et al. 2008). L'élément M11 minimal possède une activité dans r3, r5-r7 par électroporation chez le poulet, de façon similaire à l'activité de l'élément C qui présente en plus une activité dans r4 (Chomette et al. 2006, fig. 21B). L'importance fonctionnelle de certains sites de l'élément M11 a été testés par mutations ponctuelles abolissant les sites de fixation soit pour les facteurs Hox/Pbx, soit pour Krox20. La mutation des sites de fixation putatifs des facteurs Hox/Pbx abolit l'activité de M11 dans r5-r7 et diminue l'activité dans r3. Cette activité semble donc être dépendante de la fixation des facteurs Hox/Pbx dans r5, r6 et r7 et dans une moindre mesure dans r3. La mutation des sites Krox20 entraîne une absence totale d'activité dans r3 ou r5. Nous en déduisons que l'élément M11 répond aussi à la protéine KROX20. Ces données suggèrent que cet élément aurait à la fois une activité initiatrice *via* la fixation des facteurs Hox/Pbx et autorégulatrice *via* la fixation de la protéine KROX20.

Par transgénèse chez le poisson-zèbre, l'élément M11 (cloné en amont d'un promoteur minimal dirigeant la GFP) a une activité légèrement différente de celle observée dans l'embryon de poulet. Elle apparaît très précocement dans r3 puis dans r5, persiste jusqu'à 11ss dans ces deux rhombomères, puis se renforce dans r5 mais disparaît dans r3 par hybridation *in situ* contre la GFP (fig. 21C). La comparaison avec l'activité *in vivo* chez le poisson-zèbre de l'élément C nous conforte dans l'idée que l'élément M11 est un bon candidat pour l'initiation de l'expression de *Krox20* dans r3 car son activité précoce est

A

```
            HOX/PBX
CCTGTGATTCAATCCATCATGAACTCCTGTCTTCCAGGTCACTCAAAGAAAGCTTCTATTTACAGGCACTGCCCAAGGCCGGCGGG
                                                        MAFB
CTTCCCGAGAGAGTGCTTTGCTAGTCGGGATCTTTCTGAACCCCTTTAGGCAGCCTCGTGTCCATGGGAGAAGAAATCAGACGTAA
            KROX20              MEIS                                        MAFB
AGAGTCTGGAGTGGGAGGCAGATGGGGACTGTCAACACCATCCTTAAGAAGTCTTCACCAGTTTGGCAGATCAGCGGCTTGAACAT
                                                KROX20
GGCCAGGACTTGCCCCTGAGCCACAGGCATCCTGGTTTTCAAGGAGGAGAGCGAACTGCCAGTCAGAGTCTCTCTTTGAAATATGT
  MEIS                                                                              MAFB
GTGACAGACTGACTGCTGTGTCTCCCTCTCATTTTGACTCCTCTCTAGCAGTGTGTCCCTGTCTCCACCCACCTCAGTTCCGTCAG
                                                                          HOX/PBX
CATCTCCATTTTCTCCTCTCCAATGGTATCTTGACTGTTTTCTGCCAGATCAAAGGTAGAAGCGGCCCACTTGATTAATTGCCGTG
  MAFB                  MEIS                                              KROX20
CAGCTGAATCTTTTGCTTCCTTGACACTCACGGAGTGGAGACAAAGGACTGTTCTTTAGAAGCCTGGAGTTGTGTGGGTTGCTGGT
            MAFB
GGGGAGTTCATAAGCCTGAGAGAACTGAGCCTTCCCA
```

B



C

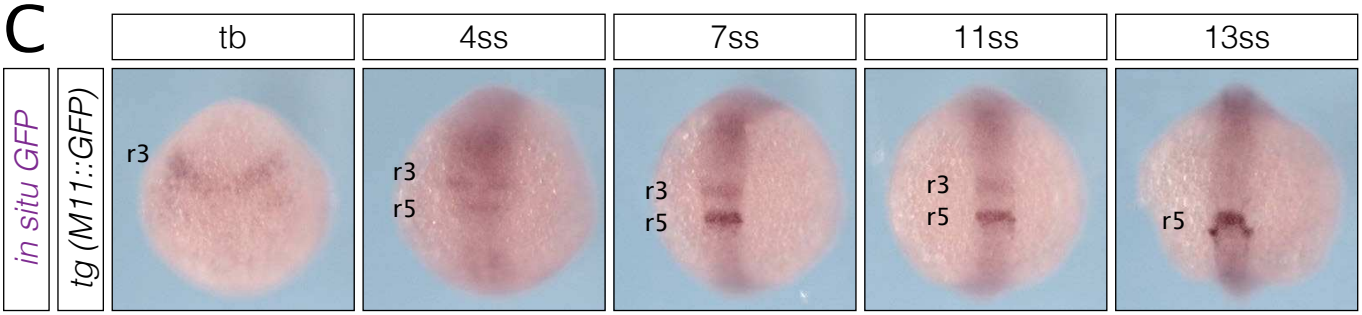


Figure 21 : Séquence et activité de l'élément M11 dans le rhombencéphale

A. Séquence minimale de l'élément M11. Les sites de fixation putatifs pour les facteurs Hox/Pbx (vert), Meis (violet), MafB (gris) et Krox20 (rouge) sont représentés.

B. Électroporation chez l'embryon de poulet de l'élément M11 en amont du gène rapporteur de la GFP montrant une activité dans r3, et de r5 à r7. (Photo : Patrick Torbey)

C. Cinétique d'activité de l'élément M11 par hybridation *in situ* GFP sur des embryons de poisson-zèbre portant le transgène *tg (M11::GFP)* du stade tailbud (tb) à 13 somites (ss). L'activité de l'élément M11 commence précocément dans r3 puis dans r5 jusqu'a 11ss, alors qu'elle se renforce dans r5 et disparait dans r3 à 13ss. (Photos : Pascale Gilardi-Hebenstreit)

similaire au patron d'expression du gène et qu'il pourrait aussi jouer un rôle dans r5. Nous avons entrepris de muter cet élément chez la souris par la technologie CRISPR/Cas9. Nous validons l'inactivation de certains sites de fixation de cet élément dans la lignée transgénique ayant intégré le transgène *tg(M11::GFP)* de façon stable dans le génome en ciblant les sites de fixation avec une combinaison d'ARN guides. Ceci nous permet de visualiser les modifications de l'activité de l'élément grâce à la GFP avant de réaliser l'inactivation de cet élément dans le génome murin. En effet, seule la délétion de l'élément *in vivo* permettra de définir le rôle de l'élément M11 dans la régulation du gène *Krox20* dans le rhombencéphale. Cependant, cet élément peut être redondant avec l'élément C et l'inactivation en fond sauvage risque de ne pas révéler de phénotype. Il faudra peut-être aussi réaliser cette inactivation en fond *Krox20*$^{\Delta C/\Delta C}$ pour apprécier la potentielle redondance entre l'élément M11 et l'élément C.

5. Les éléments *cis*-régulateurs du gène *Krox20* dans le rhombencéphale chez le poisson-zèbre

Le patron d'expression du gène *Krox20* dans le rhombencéphale est conservé entre la souris, le poulet et le poisson-zèbre. Cette conservation a notamment été utilisée dans de nombreuses études au laboratoire pour les avantages techniques que présente chacun de ces systèmes modèles (Chomette et al. 2006; Wassef et al. 2008; Labalette et al. 2011; Bouchoucha et al. 2013; Labalette et al. 2015). De plus, la conservation de séquences d'une partie des éléments *cis*-régulateurs du gène *Krox20* dans le rhombencéphale a aidé leur identification et leur caractérisation. Les alignements de séquences entre plusieurs mammifères, le poulet et le xénope ont révélé une bonne conservation de séquences entre les éléments A, B, C (Chomette et al. 2006) et M11 (Patrick Torbey, communication personnelle). Les éléments B et C (dans une moindre mesure, cf. Chapitre 3) sont aussi conservés chez le poisson-zèbre (Wassef et al. 2008; Labalette et al. 2011). La distance évolutive entre le dernier ancêtre commun de l'homme et le poisson-zèbre est d'environ 450 millions d'années (Kumar & Hedges 1998), rendant difficiles les comparaisons génomiques. Très récemment, le séquençage du génome du spotted gar (Lepisosteus oculatus) a révélé que ce poisson d'eau douce était un chaînon manquant évolutif entre l'homme et le poisson-zèbre (Braasch et al. 2016). La comparaison d'un côté entre la souris et le spotted gar et de l'autre entre le spotted gar et le poisson-zèbre a permis de révéler que les éléments A et M11 étaient aussi en partie conservés entre la souris et le poisson-zèbre (Pascale Gilardi-Hebenstreit, communication personnelle) donnant un argument supplémentaire pour la fonction de l'élément M11 dans la régulation du gène *Krox20*.

Le modèle du poisson-zèbre permet de générer un grand nombre d'embryons rapidement et les injections d'embryons au stade une cellule permettent d'obtenir facilement des lignées transgénique ainsi que des délétions. Nous avons tiré avantage de ce système pour étudier la régulation du gène *Krox20* dans le rhombencéphale. J'ai généré la délétion de l'élément C grâce à la technologie CRISPR/Cas9 et montré que, de façon surprenante, cela n'avait pas de conséquence sur l'expression du gène *Krox20* dans r3 et r5 (cf. Chapitre 3, fig. 17) contrairement à ce qui a été observé chez la souris (cf. Chapitre 2). Ceci n'a donc pas permis d'utiliser le poisson pour étudier la coopération en *cis* entre l'élément A et C comme chez la souris mais ouvre des perspectives intéressantes quand à la différence évolutive de la régulation du gène *Krox20*. C'est à présent le projet de thèse de Patrick Torbey, qui vise à identifier de nouveaux éléments *cis*-régulateurs et de caractériser le rôle de ces éléments par mutagenèse CRISPR/Cas9 chez le poisson-zèbre. L'activité des éléments *cis*-régulateurs conservés du poisson-zèbre a été observée par hybridation *in situ* GFP sur des embryons portant un transgène où l'élément est cloné en amont du gène rapporteur de la GFP. L'élément A présente une activité dans r3 et forte dans r5 pendant l'autorégulation, l'élément B uniquement dans r5, l'élément C est renforcé dans r3 à r5 et activé plus faiblement postérieurement de façon analogue à ce qui est observé chez la souris (fig. 22A). En revanche, l'élément ze2 de poisson-zèbre partiellement conservé avec l'élément M11 de souris, présente une activité forte et spécifique dans r3 et r5 à 10 somites (fig. 22A) avec une cinétique qui suit parfaitement l'expression du gène *Krox20* (données non présentées). L'élément ze2 est activé plus tardivement que l'élément souris M11 dont l'activité précède l'expression de *Krox20* dans r3 (fig. 21C). Malgré sa conservation partielle de séquence, l'élément ze2 n'a donc pas la même activité que l'élément M11, suggérant que ces éléments ont divergé de fonction entre le poisson-zèbre et la souris. La délétion de cet élément est en cours pour accéder à sa fonction au locus du gène *Krox20*.

Les conséquences de la délétion des éléments A et B sur l'expression du gène *Krox20*, chez le poisson-zèbre, ont été analysées par hybridation *in situ* d'ARNm *Krox20* (fig. 22B et 22C). La délétion de l'élément A entraîne chez le poisson-zèbre une réduction de taille du territoire r5 à 9 et 14 somites, alors que le territoire r3 semble quant à lui moins affecté (fig. 22B). Ce phénotype (en accord avec son profil d'activité) suggère que l'élément A joue un rôle important dans l'expression du gène *Krox20* dans r5 mais mineur dans r3 contrairement à son rôle chez la souris (Bouchoucha et al. 2013). La délétion de l'élément B entraîne un retard de l'expression de *Krox20* dans r5 jusqu'au stade 5 somites alors qu'elle est normale dans r5 à 12 somites. Ces résultats suggèrent que l'élément B est nécessaire à l'expression précoce du gène *Krox20* dans r5 mais peut être compensé plus tardivement par

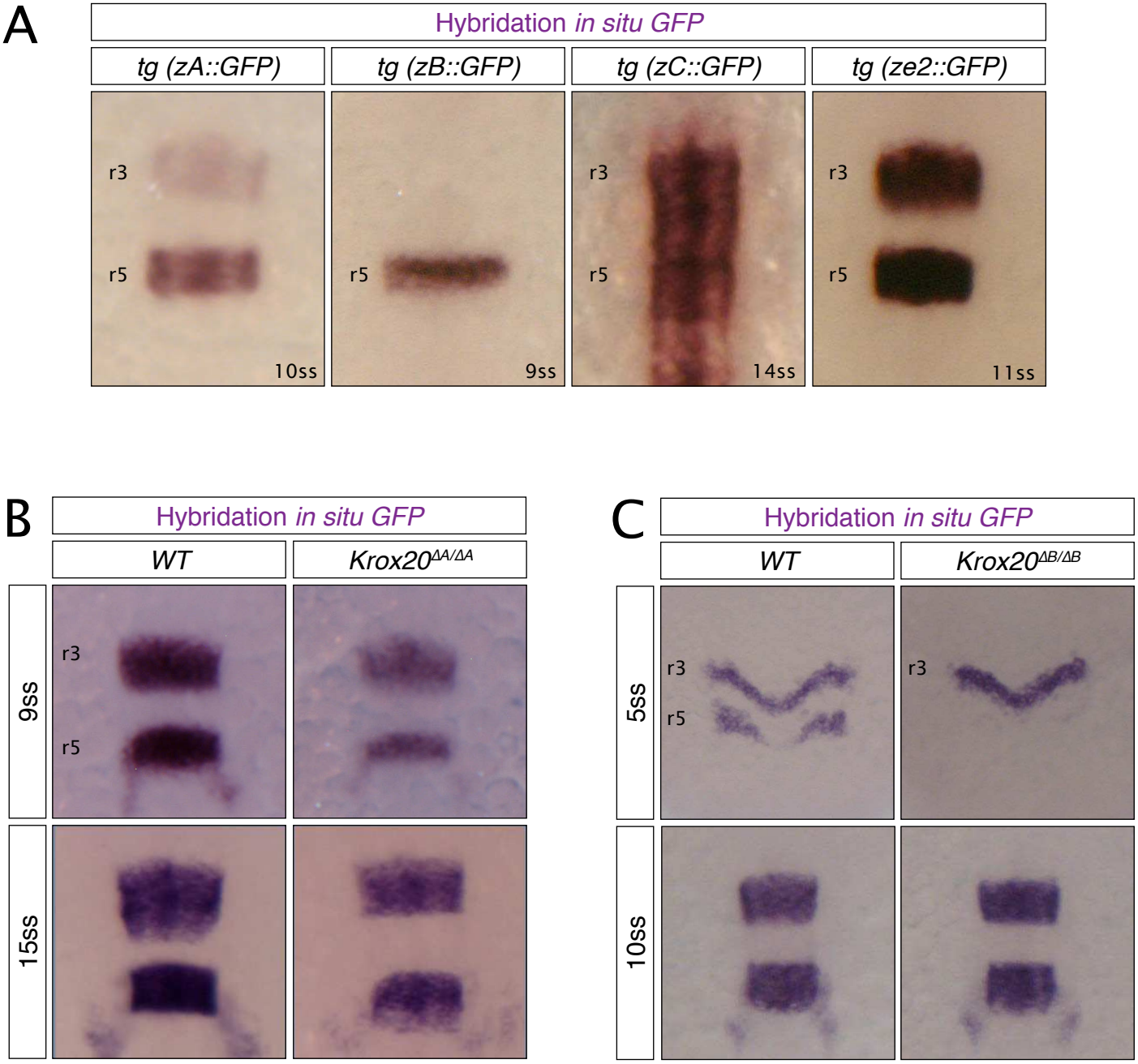


Figure 22 : Les éléments *cis*-régulateurs du gène *Krox20* chez le poisson-zèbre

A. Hybridation *in situ GFP* sur des embryons de poisson-zèbre portant les transgènes *tg (zA::GFP)* à 10 somites (ss), *tg (B::GFP)* à 9ss*, tg (zC::GFP) à et tg (ze2::GFP)* à 11ss. L'élément zA présente une activité dans r3 et r5, l'élément zB dans r5 uniquement, l'élément zC de r3 à r5 et faiblement postérieurement et l'élément ze2 (homologue du M11) dans r3 et r5.

B. Hybridation *in situ GFP* sur des embryons de poisson-zèbre sauvages et mutants *Krox20*$^{\Delta A/\Delta A}$ à 9 et 14 somites.

C. Hybridation *in situ GFP* sur des embryons de poisson-zèbre sauvages et mutants *Krox20*$^{\Delta B/\Delta B}$ à 5 et 12 somites.

(Photos : Patrick Torbey)

un autre élément régulateur. Les autres éléments actifs dans r5, notamment un élément identifié uniquement chez le poisson-zèbre nommé ze8 (Patrick Torbey, communication personnelle) peuvent donc peut-être contribuer à l'expression à la fois précoce et tardive du gène *Krox20* dans r5.

Ensemble, ces résultats montrent que, jusqu'à présent, aucune délétion d'éléments *cis*-régulateurs conservés en séquence entre la souris et le poisson-zèbre n'entraîne de phénotype similaire. Chez la souris, les délétions des éléments A et C abolissent complètement l'expression tardive du gène *Krox20* dans r3 et r5 pour l'élément A et dans r3 pour l'élément C. En revanche, chez le poisson-zèbre, la délétion de l'élément C ne modifie pas l'expression du gène *Krox20*, alors que les délétions des éléments A et B conduisent à des phénotypes subtils qui n'abolissent jamais complètement l'expression du gène *Krox20* dans les rhombomères 3 et 5. Il semble donc que la régulation du gène *Krox20* chez le poisson-zèbre fasse intervenir de nombreux élément *cis*-régulateurs partiellement redondants et dont la fonction est divergente par rapport aux éléments *cis*-régulateurs conservés chez la souris. La délétion des autres éléments *cis*-régulateurs individuellement et par combinaison permettra de caractériser en détail le système de régulation du gène *Krox20* pour l'établissement du patron d'expression dans le rhombencéphale du poisson-zèbre. Il est intéressant de souligner que ce système de régulation chez le poisson-zèbre fait intervenir plusieurs éléments *cis*-régulateurs conservés chez la souris, que leur fonction n'est pas conservée mais permet d'aboutir à un patron d'expression identique chez la souris à l'origine du développement précis et robuste du rhombencéphale et des structures qui en découlent.

6. Limitations de notre système embryonnaire, vers une étude en cellule unique ?

Notre système d'étude qu'est l'embryon n'est pas un tissu homogène puisque l'expression du gène *Krox20* est limitée à deux territoires non-adjacents contenant peu de cellules. Une première solution pour obtenir des cellules d'identité homogène serait un tri cellulaire. Pour cela, un marquage, souvent fluorescent, est requis dans les territoires d'intérêt. Nous possédons au laboratoire la lignée *Krox20*^{GFP} où la séquence codante du gène *Krox20* a été remplacée par celle de la GFP. Il serait possible de trier les cellules exprimant *Krox20* dans le rhombencéphale mais sans distinction de l'origine du territoire r3 ou r5 et le nombre de cellules resterait faible. Par ailleurs, l'utilisation de l'allèle *Krox20*^{GFP} ne permet pas l'étude de cellules triées issues d'embryons mutants homozygotes pour les délétions des éléments A et C. C'est pourquoi j'ai développé un système cellulaire reproduisant l'expression du gène *Krox20* dans l'embryon en développement à partir de cellules

embryonnaires murines. L'idée était de générer un grand nombre de cellules d'identité homogène. Nous avons réussi à obtenir jusqu'à 20% de cellules exprimant *Krox20* et une analyse préliminaire de l'identité de ces cellules indique qu'elles ont dans l'ensemble une identité rhombencéphalique. Cependant, la population était encore hétérogène en identité et le faible pourcentage de cellules exprimant *Krox20* n'était pas suffisant pour utiliser ce protocole de différenciation sur des cellules souches homozygotes pour les délétions des éléments A et C. De plus, un certain nombre de contrôles auraient été nécessaires pour confirmer que l'expression de *Krox20* dans ces cellules est bien dirigée par les éléments *cis*-régulateurs du rhombencéphale et non dus à une autre voie d'activation.

La technique de FISH (Fluorescent *in situ* Hybridization) permet de regarder la conformation de l'ADN en cellule unique en marquant différentes régions génomiques par des sondes fluorescentes sur cellules fixées et de suivre par microscopie la distance entre ces sondes. Des expériences de FISH à haute résolution permettent aujourd'hui d'accéder à des distances génomiques inférieures à 1Mb (Giorgetti et al. 2015) ce qui est encore une résolution insuffisante pour étudier le locus *Krox20*. Dans les prochaines années, il sera possible de réduire encore cette résolution et d'observer les interactions enhancer-promoteur au sein d'un locus. Les techniques actuelles permettant de regarder l'état chromatinien (ChIP-seq) et la conformation de la chromatine (3C et ses grands frères) nécessitent un nombre très important de cellules. Le 4C-seq permettant actuellement d'obtenir la meilleure résolution. Cependant, une étude a décrit en cellule unique une carte d'interaction du chromosome X par Hi-C (Nagano et al. 2013). Il est alors possible d'imaginer que les techniques de génomique actuelles vont évoluer vers une résolution grandissante et permettront d'apprécier l'architecture chromatinienne du noyau de cellules à cellules.

"Discovery depends on technology now more than it ever did." (Vicente 2015, an interview with Mike Levine).

RÉFÉRENCES BIBLIOGRAPHIQUES

Adelman, K. & Lis, J.T., 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics*, 13(10), pp.720–731.

Ali, T., Renkawitz, R. & Bartkuhn, M., 2016. Insulators and domains of gene expression. *Current Opinion in Genetics & Development*, 37, pp.17–26.

Allou, L. et al., 2012. 14q12 and severe Rett-like phenotypes: new clinical insights and physical mapping of FOXG1-regulatory elements. *European Journal of Human Genetics*, 20(12), pp.1216–1223.

Amendola, M. & Van Steensel, B., 2014. Mechanisms and dynamics of nuclear lamina-genome interactions. *Current Opinion in Cell Biology*, 28(1), pp.61–68.

Andersson, R. et al., 2014. An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493), pp.455–61.

Andersson, R., Chen, Y., et al., 2015. Human Gene Promoters Are Intrinsically Bidirectional. *Molecular Cell*, 60(3), pp.346–347.

Andersson, R., Sandelin, A. & Danko, C.G., 2015. A unified architecture of transcriptional regulatory elements. *Trends in Genetics*, 31(8), pp.426–433.

Andrey, G. et al., 2013. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science (New York, NY)*, 340(6137), p.1234167.

Aragon, F. & Pujades, C., 2009. FGF signaling controls caudal hindbrain specification through Ras-ERK1/2 pathway. *BMC developmental biology*, 9, p.61.

Arnosti, D.N. & Kulkarni, M.M., 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry*, 94(5), pp.890–898.

Avery, O.T., Macleod, C.M. & Mccarty, M., 1944. From the Hospital of The Rockefeller Institute for Medical Research. *Strain*, 79(6), pp.137–58.

Banerji, J., Olson, L. & Schaffner, W., 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*, 33(3), pp.729–740.

Banerji, J., Rusconi, S. & Schaffner, W., 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27(2 Pt 1), pp.299–308.

Barolo, S., 2012. Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *BioEssays*, 34(2), pp.135–141.

Bell, A.C., West, A.G. & Felsenfeld, G., 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, 98(3), pp.387–396.

Birgbauer, E. & Fraser, S.E., 1994. Violation of cell lineage restriction compartments in the chick hindbrain. *Development (Cambridge, England)*, 120(6), pp.1347–56.

Birney, E. et al., 2007. Identification and analysis of functional elements in 1% of the human

genome by the ENCODE pilot project. *Nature*, 447(7146), pp.799–816.

Bouchoucha, Y.X. et al., 2013. Dissection of a Krox20 positive feedback loop driving cell fate choices in hindbrain patterning. *Molecular Systems Biology*, 9(690).

Bourc'his, D. et al., 2001. Dnmt3L and the establishment of maternal genomic imprints. *Science (New York, N.Y.)*, 294(5551), pp.2536–2539.

Bouwman, B.A.M. & de Laat, W., 2015. Getting the genome in shape: the formation of loops, domains and compartments. *Genome biology*, 16, p.154.

Boyle, A.P. et al., 2008. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell*, 132(2), pp.311–322.

Braasch, I. et al., 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nature Genetics*.

Branco, M.R. & Pombo, A., 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biology*, 4(5), pp.780–788.

Buenrostro, J.D. et al., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods*, 10(12), pp.1213–8.

Cannavò, E. et al., 2016. Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Current biology : CB*, 26(1), pp.38–51.

Chambers, S.M. et al., 2009. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature biotechnology*, 27(3), pp.275–280.

Chan, Y.F. et al., 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science (New York, N.Y.)*, 327(5963), pp.302–5.

Chavrier, P. et al., 1988. A gene encoding a protein with zinc fingers is activated during G0/G1 transition in cultured cells. *The EMBO journal*, 7(1), pp.29–35.

Chavrier, P. et al., 1990. The segment-specific gene Krox-20 encodes a transcription factor with binding sites in the promoter region of the Hox-1.4 gene. *The EMBO journal*, 9(4), pp.1209–18.

Choe, S.-K., Vlachakis, N. & Sagerström, C.G., 2002. Meis family proteins are required for hindbrain development in the zebrafish. *Development*, 129(3), pp.585–595.

Chomette, D. et al., 2006. Krox20 hindbrain cis-regulatory landscape: interplay between multiple long-range initiation and autoregulatory elements. *Development (Cambridge, England)*, 133(7), pp.1253–62.

Cirillo, L.A. et al., 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular Cell*, 9(2), pp.279–289.

Cordes, S.P. & Barsh, G.S., 1994. The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transcription factor. *Cell*, 79(6), pp.1025–1034.

Core, L.J. et al., 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature genetics*, 46(12), pp.1311–20.

Core, L.J., Waterfall, J.J. & Lis, J.T., 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science (New York, N.Y.)*, 322(5909), pp.1845–8.

Cremer, M. et al., 2001. Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Research*, 9(7), pp.541–567.

Cremer, T. et al., 1982. Rabl's model of the interphase chromosome arrangement tested in Chinise hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Human Genetics*, 60(1), pp.46–56.

Crick, F., 1958. On protein synthesis. *Symposia of the Society for Experimental Biology*, 12, pp.138–163.

Crocker, J. et al., 2015. Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness. *Cell*, 160(1-2), pp.191–203.

Crocker, J., Preger-Ben Noon, E. & Stern, D.L., 2016. The Soft Touch. In *Current Topics in Developmental Biology*. Elsevier Inc., pp. 455–469.

Croft, J.A. et al., 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *Journal of Cell Biology*, 145(6), pp.1119–1131.

Davuluri, R. V. et al., 2008. The functional consequences of alternative promoter use in mammalian genomes. *Trends in Genetics*, 24(4), pp.167–177.

Deans, C. & Maggert, K. a., 2015. What Do You Mean, "Epigenetic"? *Genetics*, 199(4), pp.887–896.

Dekker, J. et al., 2002. Capturing chromosome conformation. *Science (New York, N.Y.)*, 295(5558), pp.1306–11.

Dekker, J. & Heard, E., 2015. Structural and functional diversity of Topologically Associating Domains. *FEBS Letters*, 589(20), pp.2877–2884.

Dekker, J., Marti-Renom, M. a & Mirny, L. a, 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature reviews. Genetics*, 14(6), pp.390–403.

Deng, W. et al., 2012. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, 149(6), pp.1233–44.

Deng, W. et al., 2014. Reactivation of Developmentally Silenced Globin Genes by Forced Chromatin Looping. *Cell*, 158(4), pp.849–860.

Le Dily, F. et al., 2014. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes & Development*, 28(19), pp.2151–2162.

Dixon, J.R. et al., 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518(7539), pp.331–336.

Dixon, J.R. et al., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), pp.376–80.

Doolittle, R.F., 1986. Of URFs and ORFs: A primer on how to analyze derived amino acid sequences. In University Science Books, p. 1986.

Dostie, J. et al., 2006. Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Research*, 16(10), pp.1299–1309.

Dowen, J.M. et al., 2014. Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. *Cell*, 159(2), pp.374–387.

Duan, Z.-J. et al., 2002. Developmental specificity of recruitment of TBP to the TATA box of the human gamma-globin gene. *Proceedings of the National Academy of Sciences of the United States of America*, 99(8), pp.5509–5514.

Dunham, I. et al., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74.

Duttke, S.H.C. et al., 2015. Human promoters are intrinsically directional. *Molecular Cell*, 57(4), pp.674–684.

Farley, E.K. et al., 2015. Suboptimization of developmental enhancers. *Science*, 350(6258), pp.325–328.

Farley, E.K. et al., 2016. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proceedings of the National Academy of Sciences*, p.201605085.

Fatica, A. & Bozzoni, I., 2014. Long non-coding RNAs: new players in cell differentiation and development. *Nature reviews. Genetics*, 15(1), pp.7–21.

Feingold, E. et al., 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science (New York, N.Y.)*, 306(October), pp.636–640.

Finlan, L.E. et al., 2008. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genetics*, 4(3).

Fisher, S. et al., 2006. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science*, 312(5771), pp.276–279.

Flynn, R.A. et al., 2011. Antisense RNA polymerase II divergent transcripts are P-TEFb

dependent and substrates for the RNA exosome. *Proceedings of the National Academy of Sciences of the United States of America*, 108(26), pp.10460–5.

Frankel, N. et al., 2011. Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature*, 474(7353), pp.598–603.

Frankel, N. et al., 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*, 466(7305), pp.490–493.

Fraser, P. & Bickmore, W., 2007. Nuclear organization of the genome and the potential for gene regulation. *Nature*, 447(7143), pp.413–417.

Fudenberg, G. et al., 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell reports*, 15(9), pp.2038–49.

Gaudet, J. & Mango, S.E., 2002. Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. *Science (New York, N.Y.)*, 295(February), pp.821–825.

Gerstein, M.B. et al., 2007. What is a gene, post-ENCODE? History and updated definition. *Genome Research*, 17, pp.669–681.

Ghavi-helm, Y. et al., 2014. and Are Associated With Paused Polymerase. *Nature*, 512(7512), pp.96–100.

Ghislain, J. et al., 2002. Characterisation of cis-acting sequences reveals a biphasic, axon-dependent regulation of Krox20 during Schwann cell development. *Development (Cambridge, England)*, 129(1), pp.155–166.

Ghislain, J. et al., 2003. Neural crest patterning: autoregulatory and crest-specific elements co-operate for Krox20 transcriptional control. *Development (Cambridge, England)*, 130(5), pp.941–953.

Gibcus, J.H. & Dekker, J., 2013. The hierarchy of the 3D genome. *Molecular cell*, 49(5), pp.773–82.

Gillies, S.D. et al., 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell*, 33(3), pp.717–728.

Giorgetti, L., Piolot, T. & Heard, E., 2015. High-resolution 3D DNA fish using plasmid probes and computational correction of optical aberrations to study chromatin structure at the sub-megabase scale. *Methods in Molecular Biology*, 1262, pp.37–53.

Giresi, P.G. et al., 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. , pp.877–885.

Giudicelli, F. et al., 2001. Krox-20 patterns the hindbrain through both cell-autonomous and non cell-autonomous mechanisms. *Genes & development*, 15(5), pp.567–80.

Gómez-Marín, C. et al., 2015. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proceedings of the National Academy of Sciences*, p.201505463.

Goode, D.K. et al., 2011. Minor change, major difference: divergent functions of highly conserved cis-regulatory elements subsequent to whole genome duplication events. *Development (Cambridge, England)*, 138, pp.879–884.

Goodwin, A.J. et al., 2001. In Vivo Formation of a Human ??-Globin Locus Control Region Core Element Requires Binding Sites for Multiple Factors Including GATA-1, NF-E2, Erythroid Kruppel-like Factor, and Sp1. *Journal of Biological Chemistry*, 276(29), pp.26883–26892.

Grau, D.J. et al., 2011. Compaction of chromatin by diverse polycomb group proteins requires localized regions of high charge. *Genes and Development*, 25(20), pp.2210–2221.

Greer, E.L. & Shi, Y., 2012. Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics*, 13(5), pp.343–357.

Griffith, F., 1928. Pneumococcal Types. *Journ. of Hyg.*, 27(2), pp.113–159.

Grosveld, F. et al., 1987. Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell*, 51(6), pp.975–985.

Guelen, L. et al., 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453(7197), pp.948–51.

Guerreiro, I. et al., 2013. Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. *Proceedings of the National Academy of Sciences of the United States of America*, 110(26), pp.10682–6.

Guo, Y. et al., 2015. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, 162(4), pp.900–910.

Guthrie, S., 2007. Patterning and axon guidance of cranial motor neurons. *Nature reviews. Neuroscience*, 8(11), pp.859–71.

Hadjur, S. et al., 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*, 460(7253), pp.410–413.

Handoko, L. et al., 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature genetics*, 43(7), pp.630–638.

Harrison, M.M. et al., 2011. Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genetics*, 7(10).

Heger, P. et al., 2012. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proceedings of the National Academy of Sciences*, 109(43), pp.17507–17512.

Hirayama, T. et al., 2012. CTCF Is Required for Neural Development and Stochastic Expression of Clustered Pcdh Genes in Neurons. *Cell Reports*, 2(2), pp.345–357.

Holliday, R., 1994. Epigenetics: an overview. *Developmental genetics*, 15(6), pp.453–457.

Hong, J.-W., Hendrix, D. a & Levine, M.S., 2008. Shadow enhancers as a source of evolutionary novelty. *Science (New York, N.Y.)*, 321(5894), p.1314.

Horn, S. et al., 2013. TERT Promoter Mutations in Familial and Sporadic Melanoma. *Science*, 339(6122), pp.959–961.

Hou, C. et al., 2012. Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. *Molecular Cell*, 48(3), pp.471–484.

Huang, F.W. et al., 2013. Highly Recurrent TERT Promoter Mutations in Human Melanoma. *Science*, 339(6122), pp.957–959.

Hwang, W.Y. et al., 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology*, 31(3), pp.227–9.

Ibn-Salem, J. et al., 2014. Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome biology*, 15(9), p.423.

Indjeian, V.B. et al., 2016. Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins. *Cell*, 164(1-2), pp.45–56.

Ing-simmons, E. et al., 2015. Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. , 4, pp.1–10.

Iwafuchi-Doi, M. et al., 2016. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Molecular Cell*, 62(1), pp.79–91.

Jacob, F., 1977. Evolution and tinkering. *Science (New York, N.Y.)*, 196(4295), pp.1161–6.

Jao, L.-E., Wente, S.R. & Chen, W., 2013. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences of the United States of America*, 110(34), pp.13904–9.

Jenuwein, T. & Allis, C.D., 2001. Translating the histone code. *Science (New York, N.Y.)*, 293(5532), pp.1074–80.

Jin, C. et al., 2009. H3.3/H2A.Z double variant-containing nucleosomes mark “nucleosome-free regions” of active promoters and other regulatory regions. *Nature genetics*, 41(8), pp.941–945.

Jonkers, I. & Lis, J.T., 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nature reviews. Molecular cell biology*, 16(3), pp.167–177.

Junion, G. et al., 2012. A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell*, 148(3), pp.473–486.

Kagey, M.H. et al., 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314), pp.430–435.

Katainen, R. et al., 2015. CTCF/cohesin-binding sites are frequently mutated in cancer. *Nature Genetics*, advance on(7), pp.818–21.

Kellner, W.A. et al., 2012. Genome-wide phosphoacetylation of histone H3 at Drosophila enhancers and promoters. *Genome Research*, 22(6), pp.1081–1088.

Kellum, R. & Schedl, P., 1991. A position-effect assay for boundaries of higher order chromosomal domains. *Cell*, 64(5), pp.941–950.

Khurana, E. et al., 2016. Role of non-coding sequence variants in cancer. *Nat Rev Genet*, 17(2), pp.93–108.

Kim, H. & Loparo, J.J., 2016. Multistep assembly of DNA condensation clusters by SMC. *Nature Communications*, 7, pp.1–12.

Kim, T.-K. et al., 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 465(7295), pp.182–187.

Kim, T.-K. & Shiekhattar, R., 2015. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell*, 162(5), pp.948–959.

Kind, J. et al., 2013. Single-cell dynamics of genome-nuclear lamina interactions. *Cell*, 153(1), pp.178–192.

Kobor, M.S. et al., 2004. A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. *PLoS Biology*, 2(5).

Kondo, T., Ito, S. & Koseki, H., 2016. Polycomb in Transcriptional Phase Transition of Developmental Genes. *Trends in Biochemical Sciences*, 41(1), pp.9–19.

Kumar, S. & Hedges, S.B., 1998. A molecular timescale for vertebrate evolution. *Nature*, 392(6679), pp.917–920.

Kurokawa, D. et al., 2004. Regulation of Otx2 expression and its functions in mouse forebrain and midbrain. *Development (Cambridge, England)*, 131(14), pp.3319–31.

de Laat, W. & Duboule, D., 2013. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*, 502(7472), pp.499–506.

Labalette, C. et al., 2011. Hindbrain patterning requires fine-tuning of early krox20 transcription by Sprouty 4. *Development (Cambridge, England)*, 138(2), pp.317–26.

Labalette, C. et al., 2015. Molecular dissection of segment formation in the developing hindbrain. *Development*, 142(1), pp.185–195.

Lagha, M., Bothma, J.P. & Levine, M., 2012. Mechanisms of transcriptional precision in animal development. *Trends in genetics : TIG*, 28(8), pp.409–16.

Lai, F. et al., 2015. Integrator mediates the biogenesis of enhancer RNAs. , (305).

Lam, D.D. et al., 2015. Partially Redundant Enhancers Cooperatively Maintain Mammalian Pomc Expression Above a Critical Functional Threshold. *PLOS Genetics*, 11(2), p.e1004935.

Lam, M.T.Y. et al., 2014. Enhancer RNAs and regulated transcriptional programs. *Trends in Biochemical Sciences*, 39(4), pp.170–182.

Lam, M.T.Y. et al., 2013. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature*, 498(7455), pp.511–5.

Larkin, J.D. et al., 2013. Space exploration by the promoter of a long human gene during one transcription cycle. *Nucleic Acids Research*, 41(4), pp.2216–2227.

Lee, K. et al., 2015. Dynamic enhancer – Gene body contacts during transcription elongation. *Genes & Development*, pp.1992–1997.

Lettice, L.A. et al., 2011. Enhancer-adoption as a mechanism of human developmental disease. *Human Mutation*, 32(12), pp.1492–1499.

Lewin, B. et al., 2011. Lewin's GENES X, Volume 10. , 10, p.930.

Li, W. et al., 2013. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, 498(7455), pp.516–20.

Liang, H.-L. et al., 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature*, 456(7220), pp.400–403.

Lichter, P. et al., 1988. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Human Genetics*, 80(3), pp.224–234.

Lieberman-Aiden, E. et al., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, 326(5950), pp.289–93.

Lippmann, E.S. et al., 2015. Deterministic HOX patterning in human pluripotent stem cell-derived neuroectoderm. *Stem Cell Reports*, 4(4), pp.632–644.

Lippmann, E.S., Estevez-Silva, M.C. & Ashton, R.S., 2014. Defined human pluripotent stem cell culture enables highly efficient neuroepithelium derivation without small molecule inhibitors. *Stem Cells*, 32(4), pp.1032–1042.

Lister, R. et al., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), pp.315–322.

Losada, A., 2014. Cohesin in cancer: chromosome segregation and beyond. *Nature Reviews Cancer*, 14(6), pp.389–393.

Lowe, C.B. et al., 2011. Three periods of regulatory innovation during vertebrate evolution. *Science (New York, N.Y.)*, 333(6045), pp.1019–24.

Ludwig, M.Z. et al., 2011. Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. *PLoS Genetics*, 7(11).

Ludwig, M.Z. et al., 2005. Functional evolution of a cis-regulatory module. *PLoS Biology*, 3(4), pp.0588–0598.

Lumsden, A. & Keynes, R., 1989. Segmental patterns of neuronal development in the chick hindbrain. *Nature*, 337(6206), pp.424–8.

Lupiáñez, D.G. et al., 2015. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*, pp.1012–1025.

Lynch, V.J. et al., 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature genetics*, 43(11), pp.1154–9.

Madisen, L. et al., 2010. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neuroscience*, 13(1), pp.133–140.

Maniatis, T.O.M., Goodbourn, S. & Fischer, J. a, 1987. Inducible and.

Mansfield, J.H., 2013. Cis-Regulatory Change Associated With Snake Body Plan Evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 110(26), pp.10473–10474.

Maston, G. a, Evans, S.K. & Green, M.R., 2006. Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics*, 7, pp.29–59.

Maurano, M.T. et al., 2012. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science*, 337(6099), pp.1190–1195.

Maury, Y. et al., 2014. Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes. *Nature Biotechnology*, 33(1), pp.89–96.

McNulty, C.L. et al., 2005. Knockdown of the complete Hox paralogous group 1 leads to dramatic hindbrain and neural crest defects. *Development (Cambridge, England)*, 132(12), pp.2861–71.

Meissner, A. et al., 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic acids research*, 33(18), pp.5868–77.

Melo, C.A. et al., 2013. ERNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription. *Molecular Cell*, 49(3), pp.524–535.

Mendel, G. & Bateson, W., 1866. Versuche über Pflanzenhybriden. Verhandlungen des naturforschenden Vereines in Brünn 4 Abhandlungen. *Journal of the Royal Horticultural Society*, 26(1865), pp.3–47.

Merika, M. & Thanos, D., 2001. Enhanceosomes. *Current Opinion in Genetics &*

Development, 11(2), pp.205–208.

Meuleman, W. et al., 2013. Constitutive nuclear lamina – genome interactions are highly conserved and associated with A / T-rich sequence. *Genome research*, pp.270–280.

Mifsud, B. et al., 2015. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*, 47(6), pp.598–606.

Montavon, T. et al., 2011. A regulatory archipelago controls Hox genes transcription in digits. *Cell*, 147(5), pp.1132–45.

Moore, L.D., Le, T. & Fan, G., 2012. DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, 38(1), pp.23–38.

Morgan, T.H. et al., 1915. The mechanism of Mendelian heridity. *Holt Rinehart & Winston, New York*.

Mousavi, K. et al., 2013. ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Molecular Cell*, 51(5), pp.606–617.

Nagano, T. et al., 2013. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469), pp.59–64.

Narendra, V. et al., 2015. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. , 347(6225), pp.1017–1022.

Nasmyth, K. & Haering, C.H., 2009. Cohesin: Its Roles and Mechanisms. *Annual Review of Genetics*, 43(1), pp.525–558.

Nichols, M.H. & Corces, V.G., 2015. A CTCF Code for 3D Genome Architecture. *Cell*, 162(4), pp.703–5.

Nien, C.Y. et al., 2011. Temporal coordination of gene networks by Zelda in the early Drosophila embryo. *PLoS Genetics*, 7(10).

Nonchev, S. et al., 1996. Segmental expression of Hoxa-2 in the hindbrain is directly regulated by Krox-20. *Development (Cambridge, England)*, 122(2), pp.543–54.

Nora, E.P. et al., 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398), pp.381–5.

Ocampo-Hafalla, M. et al., 2016. Evidence for cohesin sliding along budding yeast chromosomes. *Open biology*, 6(6).

Oda-Ishii, I. et al., 2005. Making very similar embryos with divergent genomes: conservation of regulatory mechanisms of Otx between the ascidians Halocynthia roretzi and Ciona intestinalis. *Development (Cambridge, England)*, 132(7), pp.1663–1674.

Ohno, S., 1972. So much "junk" DNA in our genome. In H. H. Smith, ed. *Evolution of genetic systems*. New York: Gordon & Breach, pp. 366–370.

Okano, M. et al., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de

novo methylation and mammalian development. *Cell*, 99(3), pp.247–257.

Ong, C.-T. & Corces, V.G., 2014. CTCF: an architectural protein bridging genome topology and function. *Nature reviews. Genetics*, 15(4), pp.234–46.

Osborne, C.S. et al., 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature genetics*, 36(10), pp.1065–1071.

Papantonis, A. et al., 2010. Active RNA polymerases: Mobile or immobile molecular machines? *PLoS Biology*, 8(7).

Parelho, V. et al., 2008. Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell*, 132(3), pp.422–433.

Payankaulam, S., Li, L.M. & Arnosti, D.N., 2010. Transcriptional repression: Conserved and evolved features. *Current Biology*, 20(17), pp.764–771.

Pearson, H., 2006. Genetics: what is a gene? *Nature*, 441(7092), pp.398–401.

Peric-Hupkes, D. et al., 2010. Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. *Molecular Cell*, 38(4), pp.603–613.

Perry, M.W., Boettiger, A.N. & Levine, M., 2011. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Pnas*, 108(33), pp.1–12.

Peterlin, B.M. & Price, D.H., 2006. Controlling the Elongation Phase of Transcription with P-TEFb. *Molecular Cell*, 23(3), pp.297–305.

Phillips-Cremins, J.E. et al., 2013. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, 153(6), pp.1281–95.

Pickersgill, H. et al., 2006. Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nature genetics*, 38(9), pp.1005–14.

Pinkel, D. et al., 1988. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proceedings of the National Academy of Sciences of the United States of America*, 85(23), pp.9138–9142.

Pnueli, L. et al., 2015. RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin α-subunit gene. *Proceedings of the National Academy of Sciences*, 112(14), p.201414841.

Pombo, A. & Dillon, N., 2015. Three-dimensional genome architecture: players and mechanisms. *Nature Reviews Molecular Cell Biology*, 12(March).

Poss, Z.C., Ebmeier, C. & Dylan, T., 2013. The Mediator complex and transcription elongation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(1), pp.69–75.

Pott, S. & Lieb, J.D., 2015. What are super-enhancers ? *Nature Publishing Group*, 47(1),

pp.8–12.

Preker, P. et al., 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science (New York, N.Y.)*, 322(5909), pp.1851–4.

Prescott, S.L. et al., 2015. Enhancer Divergence and cis -Regulatory Evolution in the Human and Chimp Neural Crest Article Enhancer Divergence and cis -Regulatory Evolution in the Human and Chimp Neural Crest. *Cell*, pp.68–83.

Rabl, C., 1885. Über Zelltheilung. *Morphol. Jahrb.*, 10, pp.214–330.

Ramos, A.I. & Barolo, S., 2013. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1632), p.20130018.

Rao, S.S.P. et al., 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7), pp.1665–1680.

Remeseiro, S. et al., 2012. A unique role of cohesin-SA1 in gene regulation and development. *The EMBO Journal*, 31(9), pp.2090–2102.

Reyon, D. et al., 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nature biotechnology*, 30(5), pp.460–5.

Roll-Hansen, N., 1989. The crucial experiment of Wilhelm Johannsen. *Biology & Philosophy*, 4(3), pp.303–329.

Sanborn, A.L. et al., 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences*, 112(47), p.201518552.

De Santa, F. et al., 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS biology*, 8(5), p.e1000384.

Sanyal, A. et al., 2012. The long-range interaction landscape of gene promoters. *Nature*, 489(7414), pp.109–113.

Scardigli, R. et al., 2003. Direct and concentration-dependent regulation of the proneural gene Neurogenin2 by Pax6. *Development (Cambridge, England)*, 130(14), pp.3269–3281.

Schaaf, C.A. et al., 2013. Genome-Wide Control of RNA Polymerase II Activity by Cohesin. *PLoS Genetics*, 9(3).

Schaukowitch, K. et al., 2014. Enhancer RNA facilitates NELF release from immediate early genes. *Molecular Cell*, 56(1), pp.29–42.

Schneider-Maunoury, S. et al., 1993. Disruption of Krox-20 results in alteration of rhombomeres 3 and 5 in the developing hindbrain. *Cell*, 75(6), pp.1199–214.

Schneider-Maunoury, S. et al., 1997. Segmental and neuronal architecture of the hindbrain of Krox-20 mouse mutants. *Development (Cambridge, England)*, 124(6), pp.1215–26.

Schoenfelder, S. et al., 2010. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature genetics*, 42(1), pp.53–61.

Schoenfelder, S. et al., 2015. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. , pp.1–16.

Schulz, K.N. et al., 2015. Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Research*, 25(11), pp.1715–1726.

Schwarzer, W. & Spitz, F., 2014. The architecture of gene expression: Integrating dispersed cis-regulatory modules into coherent regulatory domains. *Current Opinion in Genetics and Development*, 27, pp.74–82.

Scruggs, B.S. et al., 2015. Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Molecular cell*, 58(6), pp.1101–12.

Searls, D.B., 2002. The language of genes. *Nature*, 420(6912), pp.211–217.

Seila, A.C. et al., 2008. Divergent transcription from active promoters. *Science (New York, N.Y.)*, 322(5909), pp.1849–1851.

Seitan, V.C. et al., 2013. Cohesin-Based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Research*, 23(12), pp.2066–2077.

Seitanidou, T. et al., 1997. Krox-20 is a key regulator of rhombomere-specific gene expression in the developing hindbrain. *Mechanisms of development*, 65(1-2), pp.31–42.

Seoighe, D.M. et al., 2014. A chromosomal 5q31.1 gain involving PITX1 causes Liebenberg syndrome. *American Journal of Medical Genetics, Part A*, 164(11), pp.2958–2960.

Sérandour, A.A. et al., 2011. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Research*, 21(4), pp.555–565.

Sexton, T. et al., 2012. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148(3), pp.458–472.

Sham, M.H. et al., 1993. The zinc finger gene Krox20 regulates HoxB2 (Hox2.8) during hindbrain segmentation. *Cell*, 72(2), pp.183–196.

Shapiro, M.D. et al., 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, 428(6984), pp.717–723.

Shiraki, T. et al., 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences of the United States of America*, 100(26), pp.15776–81.

Siepel, A. & Arbiza, L., 2014. ScienceDirect Cis -regulatory elements and human evolution. *Current Opinion in Genetics & Development*, 29(Figure 1), pp.81–89.

Spielmann, M. et al., 2012. Homeotic arm-to-leg transformation associated with genomic rearrangements at the PITX1 locus. *American Journal of Human Genetics*, 91(4), pp.629–635.

Spitz, F. & Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. *Nature reviews. Genetics*, 13(9), pp.613–26.

Splinter, E. et al., 2006. CTCF mediates long-range chromatin looping and local histone modification in the ??-globin locus. *Genes and Development*, 20(17), pp.2349–2354.

van Steensel, B. & Dekker, J., 2010. Genomics tools for unraveling chromosome architecture. *Nature biotechnology*, 28(10), pp.1089–95.

Sun, Y. et al., 2015. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. *Genome Research*, 25(11), pp.1703–1714.

Symmons, O. et al., 2014. Functional and topological characteristics of mammalian regulatory domains. *Genome Research*, 24(3), pp.390–400.

Taatjes, D.J., Marr, M.T. & Tjian, R., 2004. Regulatory diversity among metazoan co-activator complexes. *Nature reviews. Molecular cell biology*, 5(5), pp.403–10.

Taberlay, P.C. et al., 2016. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. , pp.719–731.

Taft, R.J., Pheasant, M. & Mattick, J.S., 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays*, 29(3), pp.288–299.

Tahiliani, M. et al., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (New York, N.Y.)*, 324(5929), pp.930–935.

Takahashi, H. et al., 2011. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell*, 146(1), pp.92–104.

Tanabe, H. et al., 2002. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proceedings of the National Academy of Sciences of the United States of America*, 99(7), pp.4424–9.

Tang, Z. et al., 2015. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*, 163(7), pp.1–17.

Thanos, D. & Maniatis, T., 1995. Virus induction of human IFN?? gene expression requires the assembly of an enhanceosome. *Cell*, 83(7), pp.1091–1100.

Tie, F. et al., 2016. Polycomb inhibits histone acetylation by CBP by binding directly to its catalytic domain. *Proceedings of the National Academy of Sciences*, 113(6), pp.E744–E753.

Topilko, P. et al., 1994. Krox-20 controls myelination in the peripheral nervous system.

Nature, 371(6500), pp.796–9.

Verdin, E. & Ott, M., 2014. 50 Years of Protein Acetylation: From Gene Regulation To Epigenetics, Metabolism and Beyond. *Nature reviews. Molecular cell biology*, 16(4), pp.258–264.

Vermeren, M. et al., 2003. Integrity of developing spinal motor columns is regulated by neural crest derivatives at motor exit points. *Neuron*, 37(3), pp.403–15.

Vicente, C., 2015. An interview with Mike Levine. *Development (Cambridge, England)*, 142(20), pp.3453–3455.

Vietri Rudan, M. et al., 2015. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Reports*, 10(8), pp.1297–1309.

Vieux-Rochas, M. et al., 2015. Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. *Proceedings of the National Academy of Sciences of the United States of America*, 112(15), pp.4672–7.

Villar, D. et al., 2015. Enhancer Evolution across 20 Mammalian Species. *Cell*, 160(3), pp.554–566.

Visel, A. et al., 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231), pp.854–8.

Voiculescu, O. et al., 2001. Hindbrain patterning: Krox20 couples segmentation and specification of regional identity. *Development (Cambridge, England)*, 128(24), pp.4967–78.

Voiculescu, O., Charnay, P. & Schneider-Maunoury, S., 2000. Expression pattern of a Krox-20/Cre knock-in allele in the developing hindbrain, bones, and peripheral nervous system. *Genesis (New York, N.Y. : 2000)*, 26(2), pp.123–6.

Waddington, C.H., 1942. The Epigenotype. *Endeavour*, (1), pp.18–20.

Walshe, J. et al., 2002. Establishment of hindbrain segmental identity requires signaling by FGF3 and FGF8. *Current Biology*, 12(13), pp.1117–1123.

Waskiewicz, A.J., Rikhof, H. a & Moens, C.B., 2002. Reveals a Hindbrain Ground State. *Developmental Cell*, 3, pp.723–733.

Wassef, M. a et al., 2008. Rostral hindbrain patterning involves the direct activation of a Krox20 transcriptional enhancer by Hox/Pbx and Meis factors. *Development (Cambridge, England)*, 135(20), pp.3369–78.

Watson, J.D. & Crick, F.H.C., 1953. Molecular structure of nucleic acids. *Nature*, 171, pp.737–738.

Weedon, M.N. et al., 2014. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nature genetics*, 46(1), pp.61–4.

Wendt, K.S. et al., 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, 451(7180), pp.796–801.

Whyte, W. a. et al., 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, 153(2), pp.307–319.

Wiellette, E.L. & Sive, H., 2003. vhnf1 and Fgf signals synergize to specify rhombomere identity in the zebrafish hindbrain. *Development (Cambridge, England)*, 130(16), pp.3821–3829.

Wijchers, P.J. et al., 2016. Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. *Molecular Cell*, 61(3), pp.1–13.

Wilkinson, D.G. et al., 1989. Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. *Nature*, 337(6206), pp.461–4.

Wit, E. De et al., 2015. CTCF Binding Polarity Determines Chromatin Looping. *Molecular Cell*, 60(4), pp.1–9.

Wu, H. et al., 2014. Tissue-Specific RNA Expression Marks Distant-Acting Developmental Enhancers. *PLoS Genetics*, 10(9).

Xu, Q. et al., 1999. In vivo cell sorting in complementary segmental domains mediated by Eph receptors and ephrins. *Nature*, 399(6733), pp.267–71.

Yoder, J. a. & Bestor, T.H., 1998. A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Human Molecular Genetics*, 7(2), pp.279–284.

Zabidi, M. a. et al., 2014. Enhancer—core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, 518(7540), pp.556–559.

Zeng, W., Ball, A.R. & Yokomori, K., 2010. HP1: heterochromatin binding proteins working the genome. *Epigenetics*, 5(4), pp.287–92.

Zhang, Y. et al., 2012. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell*, 148(5), pp.908–921.

Zhou, V.W., Goren, A. & Bernstein, B.E., 2011a. Charting histone modifications and the functional organization of mammalian genomes. *Nature reviews. Genetics*, 12(1), pp.7–18.

Zhou, V.W., Goren, A. & Bernstein, B.E., 2011b. Charting histone modifications and the functional organization of mammalian genomes. *Nature reviews. Genetics*, 12(1), pp.7–18.

Zuin, J. et al., 2014. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci USA*, 111(3), pp.996–1001.

Résumé

Les éléments *cis*-régulateurs contrôlent l'expression spatio-temporelle des gènes au cours du développement. La segmentation du rhombencéphale en rhombomères (r) établit l'organisation du cerveau postérieur des vertébrés. La spécification et la formation des segments r3 et r5 sont gouvernées par l'expression du facteur de transcription Krox20/Egr2, sous le contrôle de plusieurs éléments *cis*-régulateurs. Les éléments B (dans r5) et C (dans r3 et r5) sont actifs au cours du démarrage de l'expression de *Krox20* alors que l'élément A (dans r3 et r5) est un élément autorégulateur. En générant la délétion de l'élément A dans le génome murin, nous avons montré qu'il est nécessaire au maintien de l'expression de *Krox20* dans r3 et r5 et qu'en son absence les cellules ne sont pas spécifiées correctement et l'architecture neuronale est perturbée. De façon inattendue, la délétion de l'élément C n'impacte pas la phase de démarrage mais abolit le maintien de l'expression de *Krox20* dans r3 de manière analogue à celle de l'élément A. Un autre élément permet donc le démarrage de l'expression de *Krox20* dans r3, peut-être de façon redondante avec l'élément C. J'ai mis en évidence par différentes approches génétiques que l'élément C coopère en *cis* avec l'élément A, lors de la phase d'autorégulation. Plusieurs approches de conformation et d'accessibilité de la chromatine au locus *Krox20* indiquent que l'élément A interagit de façon constitutive avec le promoteur et que l'élément C est capable d'affecter l'accessibilité de l'élément A à distance. Parallèlement, j'ai montré que la délétion de l'élément C dans le génome du poisson-zèbre n'entraîne pas de modification de l'expression du gène *Krox20* contrairement à la souris malgré un patron d'expression similaire du gène. Ces approches complémentaires ont permis de caractériser le rôle fonctionnel de plusieurs éléments *cis*-régulateurs contrôlant l'expression d'un gène clef du développement des Vertébrés.

Abstract

Cis-regulatory elements ensure the specifity and the maintenance of the spatio-temporal gene expression pattern throughout the development of an organism. Segmentation of the hindbrain into rhombomeres (r) establishes the posterior brain organisation in vertebrates. The specification and formation of segments r3 and r5 are governed by the expression of the transcription factor Krox20/Egr2, controlled by three *cis*-regulatory elements. Elements B (in r5) and C (in r3 and r5) are active during the initiation of *Krox20* expression while element A (in r3 and r5) is an autoregulatory element. By studying the deletion of element A in the mouse genome, we have shown that *Krox20* expression is not maintained in absence of this element and that cells are improperly specified and neuronal architecture is disturbed, emphasizing the importance of the autoregulatory phase in the hindbrain patterning. Unexpectedly, the deletion of element C does not impair *Krox20* initiation but abolishes the maintenance of expression in r3, similarly to element A knock-out. This suggests that another element allows the initiation of *Krox20* expression in r3 and could be redundant with element C. I also showed by different genetic approaches that element C cooperates in *cis* with element A. Moreover, investigation of chromatin conformation and accessibility in the *Krox20* locus indicate that element A and the promoter are engaged in a constitutive interaction and that element C is able to affect the accessibility of element A at a distance. In parallel, the deletion of element C in zebrafish did not reveal any modification in *Krox20* expression, unlike in mouse and despite a similar expression pattern. These complementary approaches lead to the characterization of the functional role of several *cis*-regulatory elements driving the expression of a key developmental gene in vertebrates.