Conserved function of the long non-coding RNA PVT1 in vertebrate development

Florian Constanty

PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Institut Curie
Laboratory of long intervening noncoding RNAs in vertebrate development

Conserved function of the long non-coding RNA *PVT1* in vertebrate development

L'ARN long non codant *PVT1*, une fonction conservée dans le développement des vertébrés

Soutenue par
Florian CONSTANTY
Le 04 novembre 2022

École doctorale n° 577
Structure et Dynamique des Systèmes Vivants

Spécialité
Biologie moléculaire et cellulaire

Composition du jury :

Dr. Marco, MARCIA EMBL Grenoble	*Président / Rapporteur*
Dr. Claire, ROUGEULLE Épigénétique et destin cellulaire	*Rapporteure*
Dr. Helena, CRUZ DE CARVALHO IBENS	*Examinatrice*
Dr. Pierre-Luc, BARDET IBPS	*Examinateur*
Dr. Alena, SHKUMATAVA Institut Curie	*Directrice de thèse*

institut Curie

Remerciements

Lorsque j'ai commencé cette aventure je n'avais pas idée de quelles découvertes scientifiques je rencontrerais en chemin. Ce périple, de plus de 3 ans, n'a été possible que grâce au soutien important de personnes auquel je suis reconnaissant et que je souhaiterais remercier.

Je souhaite remercier particulièrement Dr. Déborah Bourc'his et Dr. Pierre Léopold pour leurs aides et soutien précieux qui m'ont permis de continuer à travailler dans une situation très spéciale. Je vous remercie de m'avoir écouté et surtout entendu. Merci pour tous ce que vous avez fait afin de faciliter mon travail.

Ainsi que Virginie Bourgeois pour son écoute et ses conseils lorsque j'en avais besoin.

Dr. Silvia Fre, Dr. Raphaël Margueron et Dr. Pedro Hernandez-Cerda pour leur participation dans l'oversight committee et leurs commentaires sans filtre sur le projet.

Dr. Pablo Navarro Gil pour avoir accepté d'être mon tuteur pour ce doctorat ainsi que Dr. Ramesh Pillai et Dr. Filippo Del Bene pour leur participation à mon comité de thèse. Merci d'avoir pris le temps d'écouter mon projet et d'avoir veillé à son bon déroulement.

Dr. Claire Rougeulle, Dr. Marco Marcia, Dr. Helena Cruz de Carvalho et Dr. Pierre-Luc Bardet pour prendre de leur temps et évaluer mon travail de thèse. Merci d'être présent lors de ma défense de thèse ainsi que pour vos suggestions scientifiques.

Dr. Nicolas Servant pour tous ses conseils en bio-informatique. Dr. Michel Wassef et Dr. Daniel Holoch pour les discussions scientifiques très enrichissantes.

Dr. Alena Shkumatava pour m'avoir accepté en thèse et sa relecture de mon manuscrit. Dr. Emmanuel Barillot pour avoir accepté de Co encadrer ma thèse.

L'école doctorale SDSV pour avoir financé mon doctorat lors de ces trois années de thèse.

Je remercie chaleureusement Florent Charton pour sa présence, surtout lorsque j'en avais le plus besoin. Ainsi que mes amis Anaïs, Camille, Manon, Alexis, Aurélien et Emmanuel pour avoir échangé sur la bonne idée que nous avons eue de faire une thèse.

Je souhaite remercier les équipes des laboratoires Bourc'his, Fre et Maître pour leur accueil chaleureux pour les derniers mois de ma thèse.

Je souhaite remercier tous les membres de l'équipe Shkumatava, particulièrement Anna Nawrocka pour tous les moments passés ensemble. Ainsi que toutes les personnes qui ont de près ou de loin aider à l'élaboration de ce projet.

Je souhaite remercier, avec beaucoup d'émotions, mes parents et mon frère qui m'ont toujours soutenu et ont répondu présent au moment le plus important. Si j'ai continué et fini ma thèse c'est aussi grâce à eux. Ainsi que tous les membres de ma famille qui n'ont jamais cessé de croire en moi.

Merci à Armelle, Tarek, Valentin et Olivier pour s'être assuré du bien-être de mes modèles vivants.

Je n'oublie pas M. Jean-Philippe Eymard, ostéopathe, qui a aussi été présent, surtout afin de me remettre sur pied et de continuer ma vie à 400 à l'heure.

Je souhaite remercier toutes les personnes que j'ai croisées et avec qui j'ai pu interagir. Ce sont toutes ces interactions qui font de moi la personne que je suis aujourd'hui.

Table of contents

ABSTRACTS

English version:

Long noncoding RNAs (lncRNAs) have emerged as key regulators of physiology and disease including human cancers. Thus, it is of utmost priority for the field to characterize these transcripts as potential new therapeutic targets. While there is a growing number of annotated lncRNAs, our knowledge of their mode of action is sparse. lncRNAs conserved throughout vertebrates represent a set of putatively functional lncRNAs, making them particularly interesting.

To understand the evolution and the functional contribution of conserved lncRNAs, we studied the cancer-associated *Plasmacytoma variant translocation 1* (*PVT1*) lncRNA. The *PVT1* locus first emerged as a site of translocations in human lymphoma and later has been reported to have oncogenic functions in a considerable number of human cancers. *PVT1* has been found in all vertebrates based on its syntenic genomic position, downstream of the *MYC* protein-coding gene. *PVT1* has been mostly studied in cancers, however, its function *in vivo*, in model organisms, remains unspecified.

Conserved lncRNA functions are often supported by sequence conservation. Mammalian *PVT1* homologous transcripts share alignable sequence conservation. Outside of the mammalian class, the *PVT1* homologous transcripts do not share sequence conservation. Nonetheless, we analyzed the sequence of *PVT1* transcripts from multiple vertebrate species and identified a set of short, deeply conserved RNA motifs. Short RNA motifs emerged as a substitute for conserved sequences to drive conserved molecular functions. Thus, to determine their *in vivo* function, we genetically removed the regions containing these conserved sequence elements from the two zebrafish homologous transcripts. Zebrafish *pvt1* mutants show multiple developmental defects and poor survival to adulthood. The onset of developmental defects is linked to the activation of the p53 pathway and correlated with an upregulation of the myc protein level. The truncated *pvt1* transcripts were also upregulated, concomitant with the identification of myc and p53 binding sites in the vicinity of the zebrafish *pvt1* promoters. Similarly, to the mammalian transcripts, one of the two zebrafish *pvt1* transcripts was upregulated upon DNA damage in a p53-dependent manner. Interestingly, independently of the developmental defects, zebrafish mutants developed anemia, which is consistent with the *Pvt1* transcript mostly expressed in the erythroid lineage during mice development. Using the incPRINT technology, we analyzed proteins that bind specifically to the conserved RNA motifs

of *PVT1* and found that RAD51C and NUSAP1, two essential regulators of the DNA repair pathway, are specific binders to these motifs.

We report the first *in vivo* study of the lncRNA *PVT1* and show that it has a function during development. We found that zebrafish and human *pvt1* transcripts are under the control of the same transcriptional pathways. Overall, a better characterization of *PVT1* function in developmental and physiological processes will help to establish the role of this lncRNA in cancer.

French version:

Les ARNs longs non-codant (lncARN) ont émergé comme des régulateurs clés dans la physiologie et les maladies, incluant les cancers chez l'homme. La caractérisation de ces transcrits comme potentiel cible thérapeutique est donc une priorité. Alors qu'il y a un nombre croissant de lncARN annotés, nos connaissances sur leur mode d'action restent limitées. Les lncARN conservés chez les espèces de vertébrés représentent un groupe d'ARNs potentiellement fonctionnel, ce qui les rend particulièrement intéressants.

Afin de comprendre l'évolution et la contribution fonctionnelle des lncARN conservés, nous étudions l'ARN associé au cancer *Plasmacytoma variant translocation 1* (*PVT1*). Le locus *PVT1* a d'abord été associé à des translocations dans les lymphomes humains puis il a été démontré que *PVT1* a des fonctions d'oncogène dans un nombre considérables de cancers. *PVT1* est conservé dans toutes les espèces de vertébrés basées sur la conservation de position génomique, en aval du gène codant pour la protéine MYC. *PVT1* a été principalement étudié dans le cancer, cependant son rôle *in vivo*, dans un organisme modèle, n'a pas été spécifié.

Les fonctions conservées des lncRNA sont souvent accompagnées d'une conservation de séquence. Les transcrits de mammifères homologues de *PVT1* partagent une conservation de séquence alignable. En dehors de la classe des mammifères, les transcrits homologues de *PVT1* ne partagent pas de conservation de séquence. Néanmoins, nous avons analysé la séquence des transcrits de *PVT1* de plusieurs espèces de vertébrés et avons identifié un ensemble de motifs d'ARN courts, profondément conservés. Les motifs d'ARN courts sont apparus comme un substitut aux séquences conservées pour guider les fonctions moléculaires conservées. Pour déterminer leur fonction *in vivo*, nous avons retiré génétiquement les régions contenant ces éléments de séquence conservés des deux transcrits homologues du poisson zèbre. Les mutants *pvt1* de poisson zèbre présentent de multiples défauts de développement et une faible survie à l'âge adulte. L'apparition des défauts de développement est liée à l'activation de la voie p53 et corrélée à une régulation à la hausse du niveau de la protéine myc. Les transcrits tronqués de *pvt1* sont également régulés à la hausse, ce qui est concomitant avec l'identification de sites de liaison de myc et p53 à proximité des promoteurs de *pvt1* chez le poisson zèbre. De la même manière que pour les transcrits de mammifères, l'un des deux transcrits *pvt1* du poisson zèbre est surexprimé lors de dommages à l'ADN d'une manière dépendante de p53. De manière intéressante, indépendamment des défauts de développement, les mutants du poisson zèbre ont développé une anémie, ce qui est cohérent avec le transcrit *Pvt1* principalement exprimés dans

la lignée érythroïde pendant le développement de la souris. En utilisant la technologie incPRINT, nous avons analysé les protéines qui se lient spécifiquement aux motifs ARN conservés de *PVT1* et avons découvert que RAD51C et NUSAP1, deux régulateurs essentiels de la voie de réparation de l'ADN, se lie spécifiquement à ces motifs.

Nous rapportons la première étude *in vivo* du lncARN *PVT1* et montrons qu'il a une fonction au cours du développement. Nous avons constaté que les transcrits *pvt1* du poisson zèbre et de l'homme sont sous le contrôle des mêmes voies transcriptionnelles. Globalement, une meilleure caractérisation de la fonction de *PVT1* dans les processus développementaux et physiologiques aidera à établir le rôle de ce lncARN dans le cancer.

1. GENERAL INTRODUCTION

1.1 Diving in the human transcriptome: the RNA molecule in the flow of genetic information

1.1.1 "Junk" DNA

On April 25th, 1953, Watson and Crick published the molecular structure of DNA (Watson & Crick, 1953), helped by the work of Franklin. This revolution in genetics opened the path to understanding the flow of genetic information in a biological unit. Four years later, in September 1957, Francis Crick gave a lecture in which he outlined the principles of what became the Central Dogma of Molecular Biology. The Dogma pushes further the model of the flow of genetic information (Fig 1A) (Cobb, 2017; Crick, 1970). This dogma, commonly abbreviated as "DNA makes RNA, and RNA makes proteins", is a framework to understand the transfer of information from DNA to protein. The proteins were, back then, thought to be the principal functional unit of the cell. DNA is therefore the carrier of the genetic information, which is translated into RNA, then the RNA molecule is used as a template for protein synthesis. Interestingly, contrary to general knowledge, Crick already suggested the possibility of RNA replication, RNA driving DNA polymerization, and DNA being directly the template for protein synthesis. Crick named them "special transfers" (Fig 1A) (Crick, 1970).

In subsequent years, much effort has been made to elucidate the composition of the type of RNA molecules involved in the flow of information. Given its central role, messenger RNAs (mRNAs) have been largely studied since their discovery in 1961 (Jacob & Monod, 1961). Alongside, non-coding RNAs (ncRNAs) were discovered, and their cellular roles clarified. Such as ribosomal RNAs (rRNAs), first as part of the microsome (Claude, 1939), then as part of the ribosome (Palade, 1955), transfer RNAs (tRNAs) (Hoagland et al., 1958), and small nuclear RNAs (snRNAs) (Hadjiolov et al., 1966; Weinberg & Penman, 1968). The remaining parts of the genome were then so-called "junk" DNA (Ohno, 1972), as these DNA fragments were thought to play no role in the fitness of the species.

It was not until the 1980s / l990s that a more complex view of the transcriptome appeared. With the discovery of the first regulatory ncRNA in *E. coli* (Coleman et al., 1984), alongside the first identification of a non-coding locus, the *PVT1* locus (Graham & Adams, 1986), the first functional human long non-coding RNA (lncRNA) *H19* (Brannan et al., 1990),

together with the discovery of the lncRNA *Xist* (Brockdorff et al., 1991, 1992; Brown et al., 1992) and microRNAs (miRNAs) in *C. elegans* (Lee et al., 1993). These discoveries paved the way for a more complex role of the RNA molecule.

1.1.2 The human genome sequencing and annotation

The last two decades set the start of an explosive expansion in the understanding of the biological functions of these ncRNAs, alongside considerable improvement in technology. The sequencing of the human project took 13 years before the first publication of an "almost" complete genome (Lander et al., 2001; Venter et al., 2001). With the advent of long-read sequencing, the human genome was completed this year (Nurk et al., 2022), even if the Y chromosome remains to be finished. Ultimately, with the emergence of RNA sequencing, large consortiums (ENCODE and FANTOM) multiplied the efforts to enable a comprehensive perception of the mammalian genome and transcriptome (Carninci et al., 2005; Katayama et al., 2004; The ENCODE Project Consortium, 2007). These studies revealed that 80% of the genome is transcribed into RNAs whereas only 1% of the genome is transcribed as mRNA (Fig 1B) (Djebali et al., 2012). Even if a part of the genome can be supposedly transcribed due to transcriptional noise, this cannot explain the majority of the ncRNA transcription (Ponjavic et al., 2007). Moreover, the study of the human transcriptome identified between 20 000 and 60 000 potentially functional lncRNAs (Hon et al., 2017; Iyer et al., 2015), whereas 20 000 to 25 000 protein-coding genes were identified (Lander et al., 2001). These analyses revealed the high importance of ncRNAs in the human genome.

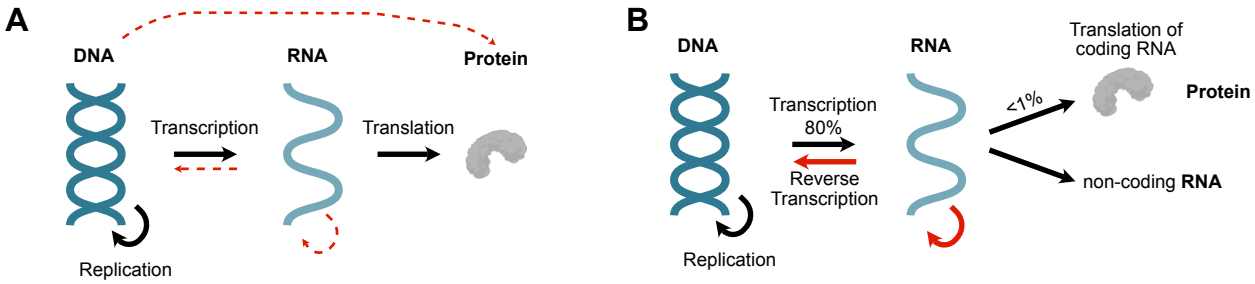


Figure 1: The modern view of the Central Dogma of Molecular Biology

A. Original Central Dogma adapted from Crick, 1970. The dotted red arrows represent the "special transfer", and the plain black arrows represent the general transfers. **B.** Modern view of the Central Dogma. The plain red arrows represent the "special transfers" with experimental proof.

1.1.3 Non-coding RNAs

As previously described, the human transcriptome is composed of coding RNAs (mRNAs) and ncRNAs. The ncRNAs are essential in the fine-tuning of gene expression and protein functions. They are the majority of RNAs in the transcriptome (Djebali et al., 2012). Interestingly, among the ncRNAs, we can find several subclasses: the housekeeping and the regulatory RNAs (Fig 2). The housekeeping RNA subclass is composed of RNAs essential for the maintenance of basic cell function. These RNAs are usually constitutively expressed. In contrast, regulatory RNA expression is spatiotemporally controlled.

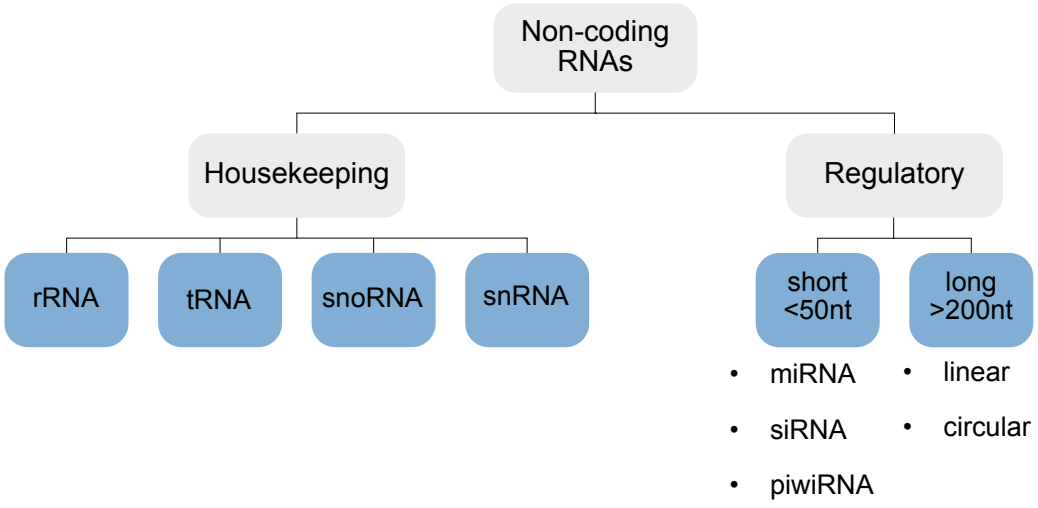


Figure 2: Non-coding RNAs classification

Schematic depicting the classification of ncRNAs into subclasses according to the function (housekeeping, and regulatory) and the size of the ncRNAs.

Housekeeping ncRNAs

Small nuclear RNAs (snRNAs) (Hadjiolov et al., 1966), are critical components of the spliceosome, catalyzing the splicing of RNA polymerase II transcripts. snRNAs associate with proteins to form RNA-protein complexes (snRNP) in the nucleus. snRNAs are often divided into two distinct groups with the Sm-class snRNAs (composed of U1, U2, U4, U4atac, U5, U7, U11, and U12) transcribed from the RNA polymerase II and processed in the cytoplasm before translocating into the nucleus. The Lsm-class snRNAs (U6 and U6atac) are transcribed by the RNA polymerase III and confined to the nucleus (Kiss, 2004). The snRNAs interact with consensus sequences on pre-RNAs by base pairing, and assist splicing events (Jurica & Moore, 2003; Morais et al., 2021). Remarkably, there are two distinct spliceosome apparatuses with the major spliceosome composed of the snRNAs U1, U2, U4, U5, U6, and over 150 proteins (Akinyi & Frilander, 2021). The minor spliceosome (much less common, ~0.5%) is composed of the U11, U12, U4atac, U5, and U6atac snRNAs and proteins (Akinyi & Frilander, 2021; Turunen et al., 2013). The respective intron types spliced are called U2-type introns or U12-type introns. The U2-type introns have GT-AG splice sites, and the U12-type introns have AT-AC splice sites (Akinyi & Frilander, 2021; Turunen et al., 2013). Intriguingly, recent work showed the non-canonical function of the U1 snRNP involved in the chromatin retention of spliced RNAs by interacting with nuclear retention factors (Azam et al., 2019; Yin et al., 2020).

Ribosomal RNAs are the most abundant form of RNA in the transcriptome, accounting for approximately 80% of cellular RNAs. rRNAs are key components of the ribosome. It makes up 60% of the ribosome, complemented by 40% of proteins. They are essential for ribosome assembly and function, serving as ribosomal protein binding sites and contributing to the binding of extra-ribosomal factors. These assemblies result in the protein translation machinery in eukaryotes and procaryotes (Fig 3A) (Noller et al., 2017; Simsek et al., 2017). Ribosome biogenesis in eukaryotes begins in the nucleolus where three of the rRNAs (18S, 5.8S, and 28S) are transcribed by the RNA polymerase I as a long primary transcript. This transcript originates from tandem repeats of ribosomal DNA genes (Grandi et al., 2002; Henras et al., 2015; Sakai et al., 1995). In contrast, the 5S rRNA is transcribed by the RNA polymerase III from independent genes not localized in the nucleolus (Ciganda & Williams, 2011). Following rRNA maturation, they interact with specific factors of the large ribosome subunit, named 60s Ribosomal Protein L (RPL), or the small ribosome subunit, named 40s Ribosomal Protein S (RPS). Interestingly, the 5S rRNA interacts with the RPL5 and RPL11 factors, forming the 5S ribonucleoprotein particle (RNP) (Sloan et al., 2013). In conditions of ribosome biogenesis

impairment, the 5S RNP interacts directly with the Mouse Double Minute 2 (MDM2) protein, a direct inhibitor of the Tumor protein 53 (TP53 / P53). This drives the activation of the P53 protein and limits proliferation (Sloan et al., 2013). Thus, rRNAs are essential for ribosomal activity and are directly involved in stress signaling pathways to control the precise flow of genetic information.

Small nucleolar RNA (snoRNA) is a family of ncRNAs composed of C/D box snoRNAs and H/ACA snoRNAs (Fig 3B) (Balakin et al., 1996; Maxwell & Fournier, 1995). These two classes of snoRNAs are defined by the boxes harbored on the transcript (Balakin et al., 1996; Maxwell & Fournier, 1995). One well-studied function of the snoRNAs is their involvement in rRNA maturation, modification, and stabilization. The C/D snoRNAs are involved in the specific 2'-O-methylation of ribose (Kiss-László et al., 1996; Nicoloso et al., 1996; Tycowski et al., 1996), whereas the H/ACA snoRNAs catalyze the isomerization of uridine to pseudouridine (ψ) (Ganot et al., 1997; Ni et al., 1997) (Fig 3B). Most of mammalian snoRNAs are localized in the introns of other genes (Kishore et al., 2013). These snoRNAs are excised from the pre-mRNA by exonucleolytic trimming and bound by proteins to protect them from degradation (Kishore et al., 2013).

Finally, tRNAs act as the link between nucleic acid and amino acids during protein synthesis. They carry an amino acid on their 3' end and interact, in the ribosome, with the mRNA by base pairing (Fig 3C) (Rich & RajBhandary, 1976). tRNAs represent the most abundant gene family in the human genome, with about 500 tRNAs and tRNA-like genes. These genes are transcribed by the RNA polymerase III (Abe et al., 2014; Parisien et al., 2013). Interestingly, while tRNAs are constitutively expressed, their cellular composition is tissue-specific (Dittmar et al., 2006; Ishimura et al., 2014). This highlights the complexity of tRNA regulation in the control of mRNA translation. Moreover, base modifications expand the complexity of tRNAs and are prominent in local environment adaptation to control tRNAs functionality (Kirchner & Ignatova, 2015; Schimmel, 2018). Stress conditions can induce the fragmentation of tRNAs. While these induced tRNA cleavages were first considered as translational inhibition mechanisms (Levitz et al., 1990), tRNAs fragments were found in all domains of life where they act as signaling factors in the stress response (Haiser et al., 2008; Hsieh et al., 2010; Jöchl et al., 2008; Thompson et al., 2008; Wang et al., 2013). These fragments are processed from tRNAs through different nucleases, driving the accumulation of different types of fragments (Xie et al., 2020). tRNAs fragments are directly involved in translation inhibition by (1)

interacting with key translational factors such as the small ribosome subunit (Gebetsberger et al., 2017), or (2) forming G-quadruplex on mRNA, blocking the translational initiation factor complex (Akiyama et al., 2020; Lyons et al., 2017). Other fragments can directly regulate mRNA expression by acting as miRNA-like and interacting with Argonaute proteins to silence gene expression, including transposable elements (Gagnon & Corey, 2012; Sharma et al., 2016). tRNA fragments can also act epigenetically, a 5' tRNA halve ranging from 30 to 34 nucleotides (nt) was found in the intergenerational inheritance of metabolic disorders (Chen et al., 2016). A more extensive view of tRNA fragment actions is described here Xie et al., 2020.

These non-exhaustive examples of housekeeping ncRNAs highlight the essential role of ncRNAs in cellular functions. These RNAs are vital for the correct functions of the cell. They act in pre-mRNA maturation and protein synthesis. They are therefore essential to facilitate the transfer of genetic information from genes to proteins.

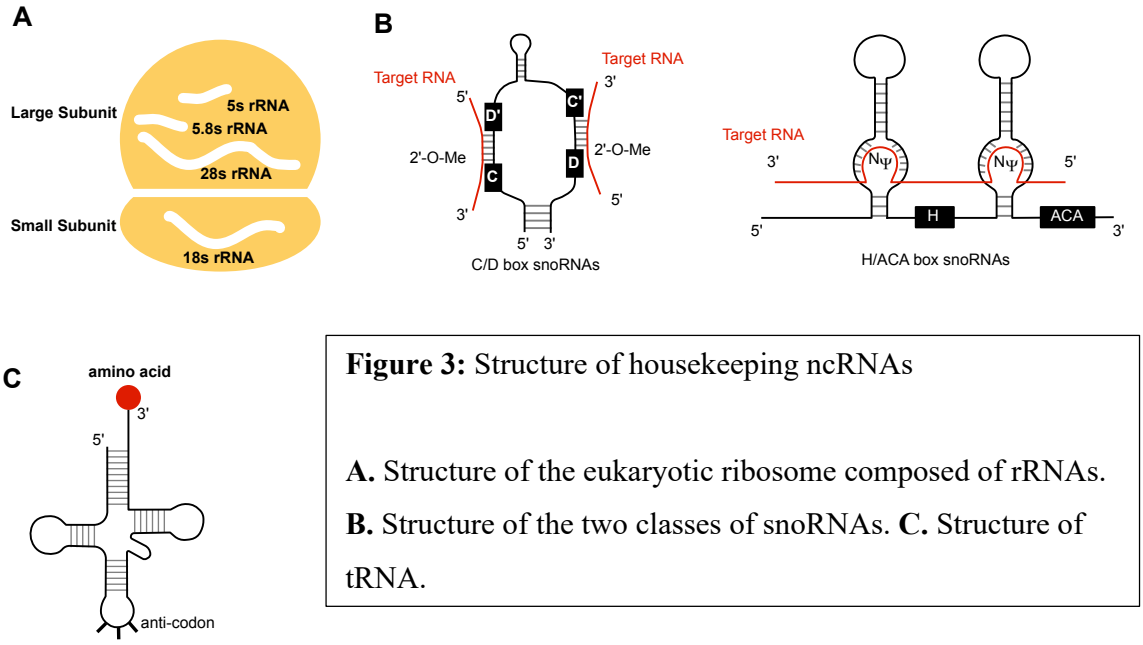


Figure 3: Structure of housekeeping ncRNAs

A. Structure of the eukaryotic ribosome composed of rRNAs. **B.** Structure of the two classes of snoRNAs. **C.** Structure of tRNA.

Regulatory ncRNAs

A tremendous effort has been put into the characterization and understanding of the regulatory ncRNAs. These RNAs are of utmost interest, especially for human health as evidenced by the large number of published reports linking miRNAs, and lncRNAs to human diseases such as cancers (Bhan et al., 2017; Calin & Croce, 2006; Carlevaro-Fita et al., 2020). Regulatory RNAs

are classified into two categories depending on their size: small (<50 nt) and long (>200 nt) ncRNAs (Fig 2).

miRNAs are small ncRNAs (19-24 nt), endogenous, and evolutionarily conserved. These RNAs associate with Argonaute (AGO) proteins, forming the RNA-Induced Silencing Complex (RISC), and act as post-transcription regulators of gene expression (Bartel, 2018; Lee et al., 1993). miRNAs bind by base complementarity to their RNA targets and interfere with the translational machinery. In animal cells, most miRNA-RNA interactions are not complementary along the entire length of the miRNA. The pairing of the nucleotides 2-8 of the miRNAs, called seed, is critical for the stabilization of the RISC complex (Bartel, 2009; Chandradoss et al., 2015; Elkayam et al., 2012; Jo et al., 2015; Nakanishi et al., 2012; Salomon et al., 2015; Schirle & MacRae, 2012). Moreover, 3' sequences of the miRNA are also essential, to a lesser extent, for the stability of the miRNA-RNA interactions (Broughton et al., 2016). Because of their partial complementarity to their target, miRNAs regulate the expression of a multitude of targets. Following the binding of the RISC complex to the targeted RNA, the formation of the silencing complex begins with the recruitment of the GW182 family of proteins (Fabian & Sonenberg, 2012; Huntzinger & Izaurralde, 2011). These proteins act as molecular bridges between the RISC complex and effectors such as the deadenylase complex (Fabian & Sonenberg, 2012; Huntzinger & Izaurralde, 2011). Therefore, miRNA induces the destabilization of the targeted RNA by deadenylation, decapping, and degradation from the 5' part of the RNA, resulting in decreased gene expression (Alberti & Cochella, 2017; Jonas & Izaurralde, 2015). miRNA functions highlight a perfect example of RNA regulation of the flow of genetic information. The tissue-specific expression of miRNAs allows local regulation of gene expression (Ludwig et al., 2016). Intriguingly, independent studies showed that miRNA stability can be impacted by the targeted RNA. This process, named Target Directed miRNA Degradation (TDMD), is driven by the near-perfect complementarity of the miRNA to its target RNA (Ameres et al., 2010; Bitetti et al., 2018; Kleaveland et al., 2018; la Mata et al., 2015; Marcinowski et al., 2012; Ulitsky et al., 2011). Remarkably, two independent studies showed that the ubiquitin ligase Zinc Finger SWIM-Type Containing 8 (ZSWIM8) drives TDMD, probably by directing proteasomal degradation of miRNA Argonaute complexes (Han et al., 2020; Shi et al., 2020). Therefore, miRNAs regulate the expression of mRNAs and lncRNAs, whereas a near-perfect complementarity of the miRNA to its target will trigger the degradation of the miRNA, adding a new layer of regulation.

Small interfering RNAs (siRNAs) (21-23 nt) were discovered in plants (Hamilton & Baulcombe, 1999). These RNAs have almost the same function as miRNAs, except that they have perfect complementarity to their target RNAs. This complementarity induces the cleavage of the targeted RNAs. The cleavage is performed by the catalytically active AGO2 protein in humans (Matranga et al., 2005). siRNAs are endogenous in plants but are used exogenously in mammals to induce RNA interference (RNAi) (Elbashir et al., 2001). Because of their perfect complementarity, siRNAs can regulate gene expressions of a single target. A few endo-siRNA have been identified in a variety of metazoan species (Okamura & Lai, 2008). Endo-siRNAs are derived from endogenous RNA precursors, processed by Dicer, and combined specifically with the AGO2 protein to cleave targeted RNAs. This class of RNA is mostly studied in *C. elegans*, drosophila, and mice (Okamura & Lai, 2008). The described endo-siRNAs act mostly by silencing transposable elements in the gonads and as an anti-viral response in stem cells (Claycomb, 2014).

Piwi-interacting RNAs (piRNAs) are another class of small ncRNAs (24-31 nt) (Aravin et al., 2006; Aravin et al., 2003; Girard et al., 2006; Grivna et al., 2006; Lau et al., 2006; Vagin et al., 2006). piRNAs interact with PIWI proteins, a germline abundant subclass of the Argonaute proteins to induce gene silencing. piRNAs are produced mainly from intergenic regions termed piRNAs clusters in a dicer-independent mechanism (Vagin et al., 2006). These regions are principally composed of transposable elements. Thus, piRNAs act, by base complementary, to repress transposable element expression and activity. Some piRNA clusters can be single-stranded and are transcribed as precursors processed by the endonuclease Zucchini, located at the surface of the mitochondria (Ipsaro et al., 2012; Nishimasu et al., 2012; Pane et al., 2007). The processed fragments are then loaded into the PIWI protein and translocated into the nucleus. Other clusters are dual-stranded, thus transcribed in both senses. In this case, two precursors are expressed and processed by Zucchini. Some processed piRNAs will be loaded into the PIWI protein, as previously described, and others will be loaded into the Aubergine protein. This last protein acts in a complementary fashion with the AGO3 protein to cleave sense and anti-sense transposon transcripts in the cytoplasm. This last biogenesis pathway, also known as the ping-pong pathway, is auto-sufficient, maintaining the production of piRNAs from expressed transposons (Czech & Hannon, 2016). After the nuclear translocation of the piRNA/PIWI complex, piRNAs direct the formation of heterochromatin by inducing the tri-methylation of the histone 3 lysine 9 (H3K9me3) at the targeted transposon loci (le Thomas et

al., 2013). Thereby, piRNAs regulate transcriptionally and post-transcriptionally transposon expressions and are the guardian of the genome during gametogenesis.

At last, the regulatory RNA group is also composed of lncRNAs. These transcripts are at least 200 nt long and transcribed by RNA polymerase II. LncRNAs are of great interest because of their functions. Short ncRNAs are specialized in the regulation of gene expression, and thus in the fine regulation of genetic information flow (Bartel, 2009). In contrast, lncRNAs have diverse molecular functions and are direct effectors of this genetic information (Constanty & Shkumatava, 2021; Hansen et al., 2013; Kopp & Mendell, 2018; Kristensen et al., 2019; Marchese et al., 2017; Memczak et al., 2013; Statello et al., 2021). This last group is the largest, consisting of different types of lncRNAs depending on their structures. There are two distinct structures of lncRNAs: circular and linear RNAs.

The circular lncRNAs (circRNAs) are produced from noncanonical splicing events, named backsplicing (the mechanism is reviewed here Kristensen et al., 2019). The first identified circRNAs were the viroids, discovered in 1976 (Sanger et al., 1976). However, viroids are not formed through backsplicing. The first mammalian circRNA identified to have a possible function was from the sex-determining region (*SRY*) in mouse testis (Capel et al., 1993). Then, different algorithms have identified circRNAs in all levels of eukaryotic life (Ivanov et al., 2015; Jeck et al., 2013; Salzman et al., 2013; Wang et al., 2014; Westholm et al., 2014). CircRNAs are expressed in a tissue-specific manner (Maass et al., 2017; Salzman et al., 2013; Xia et al., 2016). Because of their circular form, circRNAs lack cap, and polyadenylated tail. They are globally located in the cytoplasm (Hsu & Coca-Prados, 1979; Jeck et al., 2013; Memczak et al., 2013; Salzman et al., 2012). An important fraction of studied circRNAs has been associated with miRNA interactions (Hansen et al., 2013; Memczak et al., 2013; Piwecka et al., 2017; Zheng et al., 2016). The most studied example is the interaction between the circRNA *Cerebellum Degeneration-Related antigen 1 antisense* (*CDR1as)* and miR-7 (Piwecka et al., 2017; Xu et al., 2015; Yu et al., 2016). Interestingly, *CDR1as* is a brain-enriched transcript harboring more than 70 binding sites of the miR-7 (Piwecka et al., 2017). In the brain, *CDR1as* acts as a positive regulator of the miR-7. Knock-out of *CDR1as* leads to the up-regulation of miR-7 target genes, and robust sensorimotor deficits phenotype (Piwecka et al., 2017). The interaction between *CDR1as*/miR-7 is a crucial regulator of synaptic transmission. It stabilizes or facilitates the transport of the miR-7 in neurons (Piwecka et al., 2017). Other specific functions of circRNAs have been demonstrated, with circRNAs interacting with proteins (Dudekula et al., 2016). These interactions can lead to the sponging

of the proteins (Abdelmohsen et al., 2017; Ashwal-Fluss et al., 2014), enhance proteins functions (Li et al., 2015; Y. Zhang et al., 2013), mediate enzymes/substrates complex formations (Du et al., 2016, 2017; Zeng et al., 2017) or allow protein localization in specific loci in the nucleus (Chen et al., 2018). Thus, circRNAs became essential regulators of a wide range of biological processes through a diversity of modes of action.

Linear lncRNAs are RNA polymerase II transcripts, capped, polyadenylated, and spliced (Cabili et al., 2011; Guttman & Rinn, 2012). Interestingly, some lncRNAs are not spliced nor polyadenylated, such as the *Metastasis Associated Lung Transcript 1* (*MALAT1*) and the *Nuclear Paraspeckle Assembly Transcript 1 (NEAT1)*. These RNAs are processed through an unusual Ribonuclease (RNase) P cleavage directed by the tRNA-like structure of the 3' part of these transcripts (Wilusz et al., 2012). Similar to the circRNAs expression pattern, early studies of linear lncRNAs revealed a specific expression pattern in specific tissues, cell types, and subcellular compartments (Mercer et al., 2008). Intriguingly, comparative studies have shown that linear lncRNAs are generally lowly expressed but highly tissue and cell-type specific whereas mRNAs are highly expressed but lowly tissue or cell-type specific (Cabili et al., 2011; Derrien et al., 2012; Ulitsky et al., 2011). Contrary to the circular form, linear lncRNAs are predominantly nuclear. Linear lncRNAs features and functions will be the subject of the next chapter.

Over the years, the view of the human, and more generally metazoan, transcriptome has become more complex. The discovery of canonical and non-canonical functions of housekeeping ncRNAs and the diversity of regulatory ncRNAs illustrate the great importance of ncRNAs as molecules for the proper functioning of the cell and the organism. Furthermore, this is supported by the association of ncRNA misregulation with human diseases (Esteller, 2011; Lekka & Hall, 2018).

1.2 Linear long non-coding RNAs

1.2.1 Coding versus non-coding transcripts

Linear lncRNAs (hereafter referred as lncRNAs) are a complex class of ncRNAs. The main characteristic of lncRNAs is their inability to encode a protein or a functional protein. However, some transcripts that were previously described as lncRNA were recently shown to encode

functional proteins. These observations open the door to debate about the ncRNA nomenclature.

The accurate identification of lncRNAs is essential, as is precise identification between coding and non-coding transcripts. LncRNAs share characteristics with mRNAs but do have unique features that allow them to be identified. Coding potential is the obvious feature that comes to mind. If a transcript does not contain an open reading frame (ORF), it is a non-coding transcript. However, it was fast acknowledged that this feature alone was not stringent enough. The average size of lncRNA in vertebrates is ~1000 nt, so it is expected that it will by chance contain one or several ORFs (Dinger et al., 2008). With the accumulation of annotated lncRNAs in a growing number of species, new characteristics have been identified to separate coding from non-coding transcripts, first bioinformatically and then experimentally. Comparisons between mRNA and lncRNA sequences showed that lncRNA sequence evolve faster than protein-coding genes (Ulitsky & Bartel, 2013). Therefore, the general low sequence conservation can be used to separate them (Sun et al., 2013). An additional attribute is the nucleotide composition. Nucleotide frequencies in mRNA ORFs are dictated by non-random codon usage (Grantham et al., 1980). Hence, nucleotide frequencies in putative ORFs can be an indicator of transcript coding potential. While this method effectively identifies potentially coding ORFs, its application is limited to small ORFs, which are the main types of ORFs found on lncRNA transcripts. The first algorithms developed to annotate lncRNAs rely on the aforementioned features to distinguish coding from non-coding transcripts (Kong et al., 2007; Liu et al., 2006; Wang et al., 2013). Subsequently, additional layers were used to discriminate transcripts such as sequence similarity to known proteins, sequence similarities to known functional protein domains, and RNA structure (Housman & Ulitsky, 2016). As with the protein-coding gene identification, lncRNA annotation in species can also be performed by genomic comparison (Hezroni et al., 2015).

Interestingly, experimental tools have been used to evaluate the coding potential of lncRNAs. Several studies have used the ribosome profiling method to investigate the ribosome/lncRNA interactions to extract information about the coding potential of transcripts. These studies first reported that the ribosome binds to mRNAs as well as lncRNAs. Ribosome occupancy is therefore not a unique feature of mRNAs (Chew et al., 2013; Guttman et al., 2013). However, ribosome releases can be used to differentiate between coding and non-coding potential. Indeed, protein-coding genes show ribosome release at the stop codon, whereas lncRNAs show no evidence of ribosome release for any known ORFs (Guttman et al., 2013).

In parallel, ribosomal profiling of lncRNAs during zebrafish development showed that lncRNAs interact with the ribosome but with different profiles than mRNA ORFs (Chew et al., 2013). Thus, computational and experimental strategies are complementary in assessing the coding potential of transcripts. To this end, some algorithms identify lncRNAs by integrating sequence-derived and experiment-based lncRNA features (Hu et al., 2017).

Bifunctional RNAs

It is important to note that lncRNAs can harbor small ORFs. Some lncRNAs can have therefore both coding and non-coding functions (Dinger et al., 2008). Consequently, lncRNAs are considered to be a new source of micropeptides (Bazzini et al., 2014; Housman & Ulitsky, 2016; Ruiz-Orera et al., 2014). Although these micropeptides were initially considered the result of pervasive translations, resulting in not functional micropeptides (Housman & Ulitsky, 2016), an increasing number of lncRNAs have been shown encoding functional peptides. For instance, the steroid receptor RNA activator protein (SRAP) from the lncRNA *SRA* (Chooniedass-Kothari et al., 2004), the 17 amino-acid micropeptide from the lncRNA *MIR155HG* (Niu et al., 2020), the 87 amino-acid micropeptide from the circular form of the *LINC-PINT* (Zhang et al., 2018), and the 48 amino-acid micropeptide of *TUNAR/MEGAMIND* (Li et al., 2021; Senís et al., 2021) (for a more extensive view, see Li & Liu, 2019). Therefore, experimental validations are required to determine whether the translated proteins by lncRNAs are functional. Intriguingly, some mRNAs have also been found to act as non-coding molecules in bacteria, plants, and animals (Candeias et al., 2008; Kloc et al., 2005; Kumari & Sampath, 2015; Sousa et al., 2001). With the most interesting example being the role of the *P53* mRNA in the regulation of the MDM2 protein activity (Candeias et al., 2008). The MDM2 protein directly interacts with the *P53* mRNA, this impairs the binding of the MDM2 protein to the P53 protein, leading to the activation of P53 (Candeias et al., 2008). Intriguingly, this interaction also leads to the translation of the *P53* mRNA (Candeias et al., 2008). These examples are still limited, but we cannot ignore their existence and the possibility that other examples will arise with an improvement in technical efficiency. These discoveries question the classical nomenclature of coding and non-coding RNAs. The RNAs classification could therefore be modified in the coming years, for example by separating non-functional and functional RNAs. The former being mRNAs with no obvious regulatory functions as an RNA molecule and the latter composed of RNA molecules with regulatory functions. The latter group can be further subdivided into protein-coding transcripts, micropeptide coding, or without evidence of functional peptides.

1.2.2 Genomic location

LncRNA is a highly heterogeneous class of ncRNAs. As the classification of RNA molecules has contributed to a better understanding of the different ncRNAs in metazoan transcriptomes, early studies attempted to classify the different lncRNAs by exploiting their genomic location (Fig 4). LncRNAs originate from an array of genomic locations. The first class is composed of lncRNAs that do not overlap protein-coding gene sequences, such as intergenic lncRNAs (Fig 4A). These RNAs are autonomously transcribed. This is the most studied subclass of lncRNA because their functional analysis is facilitated as mutations within the lncRNA locus do not affect the sequence of protein-coding genes. Therefore, the impact of the genomic mutations can be unequivocally assigned. Interestingly, intergenic lncRNAs are mostly enriched near transcription factors and developmental genes (Amaral et al., 2018; Herriges et al., 2014; Necsulea et al., 2014; Ulitsky et al., 2011). Moreover, studies in stem cell and progenitor cells showed an important number of intergenic lncRNAs specifically expressed in these cell types (Guttman et al., 2009). This suggests that intergenic lncRNAs are important regulators of development in a tissue, cell type-specific manner. Another subclass is composed of lncRNAs in the vicinity of protein-coding gene but transcribed from the opposite direction (Fig 4B). Bidirectional lncRNAs differ from intergenic lncRNA in that they share the same promoter as the protein-coding gene, thus exhibiting similar expression regulation. Intriguingly, lncRNAs can as well originate from the intron of protein-coding genes (Fig 4C), without overlapping the sequence of the host gene. These lncRNAs are named intronic, as the *MEGAMIND* lncRNA (Ulitsky et al., 2011).

Some lncRNA transcripts overlap partially or completely the sequence of protein-coding genes in sense or antisense (Fig 4D). These lncRNAs are particularly challenging to study as genomic manipulation of the loci might impact sequences of the protein-coding gene. These transcripts are widely present as 25-40% of protein-coding genes are estimated to be transcribed in antisense (Engström et al., 2006; Katayama et al., 2005). Due to the concordant or discordant expression between the sense and antisense transcripts (Katayama et al., 2005), the natural antisense transcripts (NAT) emerged as new therapeutical targets (Schwartz et al., 2008; Wahlestedt, 2006, 2013). In the case of discordant lncRNA, knock-down of their expression can lead to the upregulation of the protein-coding gene (Katayama et al., 2005; Wahlestedt, 2006). On the contrary, the knock-down of concordant lncRNAs can lead to the reduction of the protein-coding gene expression (Wahlestedt, 2006). Thus, protein-coding gene expression can be selectively regulated in human diseases with lncRNA-targeted strategies.

Recently, a new class of lncRNA emerged, originating from enhancers, named eRNAs (de Santa et al., 2010; Kim et al., 2010) (Fig 4E). eRNAs are important epigenetic factors regulating chromatin accessibility, stabilizing enhancer-promoter loop in *cis*, and regulating chromatin landscape in trans (Sartorelli & Lauberth, 2020). These transcripts are associated to a rather low stability and a bidirectional transcription.

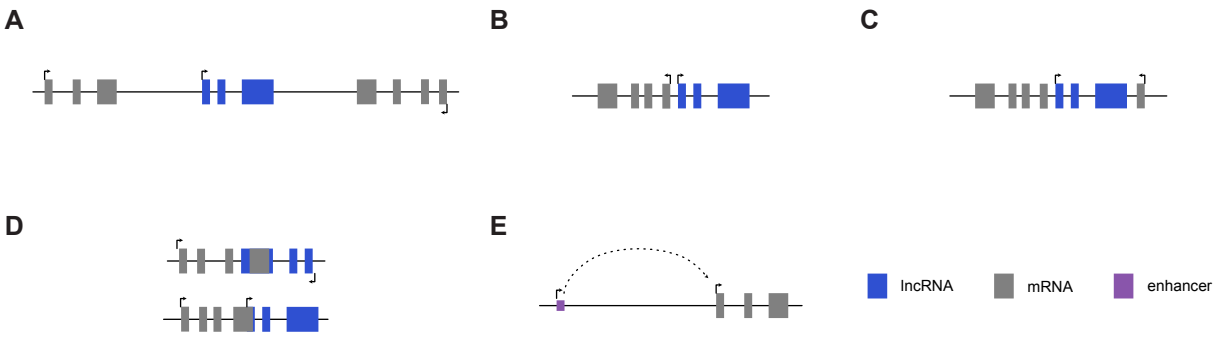


Figure 4: Genomic localisation of lncRNAs

Schematic representation of the genomic localisation of lncRNAs. In grey are represented mRNAs and in blue the lncRNA. **A.** Intergenic lncRNA. **B.** Bidirectional lncRNA **C.** Intronic lncRNA. **D.** Overlapping antisense (NAT) (top) and sense (bottom) lncRNA. **E.** Enhancer lncRNA (eRNA), the dotted line represents the action of the enhancer on the mRNA promoter.

1.2.3 LncRNA evolution

Origin of lncRNAs

LncRNA evolution is particularly remarkable. While the majority of human lncRNAs are considered humans specific (Ruan et al., 2020), it was determined, by genomic position conservation, that hundreds of lncRNA have homologs in vertebrates. Only 18 lncRNAs have been found conserved from humans to the sea urchin (Hezroni et al., 2015). The observation that most lncRNAs are poorly conserved outside the vertebrate subphylum and most of them are specie-specific shows that new lncRNA originate dynamically. Several mechanisms are at the origin of lncRNA appearance. Pseudogenization of protein-coding genes, in other words, loss of coding potential of the protein-coding gene, can lead to new lncRNAs. Mutations within protein-coding genes can drive the loss of the protein, if the transcript is still expressed and not destabilized by the mutation, then it can lead to new lncRNAs. Intriguingly, the *Xist* lncRNA

evolved as an ncRNA by pseudogenization (Duret et al., 2006). Transposable elements (TEs) are also driving the formation of new lncRNAs. TEs are strongly involved in genome innovations in mammals (Cordaux & Batzer, 2009). TEs insertions containing functional promoters can drive the expression of new lncRNAs. This is sustained as 80% of lncRNAs overlap at least one TE (Kapusta et al., 2013). TE insertions can also lead to new lncRNAs without insertions of functional promoters. Studies showed that transcriptions often occur in divergent directions from protein-coding genes and enhancers (Seila et al., 2008). However, the products of these transcriptions are rapidly degraded. TE insertions can stabilize these cryptic transcripts leading to the appearance of new stabilized transcripts coding or non-coding (Gotea et al., 2013; Wu & Sharp, 2013). Remarkably, the annotation of coding or non-coding RNAs can change throughout evolution. Transcripts can arise as lncRNA in basal vertebrates and evolve into protein-coding in mammals, as the libra/Nrep transcript, non-coding in zebrafish, and protein-coding in mammals (Bitetti et al., 2018).

Conservation of lncRNAs in vertebrates

Shortly after the explosive dynamic of lncRNA identification in different species, studies of genomic comparison arose to achieve three goals: (1) to annotate and identify lncRNAs in vertebrate species; (2) to elucidate the evolution of these transcripts; (3) to discriminate from the pervasively transcribed genes using the evolutionarily conserved lncRNAs as an indicator of the functional significance. Annotation of conserved lncRNAs in vertebrate genomes was challenging as lncRNA evolution differs from protein-coding genes. The canonical view of gene conservation is based on sequence similarities. While useful for the identification of conserved protein-coding genes, this view turned out to not be fully applicable for the identification of lncRNAs. It became clear that lncRNA sequences are poorly conserved across species (Wang et al., 2004). Nevertheless, while lncRNAs sequence conservation is lower than protein-coding genes, it is higher than intron or intergenic regions (Cabili et al., 2011; Guttman et al., 2009; Ponjavic et al., 2007), showing their putative role in biological processes. Only a few mammalian lncRNAs have been found to align with zebrafish lncRNAs but shared only patches of sequence conservation (Fig 5A) (Hezroni et al., 2015; Ulitsky et al., 2011). Conserved lncRNAs across vertebrates were identified using new layers of conservation such as expression pattern, secondary structure, and chromosomic position (Fig 5) (Amaral et al., 2018; Diederichs, 2014; Hezroni et al., 2015; Necsulea et al., 2014; Ulitsky et al., 2011).

LncRNA transcription and sequence evolved rapidly, however, tissue specificity is often conserved among primates and beyond (Necsulea et al., 2014). 47% of studied human

lncRNAs shared tissue specificity in primates, and 28% had expression pattern conservation across eutherians (Necsulea et al., 2014). Moreover, some lncRNAs show a conserved expression pattern between mice and zebrafish (Fig 5B) (Bitetti et al., 2018) and among amniotes (Chodroff et al., 2010).

Conserved RNA structure can also be a characteristic of evolutionarily conserved lncRNAs (Fig 5C). It is often hypothesized that the secondary and tertiary structures are subjected to selective pressure as it is supposed to carry the function of the RNAs (Diederichs, 2014; Guttman & Rinn, 2012; Wutz et al., 2002). The conserved structure is consistent with no conservation at the sequence level given that conservation of base-pairing properties matters more. Several studies highlighted functionally conserved RNA structure (Holmes et al., 2020; Uroda et al., 2019; Zhang et al., 2010). The identification of conserved RNA structures used to be challenging with the lack of efficient *in silico* tools. However, with deep learning more prediction tools arose (Sato et al., 2021) but their efficiencies are still limited. Experimental technics have been developed to assess, in living cells, the structure of lncRNAs (Siegfried et al., 2014; Ziv et al., 2018) but their applications are still complex and requires complex downstream analysis.

Chromosomic position, also known as synteny, is the best approach to identifying putative conserved lncRNA across vertebrates (Fig 5D). It can be used as the first layer to identify conserved lncRNAs by genomic comparison before further validations. Interestingly, using this method hundreds of lncRNA have been found conserved among vertebrates (Hezroni et al., 2015) with some from humans to zebrafish (Ulitsky et al., 2011). The zebrafish and mammalian genomes were extensively rearranged during the >400 million years of separation, however, lncRNAs synteny conservation (33.9%) was significantly higher than protein-coding gene pairs conservation (14.7%) (Ulitsky et al., 2011). Hence only with annotated genomes, chromosomic position allows the identification of conserved lncRNAs across species.

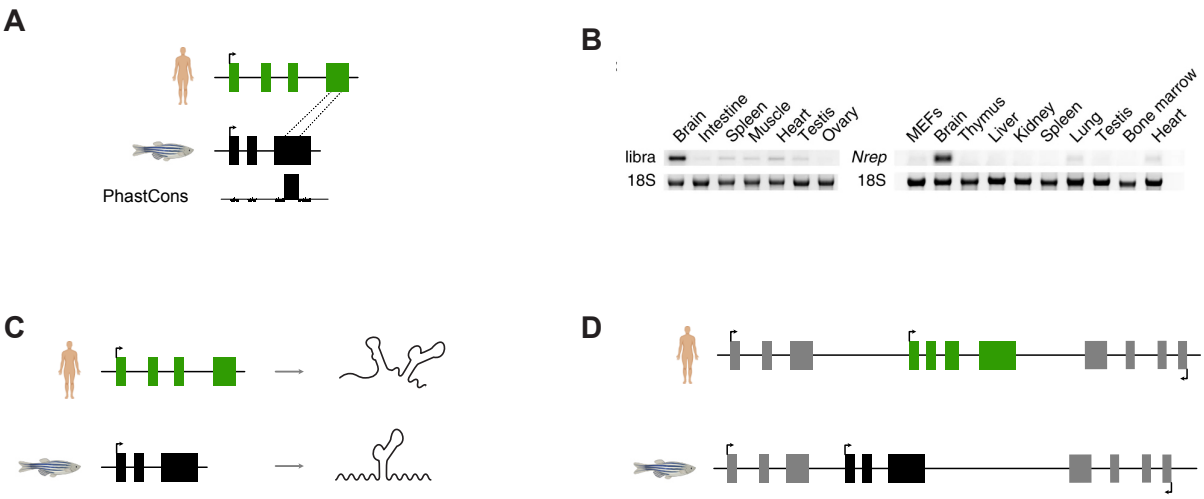


Figure 5: LncRNAs conservation

In grey are represented the mRNA, in green the human lncRNA and in black the zebrafish lncRNA. **A.** Small patches of sequence conservation in lncRNA. **B.** Tissue pattern expression conservation. Libra is the zebrafish lncRNA and Nrep the mice evolutionarily conserved transcript (figure from Bitetti et al., 2018). **C.** Conserved secondary structure **D.** Chromosomic position conservation (synteny).

It is often hypothesized that conserved lncRNAs share the same function in different species. However, the subcellular localization of the lncRNA can differ between species (Much et al., 2022), which can lead to different molecular function of the RNA. Moreover, a comparison between sequence and position conserved lncRNAs in mouse and human embryonic stem cells (ESCs) showed that different RNA processing can lead to different subcellular localization of lncRNA between species, leading to different molecular functions (Guo et al., 2020). Hence, the identification of conserved lncRNAs is not sufficient to infer the function of the homologous transcripts, further validations are required.

On the contrary, relatively similar biological processes are mediated by non-homologous ncRNAs in distant species such as the sexual chromosome dosage compensation. In Eutherian, the lncRNAs *Xist* randomly inactivate the second X chromosome by recruiting factors leading to the heterochromatinization of the X chromosome in females (Barr &

Bertram, 1949; Brockdorff et al., 1991, 1992; Brown et al., 1992). Interestingly, dosage compensation is also observed in flies with the involvement of the *rox1* and *rox2* ncRNAs (Franke & Baker, 1999), leading to the increased transcription of the unique X chromosome in males (Baker et al., 1994). In these two examples of sex chromosome dosage compensation, different mechanisms are employed, nevertheless, ncRNAs are central in their establishment.

1.2.4 LncRNA functions

LncRNA loci, regardless of their genomic location, can act by different molecular mechanisms. LncRNA loci can have a function through (1) the presence of enhancers regulating gene expression; (2) the act of transcription, and not the transcript per se; (3) the transcript itself. Interestingly, lncRNAs loci can carry one or several of these activities. The most interesting lncRNA loci are the ones where the function is carried by the transcript. LncRNA transcripts can act at different molecular levels (Constanty & Shkumatava, 2021; Statello et al., 2021). They can interact with DNA, RNA, or proteins (Fig 6). The vast majority of lncRNAs interact with proteins, it is these proteins that confer the molecular function of the lncRNAs. Therefore, deciphering the functional determinants of lncRNA driving RNA/protein interaction is dynamically studied, as this could allow the prediction of lncRNAs activity. It was initially

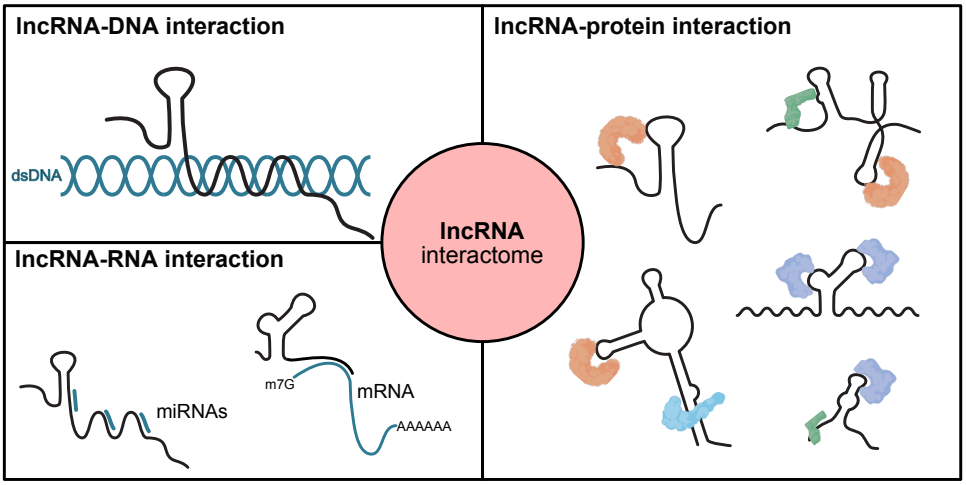


Figure 6: LncRNAs interactome

LncRNAs can interact with a wide range of molecules: with DNA, RNA, or proteins.

believed that RNA structure carried lncRNA functions (Diederichs, 2014; Guttman & Rinn, 2012; Smith et al., 2013; Wutz et al., 2002). Conserved RNA structures carry the function of a

few lncRNAs (Holmes et al., 2020; Uroda et al., 2019; Zhang et al., 2010). However, unbiased studies of RNA binding proteins (RBP) sequence and context binding sites have shown often low complexity and bi-partite RNA motifs (Dominguez et al., 2018). RBP binding specificity is, however, supported by the RNA local structure (Dominguez et al., 2018). These results show that the molecular function of lncRNA transcript is driven by the primary sequence and their local structure rather than the global RNA structure.

The identification of short RNA motifs harbored by lncRNA transcripts has become crucial in the lncRNA field. These motifs are thought to be the primary functional determinant of lncRNAs. Therefore, the identification of RNA motifs would direct toward a specific function of the lncRNA. Second, with the accumulation of identified RNA motifs this could drive to infer the functions of uncharacterized lncRNAs. Several short RNA motifs have been found to drive the function of a vast number of lncRNAs acting in a wide range of biological processes. These motifs were essential for lncRNA 3' end processing, R-loop formation, lncRNA subcellular localization, sequestration of proteins, lncRNA-proteins interactions, and regulation of miRNAs. This section is fully covered by the review Constanty & Shkumatava, 2021. Please refer to it in the annex section.

Two algorithms have emerged to identify short RNA motifs, or k-mers, on lncRNA transcripts. The SEEKR (SEquence Evaluation for Kmer Representation) (Kirk et al., 2018) and the lncLOOM (Lncrna Linear Order cOnserved Motifs) (Ross et al., 2021) algorithms. They allow the identification of RNA motifs on non-evolutionarily and evolutionarily conserved transcripts, respectively. These RNA motifs were shown to drive RNA-protein interactions (Kirk et al., 2018; Ross et al., 2021). Studying the molecular grammar of lncRNAs will greatly improve our capacity to infer lncRNA functions from the primary sequence and will contribute to a new classification of lncRNAs based on their functions.

1.3 *The Plasmacytoma Variant Translocation 1* (*PVT1*) lncRNA

The *PVT1* lncRNA was named based on the involvement of the genomic locus in variants translocation in plasmacytoma and Burkitt lymphoma (Cory et al., 1985; Graham & Adams, 1986; Webb et al., 1984). This lncRNA is of great interest as its expression is associated with multiple human diseases (Ghetti et al., 2020; Guo et al., 2022; Hanson et al., 2007) including cancers (Onagoruwa et al., 2020; Tseng et al., 2014; Tseng & Bagchi, 2015). Nevertheless, its molecular function remained not fully understood. In humans, *PVT1* is located on chromosome

8, 50 kb downstream of the protein-coding gene *MYC* (Fig 7A). Interestingly, this specific genomic feature is conserved throughout vertebrates (Fig 7A) (Hezroni et al., 2015).

1.3.1 A complex genomic locus

The human *PVT1* locus (~400 kb) is highly complex. *PVT1* map at the 8q24 chromosomal band, a locus involved in Burkitt lymphoma translocation (Davis et al., 1984; Graham & Adams, 1986). In addition, recurrent translocations of *PVT1* have been found in medulloblastoma, with the *PVT1/MYC* fusion observed in 60% of the analyzed samples (Northcott et al., 2012). *PVT1* translocation was also found in multiple solid and liquid tumors as reviewed here by Tolomeo et al., 2021. These examples show the *PVT1* locus as a hot spot of gene translocation in human cancers.

At the sequence level, enhancers are located in the *PVT1* locus regulating the expression of *PVT1* (Cho et al., 2018). In cancer, the *PVT1* promoter acts as a tumor suppressor DNA element by competing with the *MYC* promoter for interaction with these enhancers (Cho et al., 2018). Numerous mutations are spamming the *PVT1* promoter in cancer, leading the enhancers to promote *MYC* expression, thereby promoting tumorigenesis (Cho et al., 2018).

Interestingly, five miRNAs mapped within the *PVT1* intronic sequences (miR-1204, miR-1205, miR-1206, miR-1207, and miR-1208) (Fig 7B) (Huppi et al., 2012). These miRNAs are conserved in mice. Studies of the function of these miRNAs are still ambiguous but showed that they act both as positive and negative regulators of tumorigenesis (Chen et al., 2014; Huppi et al., 2008; Li et al., 2019; Liu et al., 2018). Remarkably, it was described that miR-1205 suppresses gastric cancer growth by targeting the human telomerase reverse transcriptase (hTERT) (Chen et al., 2014). miR-1205 binds to the 3' UTR of the *hTERT* mRNA, regulating its expression. Moreover, miR-1205 upregulation inhibits the growth of transplants in mice (Chen et al., 2014).

The linear lncRNA *PVT1* is transcribed by the RNA polymerase II, capped, spliced, and polyadenylated. In humans, 183 linear variants are categorized in the Ensembl Genome Browser with multiple alternative transcriptional start sites (TSSs) and alternative polyadenylation (polyA) sites. The expressions of some variants have been validated experimentally while others result from *in silico* predictions. The expression of different variants has been described in different cancer types (Guan et al., 2007; He et al., 2019; Martínez-Barriocanal et al., 2020; Pal et al., 2020; Pal & Ogunwobi, 2019), making function comparison complex between studies. On top of linear variants, circular variants of *PVT1* have

been predicted *in silico*. While a vast number have been predicted, only one variant has been functionally studied: *circPVT1* (410 nt). *CircPVT1* shares the full length of the exon 2 sequence with the linear *PVT1* (Fig 7B) (Wang et al., 2020). The linear and circular forms of *PVT1* are expressed from different promoters, thus implementing independent regulation (Verduci et al., 2017).

Intriguingly, in mice, fewer mouse *Pvt1* (*mPvt1*) linear variants have been described with 10 variants in the Ensembl Genome Browser and three mainly expressed variants (Zheng et al., 2016). However, no circular *mPvt1* forms have been described yet.

Finally, a tumor-enriched Human Leukocyte Antigen (HLA) ligand encoded by the lncRNA *PVT1* was characterized in colorectal cancer (Kikuchi et al., 2021). In their study, Kikuchi et al. showed the translation of a 10-mers micropeptide, named HF-10, from a cryptic ORF encompassing the exon 1 and 2 of the lncRNA *PVT1* (Kikuchi et al., 2021). The *PVT1* antigen was able to induce cytotoxic $CD8^+$ T-cell response showing its specific recognition by T-cells (Kikuchi et al., 2021). In patients, neoantigen formations often arise from somatic mutations, thereby, most neoantigens are specific to individuals (Rospo et al., 2019). However, the *PVT1* antigen is found in four out of six colorectal tissues tested (Kikuchi et al., 2021). This is most probably due to the high expression of *PVT1* in this cancer type (He et al., 2019).

The above examples illustrate the complexity of the *PVT1* locus. This locus is a chromosomic translocation hotspot hosting five conserved miRNAs and one conserved lncRNA that has multiple variants.

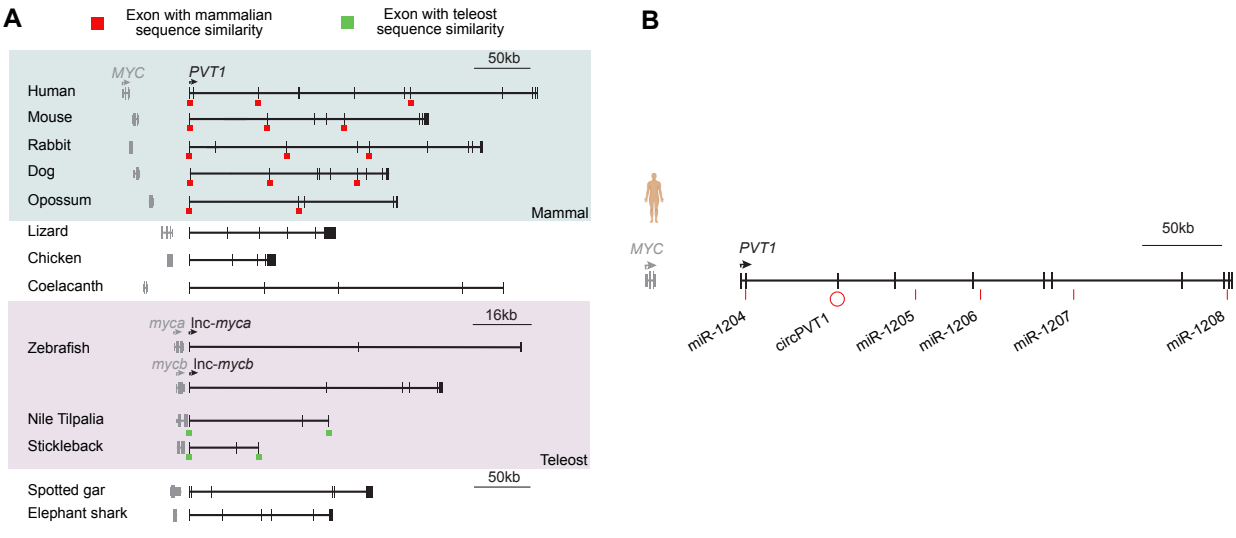


Figure 7: *PVT1* genomic localization

A. schematic representation of the *PVT1* genomic position across vertebrates (adapted from Hezroni et al., 2015). **B.** Schematic representation of the human *PVT1* locus with the mapped miRNAs and circular variant.

1.3.2 *PVT1* mode of action

Over the years, several studies have focused on the function of the *PVT1* transcript, mainly in cancer (Onagoruwa et al., 2020). A wide diversity of molecular mechanisms has been found in different models, resulting in a complex generalization of the functions of this lncRNA. Moreover, the functions of the circular and linear variants of *PVT1* can differ whereas in some cases both variants act on the same pathway, at different levels (Ghetti et al., 2020; Palcau et al., 2022; Traversa et al., 2022). For clarity, and because there is no obvious information about the evolutionary conservation of the *circPVT1*, the next parts will focus on the linear variants of the lncRNA *PVT1*.

Oncogenic functions

The lncRNA *PVT1* is upregulated in a wide range of cancer types (Derderian et al., 2019; Tseng & Bagchi, 2015). Interestingly, *PVT1* is found alongside *MYC* in 98% of the 8q24 amplicons (Tseng et al., 2014; Tseng & Bagchi, 2015). It shows that both loci are preferentially amplified in cancer. Intriguingly, an extra copy of *MYC* alone or *PVT1* alone is not sufficient to promote tumorigenesis whereas an extra copy of both loci is (Tseng et al., 2014). *PVT1* and *MYC* expression are correlated in cancers (Martínez-Barriocanal et al., 2020; Tseng & Bagchi, 2015). These data show that *PVT1* might act as a DNA element regulator by promoting *MYC* expression in cancer.

Specific downregulation of the lncRNA *PVT1* considerably impacts tumor growth and cell proliferation (Guo et al., 2018; Tseng et al., 2014; Wan et al., 2018; Wu et al., 2019). This experiment demonstrates the relevance of the *PVT1* transcript in tumorigenesis. Interestingly, high expression of *PVT1* is an indicator of poor prognosis and poor survival of patients in a wide range of cancers, including prostate cancer (Sun et al., 2020), gastric cancer (Kong et al., 2015), nasopharyngeal cancer (He et al., 2018), breast cancer (Liu et al., 2021), and in glioma (Bi et al., 2021). Diverse modes of action have been described for *PVT1* in cancers, describing *PVT1* as a proto-oncogene.

PVT1 has been described to act as a competing endogenous RNA (ceRNA) in some diseases, cancer included (Chen et al., 2019; Ogunwobi & Kumar, 2019; Qian et al., 2018; Tian et al., 2018; Yang et al., 2019). ceRNAs are RNA transcripts that compete with other RNAs for the binding of miRNAs. *PVT1* is competing with the binding of different miRNAs, thereby preventing the degradation of the initially targeted mRNAs. These interactions promote tumorigenesis (Conte et al., 2017). The miRNAs targeted by *PVT1* are highly dependent on the cancer type studied. However, some common mechanisms have been found: (1) *PVT1* leads to the upregulation of the Hypoxia-inducible factor (HIF) due to the sponging of different miRNAs in non-small cell lung cancer and gastric cancer, promoting tumorigenesis (Huang et al., 2017; Zhang et al., 2017); (2) *PVT1* regulates negatively the activity of the miR-200 family, known to prevent tumorigenesis, in breast cancer and clear renal cell carcinoma (Paci et al., 2014; Yang et al., 2017). The relevance of such a mechanism is still debated but a vast number of papers relate the involvement of *PVT1* as a ceRNA (Derderian et al., 2019; Wang et al., 2019). However, none of the studies performed serious stoichiometry analysis to evaluate how the lowly expressed lncRNA *PVT1* can sponge highly expressed miRNA families.

PVT1 is also known to drive tumorigenesis through protein interactions. Several main protein interactions have been described. However, these interactions have been shown in different cancer cell lines. In gastric cancer cell line (HGC-27), an RNA-protein pull-down assay showed the interaction between the Forkhead Box Protein M1 (FOXM1) and *PVT1* (Xu et al., 2017). *PVT1* stabilized FOXM1 post-transcriptionally (Xu et al., 2017). Interestingly FOXM1 promotes *PVT1* expression, leading to a positive feedback loop (Xu et al., 2017). This mechanism promotes tumor growth and metastasis in a FOXM1-dependent manner (Xu et al., 2017). Several studies reported *PVT1* binding to the Enhancer of Zeste Homolog 2 (EZH2), the catalytic subunit of the Polycomb repressive complex 2 (PRC2), in different cancer cell lines (Guo et al., 2018; Sun et al., 2021; Wan et al., 2016; Yang et al., 2017; Zhang et al., 2016; Zhou et al., 2016). More particularly, in hepatocellular carcinoma *PVT1* binds to EZH2 driving its stabilization (Guo et al., 2018). This leads to the binding of EZH2 to MDM2 protein, promoting tumor development by inhibiting P53 activity, inhibiting its effect on p53 (Guo et al., 2018). A similar mechanism was shown in lung adenocarcinoma where *PVT1* has been found to bind and guide EZH2 to the Large Tumor Suppressor Kinase 2 (LATS2) promoter, a positive regulator of the P53 pathway, and represses its expression (Wan et al., 2016). Nonetheless, it is known that EZH2 tends to bind to a wide range of RNAs without specificity, hence the EZH2/*PVT1* interaction must be analyzed with caution. Alternatively, *PVT1* has been found to bind directly to MYC protein in breast cancer cell lines (Tseng et al., 2014). *PVT1*

regulates MYC stability by protecting MYC from threonine 58 (T58) phosphorylation (Tseng et al., 2014). MYC T58 phosphorylation is associated with its degradation. Interestingly, MYC was found to promote *PVT1* expression, leading to a positive feedback loop, promoting tumor progression (Tseng et al., 2014). Recent studies have shown that the MYC/*PVT1* interaction appears to depend on the N6-Methyladenosine (m6A) modification of *PVT1* (Lee et al., 2021).

All these examples show the complex mechanism of *PVT1* mode of action as an oncogene. *PVT1* has been found to act at different level promoting tumor growth and metastasis progression. However, the *PVT1* molecular function may vary dependent on the type of cancer and the isoforms expressed.

Tumor suppressor functions

PVT1 was initially described as an oncogenic lncRNA. Nonetheless, recent studies showed the involvement of *PVT1* as a tumor suppressor. A P53-binding site, located in the vicinity of an alternative *PVT1* TSS leads to P53-dependant regulation of a specific *PVT1* variant (Barsotti et al., 2012). An extensive study showed that this P53-dependant *PVT1* transcription is conserved in mice, driving the expression of an isoform called *Pvt1b* (Olivero et al., 2020). Stress-induced conditions caused *Pvt1b* upregulation in a P53-dependent manner (Olivero et al., 2020). *Pvt1b* accumulated at its transcription site, resulting in *cis*-repression of Myc (Olivero et al., 2020). Remarkably, genetic ablation of the P53 binding site promoted tumor growth *in vivo* (Olivero et al., 2020). This shows the specific mode of action of *PVT1* acting as a tumor suppressor in a P53-dependant manner.

Non-cancer related functions

While *PVT1* has been mostly studied in cancer, some studies were interested in *PVT1* involvement in other diseases and physiological conditions.

PVT1 loss-of-function impaired the proliferation of trophoblast cells by increasing apoptosis (Xu et al., 2017). Conversely, increase expression of *PVT1* promoted trophoblast proliferation (Xu et al., 2017). Trophoblast cells are extraembryonic cells essential for embryonic development, it is therefore not surprising that low *PVT1* expression is associated with a higher abortion risk (Qin et al., 2019; Wang et al., 2019).

Single-cell RNA-sequencing (scRNA-seq) showed the involvement of *mPvt1* in muscle atrophy (Alessio et al., 2019). Interestingly, it was shown that *mPvt1* regulates Myc T58 phosphorylation in muscle atrophy. Upregulation of *mPvt1* promoted Myc stabilization, driving autophagy and muscle atrophy (Alessio et al., 2019). However, no experiments such as

RNA immunoprecipitation (RIP) or pull-down assay were performed, resulting in an incomplete interpretation.

mPvt1 was shown to govern epidermal cell differentiation through Myc/*mPvt1* interaction (Lee et al., 2021). Ablation of m*Pvt1* impaired the self-renewal and wound healing capacity of skin cells (Lee et al., 2021). Interestingly, Myc/*mPvt1* interaction was dependent on m6A modification of m*Pvt1* (Lee et al., 2021). Although this study showed a direct interaction between m*Pvt1* and Myc, the RIP was performed by mixing cell extracts and thus was not performed *in cellulo*.

PVT1 mode of action is highly complex. *PVT1* locus act as a DNA regulatory element and as an active lncRNA. The transcript of *PVT1* has been found to interact with different miRNAs and proteins. Generalization of the *PVT1* mode of action is therefore challenging, mostly due to the different models studied. Interestingly, one mode of action seems to be generalized in cancer and non-cancer studies: *PVT1* impacts MYC expression at the protein level.

Tissue/cell type	Biological implication	Cellular outcome	Molecular function	Refs
Trophoblast cells	Downregulation of *PVT1* contributes to gestational diabetes mellitus, preeclampsia, and multiple abortions.	*PVT1* promotes cell viability, proliferation, and migration.	*PVT1* overexpression led to AKT phosphorylation, influencing the PI3K/AKT pathway. *PVT1* was described to interact with the epigenetic factor EZH2. It is supposed that *PVT1* may act by directing EZH2 to the *ANGPTL4* promoter.	Wang et al., 2019 Xu et al., 2017
Podocytes	*PVT1* high expression is linked to diabetic nephropathy in patients.	*PVT1* expression induces apoptosis.	It was supposed that PVT1 regulates *FOXA1* expression through EZH2 interaction.	Liu et al., 2019
Mesangial cells	*PVT1* level expression increase in hyperglycemic conditions and is associated with end-stage renal disease in type 1 and 2 diabetes.	*PVT1* high expression was correlated with the expression of extracellular matrix factors.	Downregulation of *PVT1* expression significantly impacted the expression and secretion of ECM factors (FN1, COL4A1, TGFB, and PAI-1).	Alvarez & DiStefano, 2011
$CD4^+$ T cells	*PVT1* expression is positively correlated to $CD4^+$ T cell activation. *PVT1* was found upregulated in the salivary gland tissue of Sjögren's syndrome patient.	Maintenance of the proliferation in activated $CD4^+$ T cells.	The lncRNA *PVT1* is a T cell receptor-responsive gene through the PI3K/AKT pathway. *PVT1*-depleted cells have a lower activation of glycolytic genes upon T cell activation. Moreover, PVT1 downregulated led to a decrease in MYC protein level. Therefore, it is suggested that *PVT1* regulates MYC expression to promote glycolytic metabolism upon CD4+ T cell activation.	Fu et al., 2020

Tissue/cell type	Biological implication	Cellular outcome	Molecular function	Refs
Skin	*PVT1* regulates the wound-healing capacity of the skin.	*PVT1* ensures the colony formation of epidermal stem cells and sustains epidermal stemness *in vivo*.	*PVT1* interacts with MYC through its methylated sites (m6A).	Lee et al., 2021
Fibroblast and lung adenocarcinoma	The specific *PVT1b* isoforms inhibit tumorigenesis.	Genotoxic and oncogenic stress induces *PVT1b* expression to repress cell proliferation	*PVT1b* indirectly suppresses MYC expression *in cis*	Olivero et al., 2020
Multiple tissues	High expression of *PVT1* takes part in tumorigenesis.		*PVT1* interacts with different miRNAs, dependent on the cancer type.	Wang et al., 2019
Breast cancer	*PVT1* upregulation promotes cancer development.	*PVT1* sustains cells proliferation	*PVT1* interacts with MYC protein, inhibiting its phosphorylation, therefore its subsequent degradation. This leads to MYC stabilization. MYC induces *PVT1* expression, driving a positive feedback loop.	Tseng et al., 2014
Gastric cancer	High *PVT1* expression is associated with advanced progression and poor prognosis in patients.	The lncRNA increases cell proliferation and drives tumor growth and angiogenesis.	*PVT1* interacts with the activated STAT3 protein and increases its transcriptional activity. STAT3 induces *PVT1* expression, leading to a positive feedback loop.	Zhao et al., 2018
Gastric cancer	*PVT1* expression correlated to poor prognosis in patients.	*PVT1* enhanced gastric cancer cells proliferation and invasion	*PVT1* directly binds to FOXM1, leading to its stabilization. FOXM1 induces *PVT1* expression, leading to a positive feedback loop.	Xu et al., 2017

Tissue/cell type	Biological implication	Cellular outcome	Molecular function	Refs
Nasopharyngeal carcinoma	Upregulation of *PVT1* was associated with reduced survival of patients.	*PVT1* promotes cell proliferation, colony formation, and tumorigenesis of mouse xenografts.	*PVT1* acts as a scaffold for the acetyltransferase KAT2A, promoting the TIF1B/H3K9ac complex-mediated NF90 transcription, which regulates HIF-1a stability.	Wang et al., 2020
Nasopharyngeal carcinoma	*PVT1* is highly expressed and associated with poor prognosis in patients.	*PVT1* confers cancer cell resistance to radiotherapy by inhibiting apoptosis.	*PVT1* expression sustain the DNA repair capability of the cells through the ATM-p53 pathway and impacted caspase activation.	He et al., 2018
Multiple tissues	*PVT1* upregulation is associated with tumorigenesis.		*PVT1* interacts with EZH2, driving epigenetics repression of transcription of specific locus.	Guo et al., 2018; Sun et al., 2021; Wan et al., 2016; Yang et al., 2017; Zhang et al., 2016; Zhou et al., 2016

Table 1: non-exhaustive list summarizing *PVT1* transcript function.

1.1 Thesis objectives

PVT1 is one of the first lncRNAs identified in the human genome. An exploding number of studies tried to determine its function. However, most studies focused on the involvement of *PVT1* in cancer. Characterizing the *PVT1* mode of action in cancer remains challenging. Thus, deciphering the *PVT1* mode of action in other models will considerably help to better characterize its mode of action. *PVT1* represents a compelling case of lncRNA evolution. It is conserved among vertebrates by genomic position conservation, located downstream of the *MYC* protein-coding gene (Hezroni et al., 2015). Interestingly, mammalian *PVT1* homolog transcripts share alignable sequence conservation whereas outside of the mammalian class the sequence of the transcripts is not conserved. Thus, deciphering the conserved function of *PVT1* in vertebrates will greatly improve the characterization of *PVT1* function and shed light on conserved mechanisms of lncRNA across species without alignable sequence conservation. We are using the conserved characteristics of the zebrafish model to study *PVT1*. The zebrafish model is an excellent genetic vertebrate, non-mammalian, model. Its complete genome sequencing and annotation allowed the characterization of *PVT1* homologous transcripts conserved through genomic position (Hezroni et al., 2015). Available molecular tools allow specific genome modification with the Clustered Regularly Interspaced Short Palindromic Repeats – CRSPR associated Protein 9 (CRISPR-Cas9) technology.

2. RESULTS

2.1 Functional characterization of the lncRNA *PVT1 in vivo*

Florian Constanty[1], Perrine Lavalou[1], Carine Vias[1], Alena Shkumatava[1]

[1] Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, Paris 75005, France

2.1.1 Introduction

Long non-coding RNAs (lncRNAs) emerged as key regulators of a wide range of biological processes, such as development and cell differentiation (Constanty & Shkumatava, 2021; Marchese et al., 2017; Perry & Ulitsky, 2016; Statello et al., 2021; Ulitsky & Bartel, 2013). Their expressions have been linked to a vast number of human diseases, including almost all types of cancer (Esposito et al., 2019; Gutschner & Diederichs, 2012; Huarte, 2015). Relevant lncRNAs in cancer are mostly studied in specific cancer cell lines, considerably reducing the extrapolation of their functions. While lncRNAs represent an untapped source of therapeutic targets, only a handful have been thoroughly studied, with their mode of action elucidated (Bitetti et al., 2018; Constanty & Shkumatava, 2021; Ross et al., 2021; Statello et al., 2021; Taiana et al., 2020; Tripathi et al., 2010; Uroda et al., 2019; Ziv et al., 2021). Deciphering the physiologic role of lncRNAs would help to characterize their deregulation in diseases. *In vivo* studies are of utmost importance to characterize the functionality of lncRNAs in physiologic conditions (Andergassen & Rinn, 2022; Bitetti et al., 2018; Kopp et al., 2019; Nakagawa et al., 2012). Although these studies are limited to the evolutionarily conserved lncRNAs, hundreds of lncRNAs are conserved from humans to zebrafish (Hezroni et al., 2015). These represent a compelling source of lncRNAs to be characterized.

To understand the evolution and the functional contribution of conserved lncRNAs, we studied the cancer-associated *Plasmacytoma variant translocation 1* (*PVT1*) lncRNA. This lncRNA, was identified as a locus involved in human lymphoma (Cory et al., 1985; Graham & Adams, 1986; Webb et al., 1984). Its expression is upregulated in a wide range of human diseases (Ghetti et al., 2020; Guo et al., 2022; Hanson et al., 2007), cancers included (Onagoruwa et al.,

2020; Tseng et al., 2014; Tseng & Bagchi, 2015). *PVT1* expression is associated with poor prognosis and poor survival of cancer patients (Bi et al., 2021; He et al., 2018; Kong et al., 2015; Yang et al., 2017). *PVT1* is located downstream of the *MYC* protein-coding gene in vertebrates (Hezroni et al., 2015). Intriguingly, *PVT1* is deeply conserved in vertebrates without sequence conservation (Hezroni et al., 2015). Reports showed that *PVT1* acts as an oncogene by interacting with miRNAs (Wang et al., 2019). Other more relevant studies showed that *PVT1* interacts with proteins including the MYC proto-oncogene protein. This interaction protects MYC from degradation and promotes tumorigenesis (Guo et al., 2018; Tseng et al., 2014; Tseng & Bagchi, 2015; Xu et al., 2017; Xu et al., 2017; Yang et al., 2017; Zhou et al., 2016). Remarkably, recent studies showed that a specific variant of *PVT1* act as a tumor suppressor in a p53-dependent manner, in human and mouse cells (Olivero et al., 2020). Although *PVT1* is the first ncRNA locus identified in the human genome and is a therapeutical target, its mode of action is still understudied.

In this study, we characterize the functional conservation of the lncRNA *PVT1* in development. We show that despite the lack of sequence conservation in vertebrate species, *PVT1* homologous transcripts share two deeply conserved short RNA motifs. The genetic removal of these motifs from the zebrafish transcripts led to multiple developmental defects. The comprehensive study of zebrafish *pvt1* gene regulation showed that the zebrafish and human *pvt1* transcripts are under the control of the same transcriptional pathways.

2.1.2 Results

PVT1 transcripts harbor deeply conserved RNA-motifs

In vertebrates, *PVT1* is located downstream of the protein-coding gene *MYC* (Hezroni et al., 2015). In zebrafish, due to a full genome duplication during teleost fish evolution, two *PVT1* homologous genes were identified (Hezroni et al., 2015). These two lncRNAs are localized downstream of the *myca* and *mycb* genes, thus named lnc-*myca* and lnc-*mycb* respectively (Fig 8A). Interestingly, lnc-*myca* and *myca* expression were correlated during zebrafish development (SuppFig 8A). The correlation between lnc-*mycb* and *mycb* expression was not significant throughout development but we observed a correlation at specific developmental stages (SuppFig 8B). These suggest that *pvt1* homologous genes share common DNA regulatory elements with the *myc* protein-coding genes in zebrafish. Since subcellular localization defines the molecular function of lncRNAs (Chen, 2016; Guo et al., 2020), we

defined the sublocalization of the lnc-*myc* RNAs. Interestingly, as the human *PVT1* (Kong et al., 2015), both lnc-*myca* and lnc-*mycb* transcripts are predominantly present in the nucleus (Fig 8B, SuppFig 8C). This results in the possible conservation of molecular functions. Linear sequence conservation often supports conserved functions. However, mammalian and zebrafish homologous transcripts do not share alignable sequence conservation (Fig 8A) (Hezroni et al., 2015). Nonetheless, short RNA motifs (k-mers) emerged as an alternative to decode conserved molecular functions between fast-evolving transcripts (Kirk et al., 2018; Ross et al., 2021). Using the newly published lncLOOM algorithm (Ross et al., 2021), two deeply conserved RNA motifs were identified (Fig 8C). Remarkably, these two k-mers were found in all species analyzed (11) (Supp Data 2). One k-mer is missing on the lnc-*myca* transcript (Fig 8C). These two 6-mers (TGGATT and TTTTTG) are present multiple times on the different *PVT1* homologous transcripts (Fig 8C). To test their functionality in zebrafish, we performed genetic ablation of the regions containing the motifs using CRISPR-Cas9 (Fig 8C, SuppFig 8D-G). Genetic removal of these regions did not impact the expression of the lnc-*myc* transcripts. The remaining parts of both transcripts were stably detected by RNA-seq and quantitative real-time polymerase chain reaction (qRT-PCR) at different developmental stages (Fig 8D, SuppFig 8H). The specific removal of the RNA motifs in the zebrafish transcripts did not impact the expression of the *myc* mRNA (SuppFig 8I). These results show that the two identified motifs are not involved in transcript stabilization nor act as enhancers for myc expression during development.

Genetic ablation of the RNA motifs in zebrafish leads to multiple developmental defects

To eliminate any potential compensation by homologous transcripts, we generated double homozygous *pvt1* mutants (SuppFig 8D-G). When crossing double heterozygous fish, the expected ratio is 1/16 of fish being Wild-Type (WT) and 1/16 of double homozygous mutants. At the adult stage of at least 2-months, we repetitively retrieved (from 5 independent crosses) a lower number of double homozygous mutants. On average, we retrieved 1/50 double homozygous mutants (Fig 9A). The non-Mendelian distribution was specific to the double homozygous mutants and, to some extent, to the lnc-*myca* heterozygous and lnc-*mycb* homozygous mutants (Fig 9A). Mendelian distribution of the genotypes was observed at 2 days post-fertilization (dpf) but not at the adult stage (SuppFig 9A & B). Thus, the lnc-*myc* partial deletion impaired the survival of embryos to adulthood. Intriguingly, the impaired survival of the embryos was exacerbated as the parents were aging (SuppFig 9C). 3-months-old fish gave rise to a limited number of double homozygous mutant embryos surviving to adulthood,

whereas 12-months-old fish did not give rise to double homozygous mutant adult fish (SuppFig 9C). This was also observed for the lnc-*myca* heterozygous and lnc-*mycb* homozygous mutants as their number considerably decreased in the progeny upon the aging of the parents (SuppFig 9C).

A fraction of double homozygous mutant embryos developed multiple developmental defects. They first appeared at 3 dpf, with pericardial edema, yolk retention, and general body misshape (Fig 9B, SuppFig 9D). At 6 dpf these embryos displayed aggravated developmental defects, with pericardial and yolk edema, eye edema, no inflation of the swim bladder, malformation of the head, eye, and jaw, and general body misshape (Fig 9B, SuppFig 9D). The pericardial, yolk, and eyes edema were not present in all embryos. Therefore, the other observed features were used to determine if the embryos had developmental defects (SuppFig 9D). Although not all double homozygous mutant embryos displayed these defects, an important proportion of embryos did, around 30 to 40% (Fig 9C). As expected, simple homozygous embryos did not display the same proportion of embryos with defects (~5-10%) (Fig 9C). Thus, there is a possible compensation mechanism between the two transcripts. From 3 dpf, two groups of embryos can be distinguished: embryos with developmental defects (named DD) and embryos with no developmental defects (named nDD).

The embryonic developmental defects are associated with an increase of the p53 pathway and apoptosis

To decipher the impact of the RNA motifs deletion at the molecular level, we analyzed the transcriptome of 1 dpf embryo by RNA-seq. At this development stage, no developmental defects could be observed at the organism level. However, we reasoned that the transcriptome could be impacted before the appearance of the first developmental defects. Because of the incomplete penetrance, we sequenced single embryos, 5 WT, and 14 double homozygous mutants. While PCA analysis showed WT embryos clustering separately from the double homozygous mutants, two distinct clusters were observed among the double homozygous mutants (Fig 10A). The two groups were analyzed separately, with group 1 composed of 11 double mutant embryos (80% of the embryos) and group 2 with 3 embryos (20%) (Fig 10A). The proportion of embryos in group 2 corresponds to the proportion of embryos that develop developmental defects at later stages (Fig 9C & 10A). One pathway was commonly upregulated in both double mutant groups: ribosome biogenesis (Fig 10B-E, SuppFig 10A & B). Other misregulated pathways were specific to each group with the upregulation of the p53 pathway in the embryos of group 2 (Fig 10E, SuppFig 10C). The p53 pathway is a regulator of

the stress response, its activation is triggered by cellular stress. Among others, ribosome biogenesis stress is known to trigger the p53 pathway (Golomb et al., 2014). Intriguingly, p53 upregulation during zebrafish development is known to lead to developmental defects (Chakraborty et al., 2009; Danilova et al., 2010; Liu et al., 2003; Plaster et al., 2006).

To determine the differentially expressed genes associated with the appearance of developmental defects, we performed an RNA-seq of embryos with developmental defects (DD) at 3 dpf (Fig 10F). Despite multiple developmental defects, we did not retrieve a considerable amount of misregulated genes in the DD embryos, with 499 (3,9%) genes upregulated and 937 (7,3%) genes downregulated (Fig 10G). This could suggest that specific pathways are impacted upon the deletion. The p53 pathway and apoptosis were upregulated in DD embryos (Fig 10H, SuppFig 10D-F). Interestingly, in the same embryos, we detected a downregulation of glycolysis and purine metabolism (SuppFig 10D-F). This is concordant with the p53 pathway upregulation as p53 is known to directly repress the transcription of glucose transporters (Schwartzenberg-Bar-Yoseph et al., 2004). P53 also regulates glucose metabolism by direct or indirect regulation of key enzymes of the glycolysis pathway (Bensaad et al., 2006; Franklin et al., 2016; Jiang et al., 2011; Ros et al., 2017; Wang et al., 2014; Zhang et al., 2014). The downregulation of the purine metabolism could be interpreted by p53 inhibition of the pentose phosphate pathway (PPP) (Jiang et al., 2011), an essential precursor for the biosynthesis and metabolism of purines (Xu et al., 2021). In order to evaluate the specificity of the p53 pathway upregulation with the appearance of developmental defects, we analyzed the expression of the p53 pathway in nDD embryos. The p53 pathway was specifically upregulated in DD embryos and not in nDD embryos (Fig 10I). The increase of apoptosis was confirmed through Terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) assay in the retina of embryos at 3 dpf. We found on average 11 apoptotic cells in DD embryos whereas an average of 2 apoptotic cells in WT was observed (Fig 10J & K). In nDD embryos, 1 apoptotic cell per retina was observed, on average (Fig 10J & K). These results show that the onset of developmental defects in embryos was specifically associated with the upregulation of the stress pathway.

Zebrafish embryos develop externally, as such, they are more susceptible to their environment which can lead to DNA damage. Moreover, fast-dividing cells are more prone to DNA damage, which can lead to apoptosis. Thus, we tested if the increase of apoptosis in DD embryos was due to an accumulation of DNA damage in cells. We did not observe an increase in DNA damage sites in DD embryos, with an average of 3 γH2AX positive cells per retina whereas 2 γH2AX positive cells were found in WT, on average (Fig 10L, SuppFig 10G).

However, these differences might be enhanced if the embryos are previously treated with a DNA damaged agent. nDD embryos had significantly more γH2AX positive cells (4 per retina on average) (Fig 10L, SuppFig 10G). This can be explained by the accumulation of cells with DNA damage throughout development due to the slightly lower rate of apoptosis in these embryos (Fig 10J & K).

These results highlight the specificity of the p53 pathway and apoptosis upregulation associated with developmental defects. These results are in accordance with previous findings that linked p53 and apoptosis upregulation to developmental defects in zebrafish (Chakraborty et al., 2009; Danilova et al., 2010; Liu et al., 2003; Plaster et al., 2006).

Myc proteins level is upregulated at the onset of the developmental defects

The *PVT1* gene is located downstream of the *MYC* protein-coding genes in vertebrates (Hezroni et al., 2015). LncRNAs are known to impact gene expression in *cis* (Gil & Ulitsky, 2020). Thus, *myc* is a potential gene impacted by the *pvt1* partial deletion in zebrafish. In addition, *mPvt1* accumulates at its transcription site (Olivero et al., 2020), and binds to and control myc protein expression level (Tseng & Bagchi, 2015)

The partial deletion of the lnc-*myc* RNAs had a limited impact on *myc* gene expressions. At 1 dpf, *mycb*, but not *myca*, was upregulated in embryos of group 2 and no impairment of *myc* expression was observed in embryos of group 1 (Fig 11A). HOMER analysis revealed an enrichment of myc binding sites at the promoter of transcripts showing higher RNA levels in embryos from both groups of double homozygous mutants (SuppFig 11A). This suggests the transcriptional activity of myc proteins in both groups of mutants that was not accompanied by an upregulation of *myc* gene expression in both groups (Fig 11A). The protein level of myc was not quantifiable at 1 dpf by western blot due to the low protein yield at this developmental stage. MYC hyperactivity is often linked to increased translation and cell proliferation as MYC drives essential pathways like ribosome biogenesis (Dang, 2012; Mori et al., 2021; Poortinga et al., 2004, 2011), pathway found upregulated in both groups, at 1 dpf (Fig 10D & E).

With the appearance of the first developmental defects at 3 dpf, the myc protein level was upregulated in the DD embryos (Fig 11B-C). This upregulation was not along with an increase of the expression as *myca* and *mycb* expression were stable in DD embryos (SuppFig 11B & C). Interestingly, the myc protein level increased was accompanied by an upregulation of the truncated lnc-*myc* RNAs, specifically in the DD embryos (Fig 11D). Consistently, sequence analysis in the vicinity of the lnc-*myc* promoters highlighted the presence of a myc binding site (E-Box) localized at the promoter of lnc-*myca* and lnc-*mycb* genes (Fig 11E,

SuppFig 11D & E). A p53 response elements (REs) was also found 3 kb downstream of the lnc-*myca* promoter (Fig11E, SuppFig 11D & E). To establish confidence in the functionality of the identified binding sites, we performed the same sequence analysis using the mouse *Pvt1* promoter. We identified, with similar confidence, the presence of Myc and P53 binding sites, previously identified as functional (Carramusa et al., 2007; Olivero et al., 2020) (Fig 11E, SuppFig 11D & E). This suggests that, as in humans (Tseng & Bagchi, 2015), zebrafish *pvt1* expressions might be driven by myc.

Later in development, at 6 dpf, DD embryos have important developmental defects (Fig 9B). Interestingly at this developmental stage *myc* expression was downregulated in DD embryos at the RNA level (SuppFig 11F), accompanied by a significantly impaired protein level (SuppFig 11G & H). Because DD embryos are severely impacted at this development stage, cell proliferation and gene expression globally might be impaired.

In summary, our data show an upregulation, at the protein level, of the protein myc. This upregulation happened at the onset of the developmental defects, with myc being transcriptionally active before the first appearance of the defects. Interestingly, the lnc-*myc* RNAs are also upregulated. This could be due to the transcriptional regulations of these genes by myc as myc binding sites are found at the promoters of these genes.

Zebrafish *pvt1* homolog expression is induced by stress

Motifs analysis of the lnc-*myc* promoters identified p53 REs specifically in the vicinity of the lnc-*myca* promoter (Fig 11E, SuppFig 11D & E). Interestingly, this suggests that, as in mice and humans, the zebrafish transcript expression could be driven by p53 in stress conditions. To test this, embryos were treated with camptothecin (CPT), a topoisomerase I inhibitor, leading to the accumulation of single and double DNA breaks in the genome, which induces the p53 pathway (Langheinrich et al., 2002). The embryos were treated with different concentrations of CPT, inducing the p53 pathway in a non-concentration-dependent manner (SuppFig 12A). As expected, only the lnc-*myca* expression was responsive to CPT treatment and not the lnc-*mycb* expression (Fig 12A). tp53 morpholino were injected in embryos to determine if p53 directly induces the expression of lnc-*myca* in stress-induced conditions (SuppFig 12B). tp53 morpholino successfully limited p53 and p21 induction upon CPT treatment (Fig 12B, SuppFig 12C & D). lnc-*myca* induction was abolished with p53 knockdown in CPT-treated embryos (Fig 12B, SuppFigg 12E). As anticipated, tp53 morpholinos injection did not impact lnc-*mycb* expression in stress-induced conditions (SuppFig 12F). These results show that the zebrafish *pvt1* homologous transcript, lnc-*myca*, expression is induced upon DNA damages in a p53-

dependent manner. P53 is a master regulator of stress, it is activated upon stress signal and acts to induce the expression of target genes to reduce cellular stress (Williams & Schumacher, 2016). Thereby, lnc-*myca* might act in the stress response pathway upon DNA damage. Interestingly, these results are concordant with previously published results in humans and mice where *PVT1* is involved in the stress response pathway in a p53-dependent manner (Olivero et al., 2020).

Partial deletion of *pvt1* impacts the erythroid lineage

LncRNAs are known to be expressed in a tissue-specific manner. As such, we analyzed available scRNA-seq data across mice development (Cao et al., 2019; Pijuan-Sala et al., 2019). Interestingly both available datasets, which include data from mice gastrulation and organogenesis (from E6.5 to E13.5) (Cao et al., 2019; Pijuan-Sala et al., 2019), showed that *mPvt1* is specifically expressed in erythroid lineage during mice development (SuppFig 12G). MYC is also expressed in the erythroid lineage, with reports showing that MYC expression is essential for erythropoiesis (Guo et al., 2009) and other demonstrating that MYC downregulation is indispensable for terminal erythroid maturation (Jayapal et al., 2010). The correct erythroid formation is vital for oxygen transport in the organism, a lower rate of red blood cells could drive anemia. Double homozygous mutants showed a significantly higher proportion of larvae with lower hemoglobin staining (41% vs 17% in WT) at 2 dpf (Fig 12C & D). This could drive the development of anemia in double homozygous mutants. Erythroid are known as the sink of oxidative stress in the organism due to their primary function as oxygen transporter (Kramer et al., 2017; Scott et al., 1991). We tested if oxidative resistance in the double homozygous mutants was impaired. To avoid any bias due to the developmental defects, we selected only nDD embryos at 3 dpf. nDD embryos were more sensitive to oxidative stress (Fig 12E) with roughly two times more embryos (38%) dying upon oxygen peroxide treatment than WT embryos (22%) (Fig 12E). These results show that nDD mutants, while not having visible developmental defects, are more susceptible to oxidative stress.

These results show that a larger proportion of double homozygous mutants might develop anemia and are sensitive to oxidative stress. Interestingly, the proportion of embryos with abnormal hemoglobin staining at 2 dpf (41%) is approximately equal to the proportion of nDD embryos with survival impairment upon oxidative stress (38%). The same observation can be made for the WT embryos (17% vs 22% respectively). This suggests that the embryos with lower hemoglobin staining at 2 dpf are not embryos that will specifically have

developmental defects later in development. Thus, the erythroid impairment upon the RNA motifs deletion might be independent of myc and p53 upregulation.

2.1.3 Discussion

Here, we established the first *in vivo* model to study *PVT1* lncRNA functionality in physiology. We showed that vertebrate *PVT1* homologous transcripts share two deeply conserved RNA motifs. Interestingly, these two motifs were present multiple times on the different *PVT1* homologous transcripts showing their putative involvement as protein binding sites. While some motifs are distant, the secondary structure of the RNA might bring the motifs in the vicinity of each other. These conserved motifs could drive a conserved molecular function of *PVT1* in vertebrates as no alignable sequence conservation is detected in vertebrate *PVT1* homologous transcripts (Figure 13).

The specific genetic ablation of the region containing the motifs allows a minimally invasive approach. It is known that the 5' part of the *PVT1* locus contains enhancers that drive *MYC* and *PVT1* expression (Cho et al., 2018). Therefore, selectively impaired *PVT1* expression without impacting *MYC* expression is challenging but our approach allowed it. Remarkably, the remaining parts of the transcripts are still expressed, indicating that the deletion did not impact the transcription of the locus.

Surprisingly, we observed some variability in the number of retrieved fish at 2-months-old in each genotype, with the most variable genotype being the lnc-*myca* heterozygous and lnc-*mycb* homozygous mutants. This can be due to the difference in the parents' ages used for the independent crosses. Indeed, the impaired survival was dependent on the age of the parents.

It is known that lncRNAs are expressed in a tissue-specific manner. In this study, we did not determine the tissues where the lnc-*myc* RNAs were expressed in zebrafish. The relatively low expression of the lnc-*myc* RNAs made the detection of the transcripts, by in situ hybridization, challenging. Hence, scRNA-seq would be required to determine the tissue where these transcripts are expressed. While some scRNA-seq are available from different stages of zebrafish development, the lnc-*myc* genes are not annotated in the zebrafish latest genome release, limiting the use of these data. It can be surprising to observe multiple developmental

defects in a multicellular organism upon the partial deletion of the zebrafish *pvt1* lncRNAs. Interestingly, previous reports already demonstrated that the upregulation of the p53 pathway and apoptosis during zebrafish development led to multiple developmental defects (Chakraborty et al., 2009; Danilova et al., 2010; Liu et al., 2003; Plaster et al., 2006).

The observed phenotype might be therefore the consequences of the p53 pathway upregulation upon the partial lnc-*myc* deletion. Indeed, we observed an increase of the p53 pathway prior to the visible developmental defects. A rescue experiment, injecting tp53 morpholino, would be required to confirm it. This experiment was not performed in our study as we noticed that the proportion of double homozygous mutants with developmental defects was highly variable between the parents crossed, making the rescue experiments challenging. This variation could be due to the selection of escapees "resistant" to the p53 upregulation. We observed a sex ratio bias in the double homozygous mutants with a higher proportion of males than females at the adult stage. This is concomitant with previous results showing that upregulation of the p53 pathway and apoptosis during sex reversal led to a sex ratio bias in zebrafish towards males (Rodríguez-Marí et al., 2010). Suggesting that the fish recovered at the adult stage are escapees "resistant" to the upregulation of the p53 pathway leading to a higher proportion of embryos that do not have developmental defects in the progeny.

Interestingly, we observed a correlation between the upregulation of the myc proteins level and the lnc-*myc* expression. However, a western blot of the myc protein in nDD embryos would be required to confirm it. This could be due to the induction of the lnc-*myc* expression by myc as we identified myc binding motifs at the promoters of the lncRNAs. Moreover, the lnc-*myc* transcripts may participate in the stabilization of the myc proteins. Indeed, it was previously shown that *PVT1* stabilizes MYC at the protein level (Alessio et al., 2019; J. Lee et al., 2021; Tseng et al., 2014; Tseng & Bagchi, 2015). Here we can suggest that the deletion did not impact the *pvt1*/myc interaction, showing positive feedback between lnc-*myc* lncRNAs and myc proteins as in human cancer (Tseng & Bagchi, 2015). A complete inhibition of the zebrafish *pvt1* expression would be required to confirm the involvement of the lncRNAs in myc proteins stabilization in zebrafish.

This study represents the first *in vivo* analysis of the physiological function of the cancer-associated lncRNA *PVT1*. Deletion of the deeply conserved RNA motifs leads to developmental defects. Interestingly these defects might be due to the higher myc proteins level in mutants and the p53 pathway upregulation. However, scRNA-seq would be required as

conclusions from a multicellular organism are challenging. We found that features and function of the lncRNA *PVT1* might be conserved from humans to the zebrafish with: (1) the presence of myc recognition sites at the promoters of the lnc-*myc* genes, showing the possible involvement of the lnc-*myc* RNA in the myc pathway; (2) the involvement of the lnc-*myc*a transcript in the p53 pathway upon DNA damage; (3) the partial deletion of the lnc-*myc* RNA impacting the development of the erythroid lineage (Figure 13).

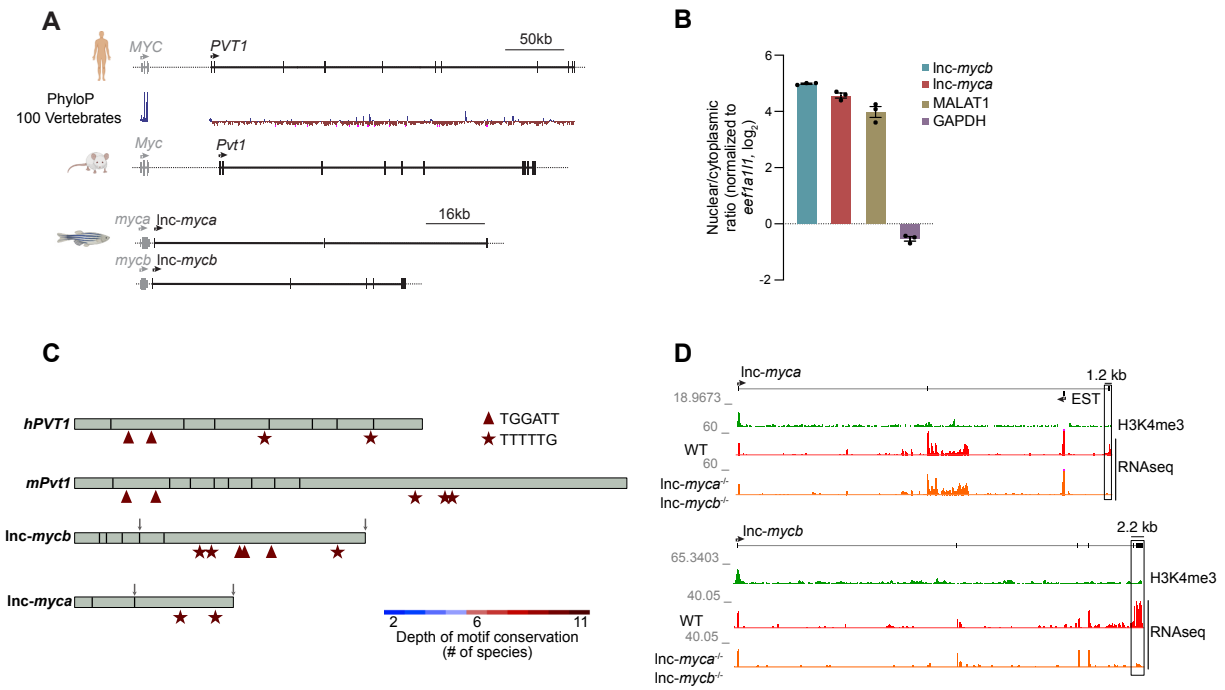
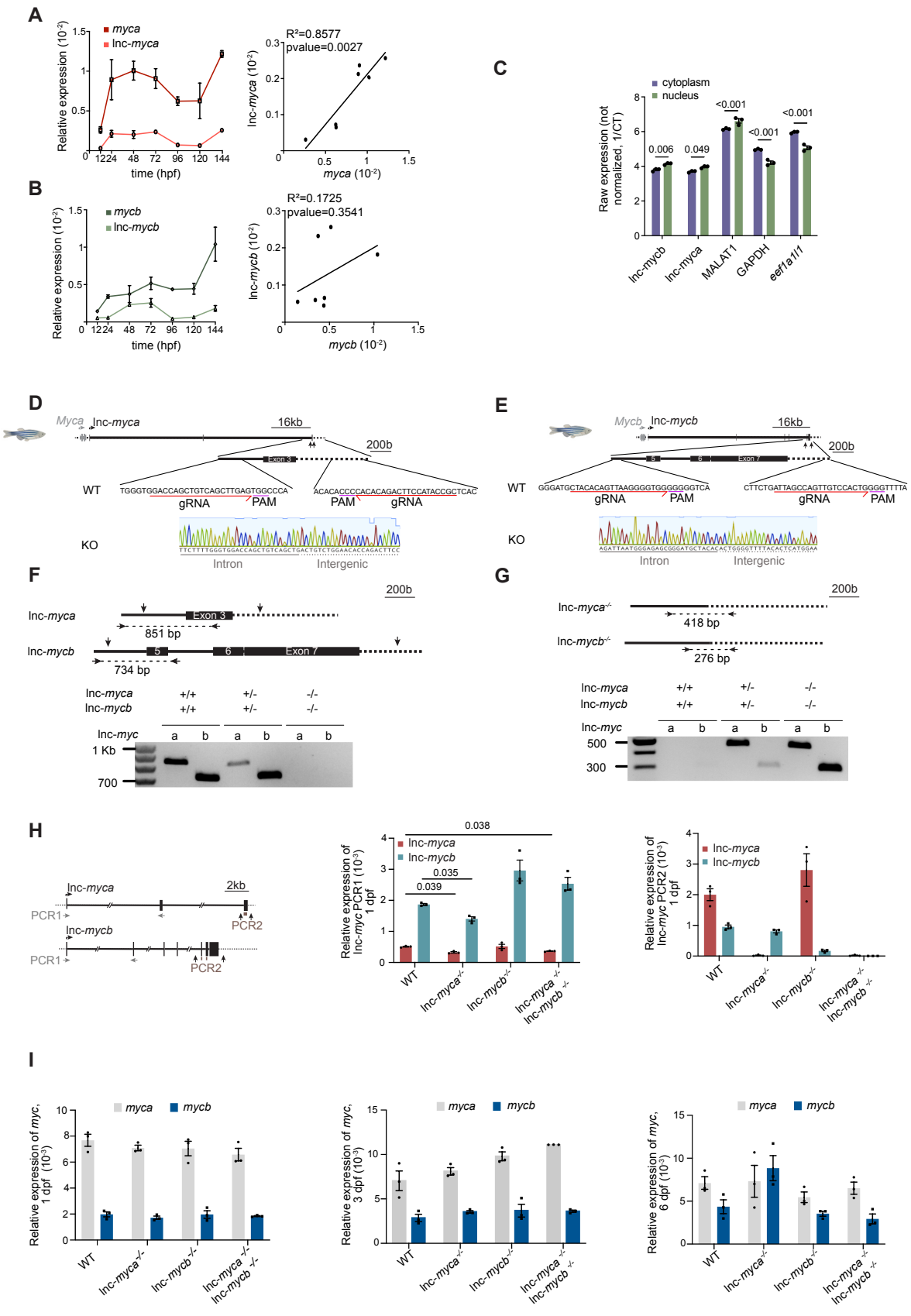


Figure 8: *PVT1* harbors two deeply conserved RNA motifs

A. Schemes representing the *MYC*/*PVT1* locus in humans, mice, and zebrafish. The PhyloP track was obtained on the UCSC Genome Browser with the "Vertebrate Multiz Alignment & Conservation (100 species)" track. **B.** Ratio of the nuclear over the cytoplasmic qRT-PCR fraction. The data presented on the graph are normalized. **C.** Localization of the motifs identified by the lncLOOM algorithm on the human, mice, and zebrafish transcripts. The vertical black arrows represent the position of the guide RNAs (gRNAs) used to delete the region of the zebrafish transcripts containing the conserved motifs. **D.** Genomic locus of the zebrafish *pvt1* genes. The H3K4me3 tracks were obtained from Zhang et al., 2014. The RNA-seq track was obtained from our RNA-seq, uploaded in the UCSC Genome Browser. The vertical boxes represent the regions deleted. The scale bars represent the size of the deleted regions.


A
myca
lnc-myca
Relative expression (10-2)
time (hpf)
R²=0.8577
pvalue=0.0027
lnc-myca (10-2)
myca (10-2)
B
mycb
lnc-mycb
Relative expression (10-2)
time (hpf)
R²=0.1725
pvalue=0.3541
lnc-mycb (10-2)
mycb (10-2)
C
cytoplasm
nucleus
Raw expression (not normalized, 1/CT)
0.006
0.049
<0.001
<0.001
<0.001
lnc-mycb
lnc-myca
MALAT1
GAPDH
eef1a1l1
D
Myca
lnc-myca
16kb
200b
Exon 3
WT
TGGGTGGACCAGCTGTCAGCTTGAGTGGCCCA
ACACACCCCACACAGACTTCCATACCGCTCAC
gRNA
PAM
PAM
gRNA
KO
Intron
Intergenic
E
Mycb
lnc-mycb
16kb
200b
5
6
Exon 7
WT
GGGATGCTACACAGTTAAGGGGTGGGGGGGTCA
CTTCTGATTAGCCAGTTGTCCACTGGGGTTTTA
gRNA
PAM
gRNA
PAM
KO
Intron
Intergenic
F
200b
lnc-myca
Exon 3
851 bp
lnc-mycb
5
6
Exon 7
734 bp
lnc-myca +/+ +/- -/-
lnc-mycb +/+ +/- -/-
lnc-myc a b a b a b
1 Kb
700
G
200b
lnc-myca-/-
418 bp
lnc-mycb-/-
276 bp
lnc-myca +/+ +/- -/-
lnc-mycb +/+ +/- -/-
lnc-myc a b a b a b
500
300
H
lnc-myca
2kb
PCR1
PCR2
lnc-mycb
PCR1
PCR2
Relative expression of lnc-myc PCR1 (10-3) 1 dpf
0.038
0.035
0.039
lnc-myca
lnc-mycb
WT
lnc-myca-/-
lnc-mycb-/-
lnc-myca -/- lnc-mycb -/-
Relative expression of lnc-myc PCR2 (10-3) 1 dpf
I
Relative expression of myc. 1 dpf (10-3)
Relative expression of myc. 3 dpf (10-3)
Relative expression of myc. 6 dpf (10-3)
myca
mycb
WT
lnc-myca-/-
lnc-mycb-/-
lnc-myca -/- lnc-mycb -/-

Supplementary Figure 8: Genetic removal of the RNA motifs in zebrafish *pvt1* transcripts

A-B. qRT-PCR and linear regression of lnc-*myca* and *myca* (A), and lnc-*mycb* and *mycb* (B) during zebrafish development. n=3 biological replicates per developmental stage. Correlation test with the Pearson coefficient. **C.** Raw qRT-PCR data from subcellular fractionation. The data are represented as 1/CT_values. n=3 independent biological replicates, Two-way ANOVA, and multiple T-tests with Sidak correction. **D-E.** Schematic representation of the RNA motifs deletion on lnc-*myca* (D) and lnc-*mycb* (E) genes with the gRNA sequences. The sequencing results are a screenshot of Sanger sequencing quality file. **F-G.** Genotyping strategy by PCR of the WT allele (F) and the mutant allele (G). The vertical black arrows show the position of the gRNAs, and the horizontal arrows with dotted lines represent the position of the amplicon for the PCR. **H.** qRT-PCR strategy to assess the lnc-*myc* RNA expressions. The vertical black arrows show the position of the gRNAs, and the horizontal arrows illustrate the position of the primers for qRT-PCR. The PCR1 amplicon is within the 5' part of the transcripts, and the PCR2 amplicon is within the deleted region. The graphics represent the qRT-PCR of the lnc-*myc* transcripts at 1 dpf in the given genotype. n=3 independent biological replicates, Two-way ANOVA, and multiple T-tests with Dunnett correction. **I.** qRT-PCR of *myc* transcripts at different developmental stages (1, 3, and 6 dpf) in the given genotype. n=3 independent biological replicates, Two-way ANOVA and multiple T-tests with Dunnett correction.

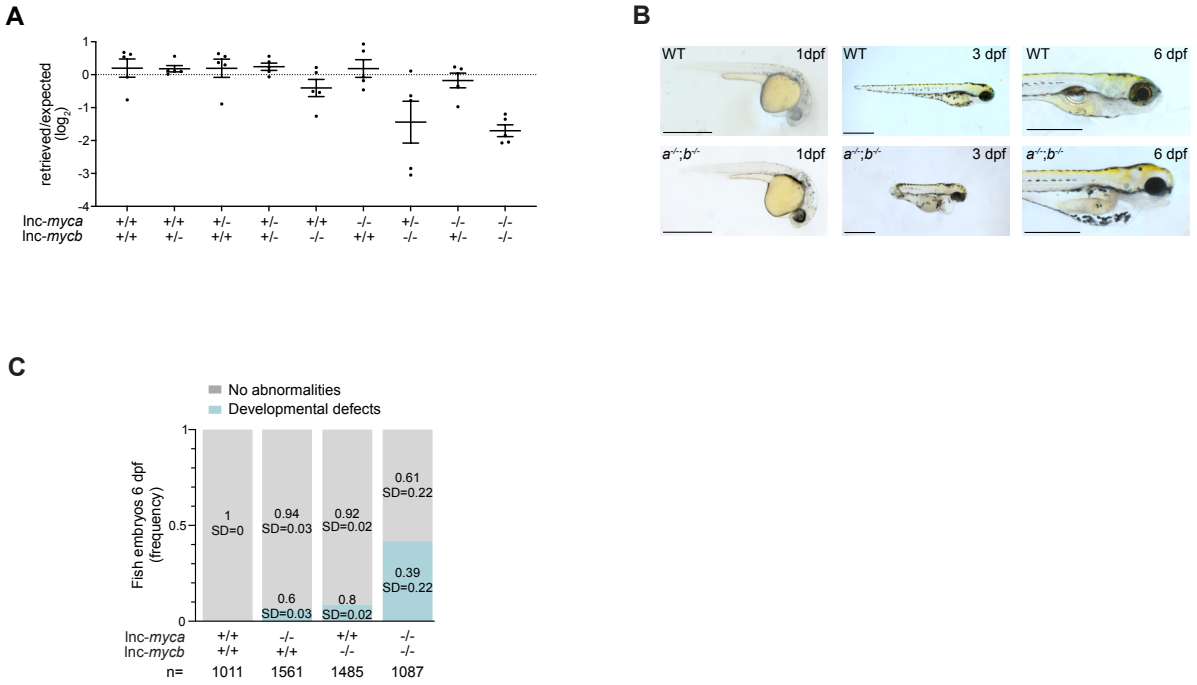
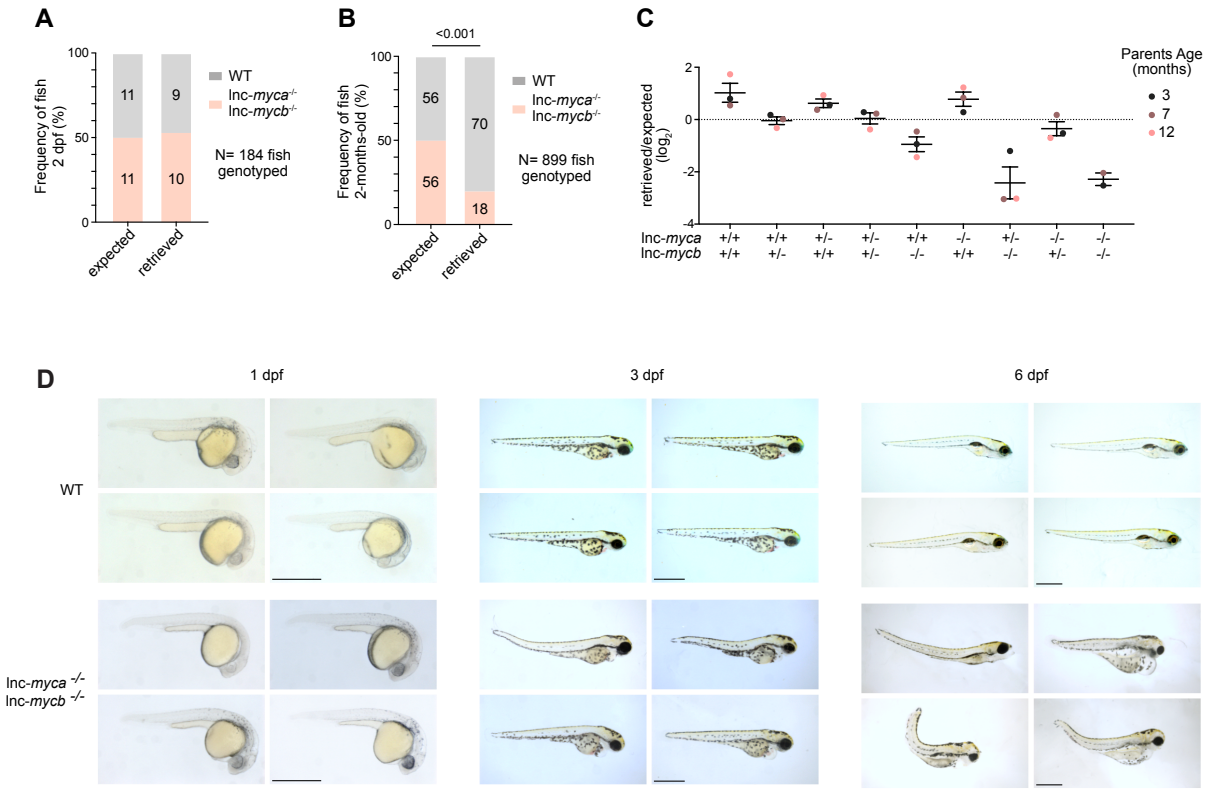


Figure 9: Double homozygous mutants develop striking developmental defects

A. Ratio of the number of fish retrieved over the number of fish expected at the adult stage from a double heterozygous cross. The dotted line represents the ratio value where the retrieved and expected numbers are equal. n=5 independent crosses. One-way ANOVA F(8,27)=2.912 P=0.0177. **B.** Representative pictures of WT and double homozygous mutant embryos at the indicated developmental stage. a stands for lnc-*myca* and b stands for lnc-*mycb*. The scale bar is 0.54 mm. **C.** Frequency of embryos with developmental defects or with no abnormalities in the given genotype, at 6 dpf. The total number of embryos analyzed is indicated underneath the respective genotype. This observation was performed over 3 independent crosses of each genotype, SD represents the variability of frequency between the crosses.



Supplementary Figure 9: Decreased survival to adulthood of the double homozygous mutants

A-B. Frequency of WT and double homozygous mutants at 2 dpf (A) and 2-months-old (B). The number of embryos or fish is indicated in the bar plot. Fisher's exact test. **C**. Ratio of the number of fish retrieved over the number of fish expected upon aging of the parents, at the adult stage. The dotted line represents the ratio value where the retrieved and expected numbers of fish are equal. n=3 independent crosses of the same parents. **D.** Representative pictures of WT and double homozygous embryos at 1, 3, and 6 dpf. The scale bar is 0.54 mm.

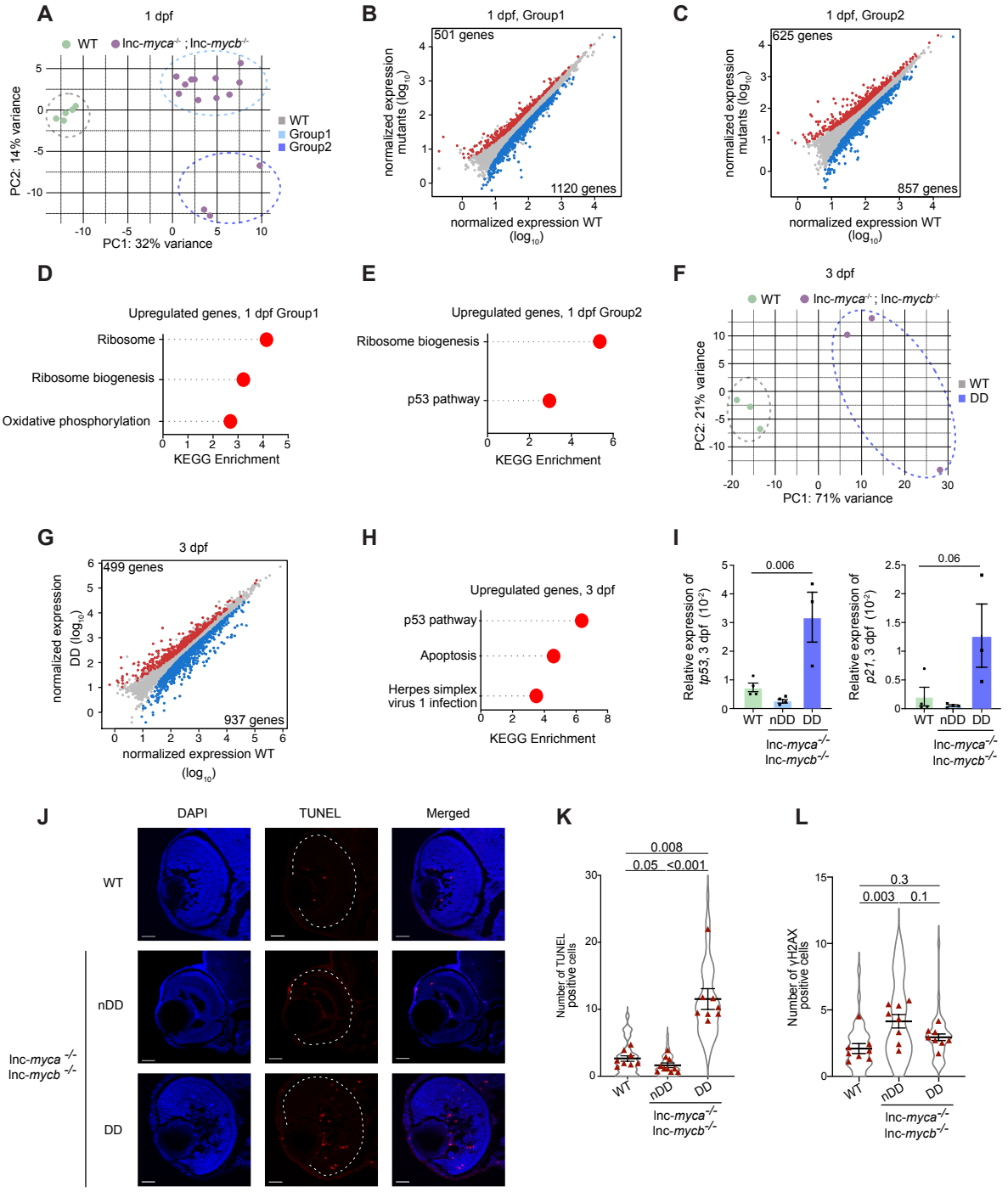
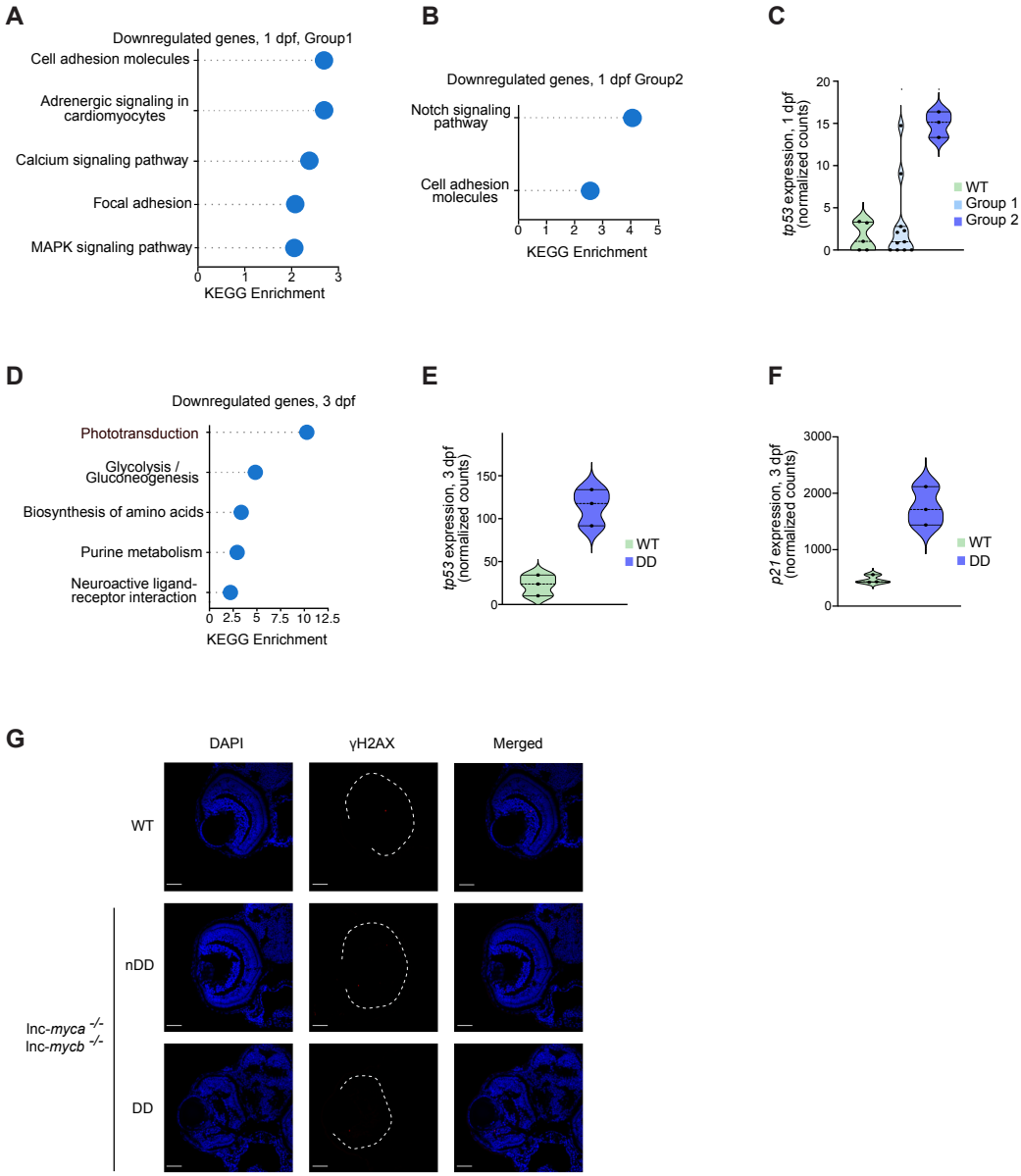

A
1 dpf
WT
lnc-myca-/- ; lnc-mycb-/-
PC2: 14% variance
PC1: 32% variance
WT
Group1
Group2
B
1 dpf, Group1
501 genes
1120 genes
normalized expression mutants (log10)
normalized expression WT (log10)
C
1 dpf, Group2
625 genes
857 genes
normalized expression mutants (log10)
normalized expression WT (log10)
D
Upregulated genes, 1 dpf Group1
Ribosome
Ribosome biogenesis
Oxidative phosphorylation
KEGG Enrichment
E
Upregulated genes, 1 dpf Group2
Ribosome biogenesis
p53 pathway
KEGG Enrichment
F
3 dpf
WT
lnc-myca-/- ; lnc-mycb-/-
PC2: 21% variance
PC1: 71% variance
WT
DD
G
3 dpf
499 genes
937 genes
normalized expression DD (log10)
normalized expression WT (log10)
H
Upregulated genes, 3 dpf
p53 pathway
Apoptosis
Herpes simplex virus 1 infection
KEGG Enrichment
I
0.006
Relative expression of tp53, 3 dpf (10-2)
0.06
Relative expression of p21, 3 dpf (10-2)
WT
nDD
DD
lnc-myca-/-
lnc-mycb-/-
J
DAPI
TUNEL
Merged
WT
nDD
DD
lnc-myca -/-
lnc-mycb -/-
K
0.008
0.05
<0.001
Number of TUNEL positive cells
WT
nDD
DD
lnc-myca-/-
lnc-mycb-/-
L
0.3
0.003
0.1
Number of γH2AX positive cells
WT
nDD
DD
lnc-myca-/-
lnc-mycb-/-

Figure 10: The developmental defects are linked with the p53 pathway and apoptosis upregulation

A. Representation of the PCA from single embryo RNA-seq at 1 dpf. The colors of the dots determine the genotype of the biological replicate, the dotted circles represent the three clusters. **B-C.** Scatter plots showing the averaged gene expression in the WT embryos correlated with the averaged gene expression of the embryos from group 1 (B) and group 2 (C) of the double homozygous mutants. the red dots represent the genes significantly upregulated and in blue the significantly downregulated. **D-E.** The graphics illustrate the pathways significantly enriched, using the KEGG dataset. Only significantly upregulated genes were used from group 1 (D) and group 2 (E). **F.** Representation of the PCA from the RNA-seq at 3 dpf. The colors of the dots determine the genotype of the biological replicate, the dotted circles represent the clusters. **G.** Scatter plots showing the averaged gene expression of the WT embryos correlated with the averaged gene expression of the double homozygous embryos. the red dots represent the genes significantly upregulated and in blue the significantly downregulated. **H.** Graphic illustrating the pathways significantly enriched at 3 dpf, using the KEGG dataset. Only significantly upregulated genes from the RNA-seq were used for this analysis. **I.** qRT-PCR of genes from the p53 pathway in the WT embryos, nDD, and DD double homozygous mutants at 3 dpf. One-way ANOVA and multiple T-test with Dunnett correction. **J.** Representative pictures of TUNEL assay of zebrafish embryo retina at 3 dpf. The scale bar is 25 microns. **K-L.** Quantification of TUNEL positive cells (K) and γH2AX positive cells (L). The violin plots represent the variation of the number of positive cells per slice (4-6 per biological replicates) and the red dots are the average of positive cells per biological replicates. One-way ANOVA, multiple T-test with Dunnett correction. The statistical tests were performed on the biological replicates.



Supplementary Figure 10: p53 is upregulated before the appearance of the developmental defects at 1 dpf

A-B. Graphics illustrating the pathways significantly enriched, using the KEGG dataset. Only significantly downregulated genes were used for this analysis from group 1 (A) and group 2 (B). **C.** Violin plot of the *tp53* expression with normalized counts from the RNA-seq at 1 dpf. **D.** The graph illustrates the pathways significantly enriched at 3 dpf, using the KEGG dataset. Only significantly downregulated genes from the RNA-seq were used for this analysis at 3 dpf. **E-F.** Violin plot of the normalized counts from the RNA-seq of *tp53* (E) and *p21* (F) transcripts at 3 dpf. **G.** Representative pictures of γH2AX immunostaining of zebrafish embryo retina at 3 dpf. The scale bar is 25 microns.

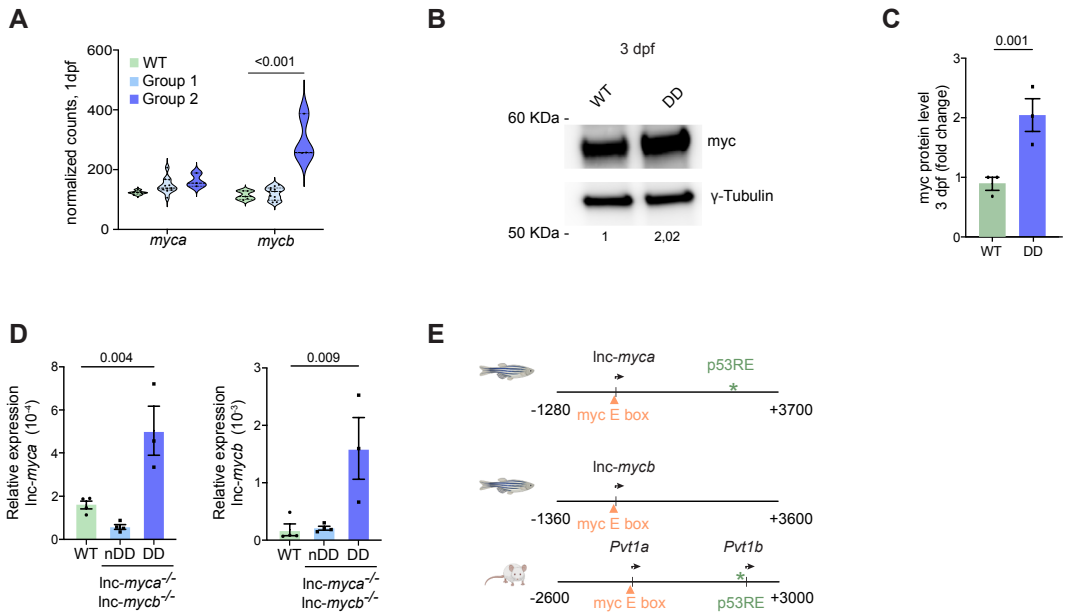
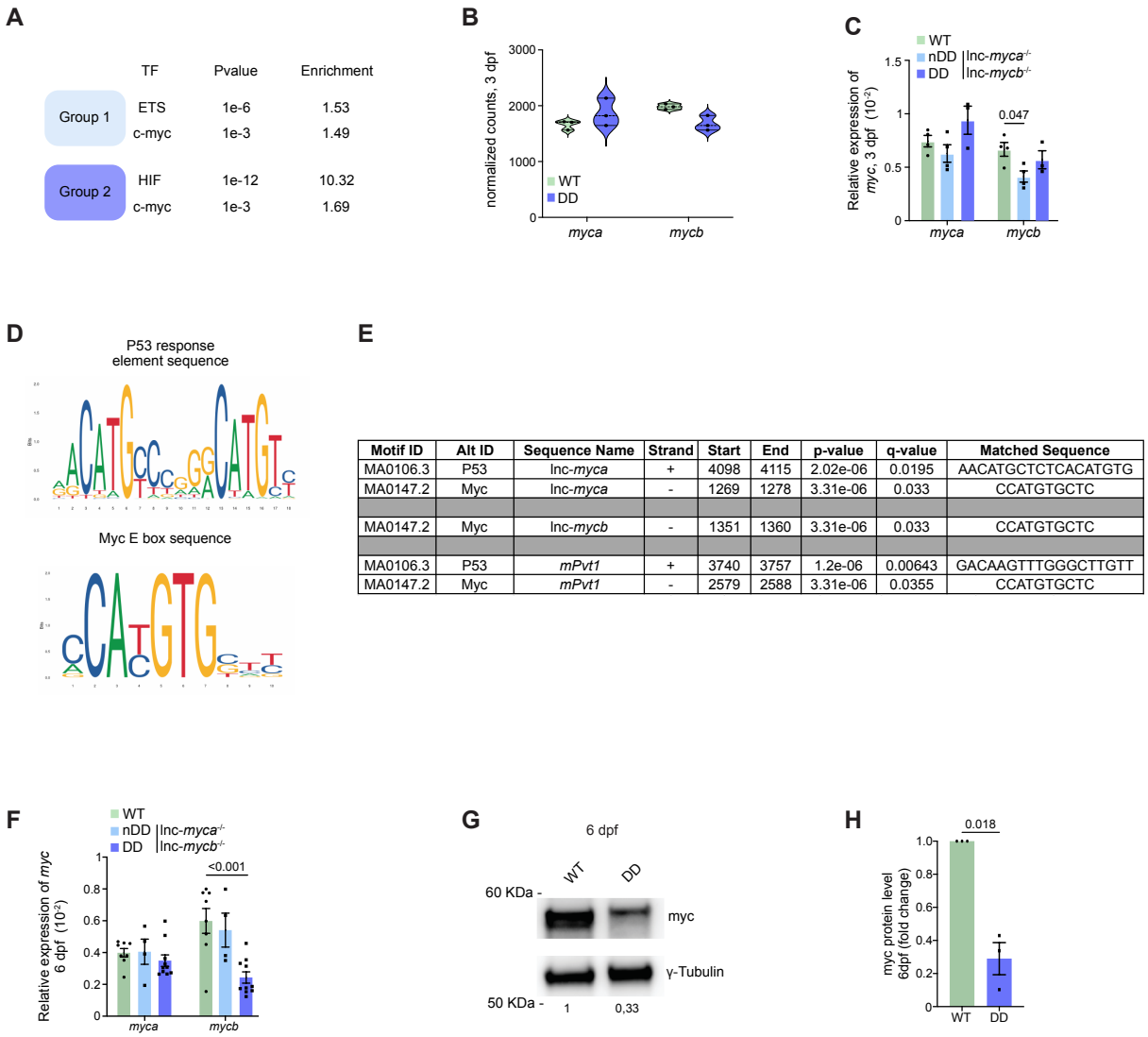


Figure 11: The increased myc protein level and lnc-*myc* lncRNA expressions were correlated with the onset of the developmental defects in mutants

A. Violin plot of the normalized counts from the RNA-seq at 1 dpf of the myc transcripts. **B.** Western blot of the myc proteins in the WT and DD embryos at 3 dpf. The quantification of myc, normalized to γ-Tubulin, is indicated underneath. **C**. Quantification of the western blots, n=3 independent experiments. T-test. **D.** qRT-PCR of the lnc-*myc* transcripts in the WT embryos, nDD, and DD double homozygous mutants at 3 dpf. One-way ANOVA and multiple T-test with Dunnett correction **E.** Schematic illustrating the transcription factor motifs found in the vicinities of the zebrafish and mouse *Pvt1* promoters.



Motif ID	Alt ID	Sequence Name	Strand	Start	End	p-value	q-value	Matched Sequence
MA0106.3	P53	lnc-*myca*	+	4098	4115	2.02e-06	0.0195	AACATGCTCTCACATGTG
MA0147.2	Myc	lnc-*myca*	-	1269	1278	3.31e-06	0.033	CCATGTGCTC
MA0147.2	Myc	lnc-*mycb*	-	1351	1360	3.31e-06	0.033	CCATGTGCTC
MA0106.3	P53	*mPvt1*	+	3740	3757	1.2e-06	0.00643	GACAAGTTTGGGCTTGTT
MA0147.2	Myc	*mPvt1*	-	2579	2588	3.31e-06	0.0355	CCATGTGCTC

Supplementary Figure 11: The lnc-*myc* expression might be driven by myc

A. Recapitulative table of the promoter analysis of the genes significantly upregulated at 1 dpf in both groups. The table shows the transcription factor (TF), the P-value, and the enrichment over the background. **B.** Violin plot of the normalized counts from the RNA-seq of the *myc* transcripts at 3 dpf. **C.** qRT-PCR of the *myc* transcripts at 3 dpf in the different groups of mutants. Two-way ANOVA, multiple T-test with Dunnett correction. **D.** Sequences of the P53 and Myc binding motifs used in the FIMO algorithm. The motifs were downloaded from the JASPAR motifs database. **E.** Recapitulative table of the results obtained with the FIMO motifs scanning algorithm. **F.** qRT-PCR of the *myc* transcripts at 6 dpf in the different groups of mutants. Two-way ANOVA, multiple T-test with Dunnett correction. **G.** Western blot of the myc proteins in the WT and DD embryos at 6 dpf. The quantification of the myc band, normalized with the γ-Tubulin band, is indicated at the bottom. **H.** Quantification of the western blots, n=3 independent experiments. Welch T-test.

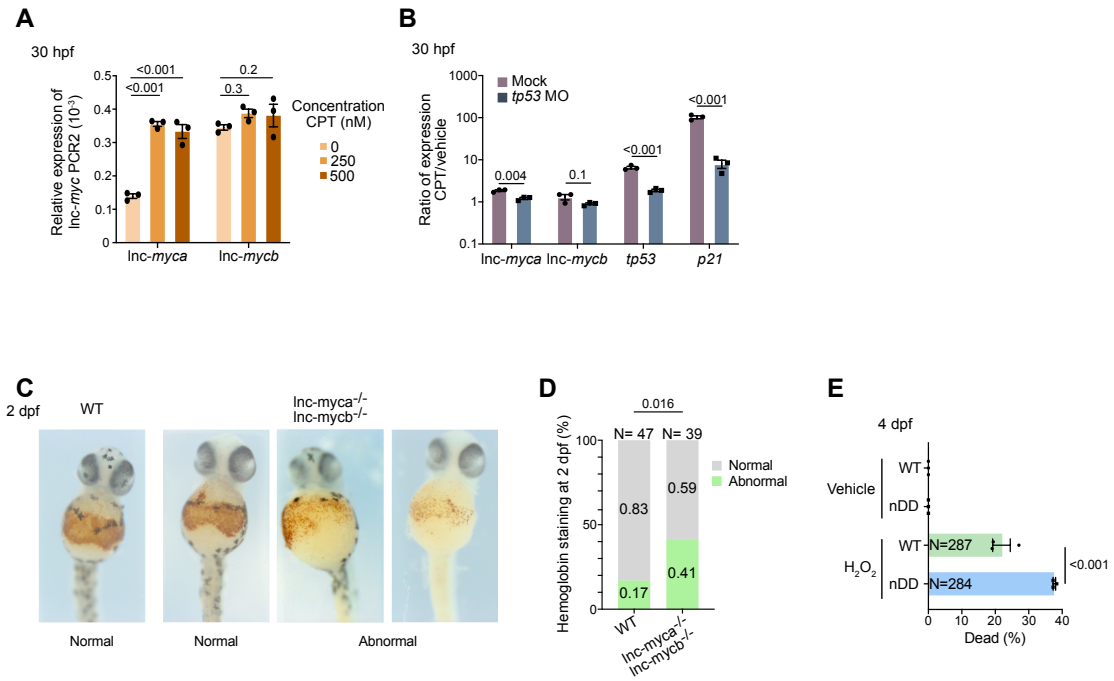
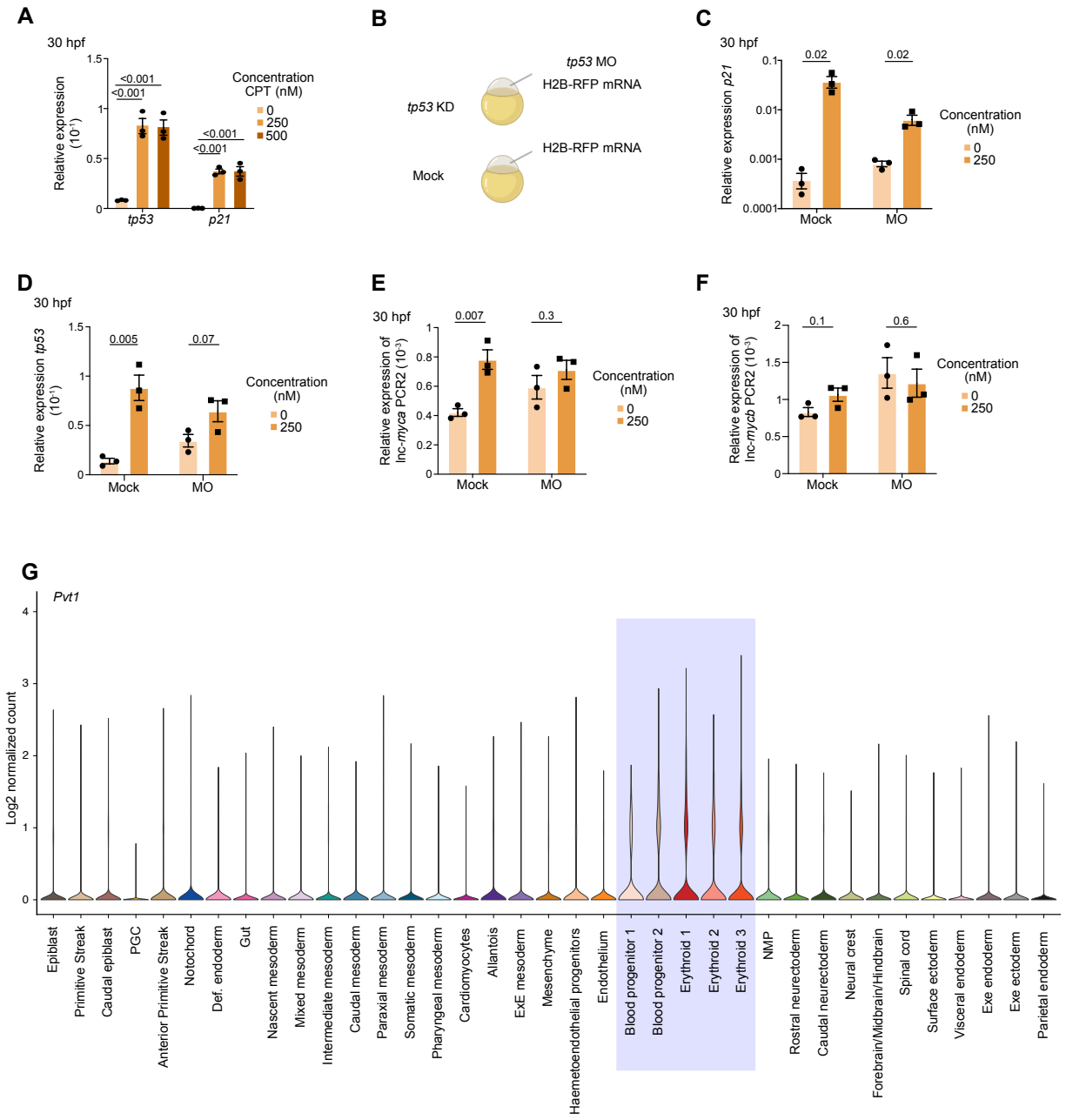


Figure 12: Double homozygous mutants develop anemia

A. qRT-PCR of the lnc-*myc* transcripts at 30 hpf at the indicated CPT concentration. Two-way ANOVA and multiple T-tests with Dunnett correction. **B.** Ratio of expression of the treated over the control condition in mock or p53 morpholino injected embryos. Multiple T-test with Holm-Sidak correction. **C.** Representative pictures of 2 dpf embryos with hemoglobin staining. The pictures illustrate normal and abnormal hemoglobin staining. **D.** Quantification of hemoglobin staining into normal or abnormal staining in the WT and double homozygous embryos. The proportion of embryos is indicated within the bar plot and the total number of examined embryos is indicated on top of the bar plot. Fisher's exact test. **E.** Proportion of dead embryos after that 3 dpf embryos were incubated in vehicle or H202. The total number of analyzed embryos is indicated in the graph. n=3 independent experiments. One-way ANOVA, multiple T-test with Tukey correction.



Supplementary Figure 12: lnc-*myca* expression is driven by p53 in stress-induced conditions

A. qRT-PCR of *tp53* and *p21* transcripts at 30 hpf at the indicated conditions. Two-way ANOVA and multiple T-tests with Dunnett correction. **B.** Scheme representing the morpholino injection strategy. **C-F.** qRT-PCR of *p21* (C), *tp53* (D), lnc-*myca* (E), and lnc-*mycb* (F) at 30 hpf embryos mock or morpholino injected and treated with the indicated CPT concentration. Multiple T-test with Holm-Sidak correction. **G.** Violin plot of m*Pvt1* expression in the different cell types identified by scRNA-seq. The graph was downloaded on https://marionilab.cruk.cam.ac.uk/MouseGastrulation2018/. The analyzed dataset is coming from Pijuan-Sala et al., 2019.

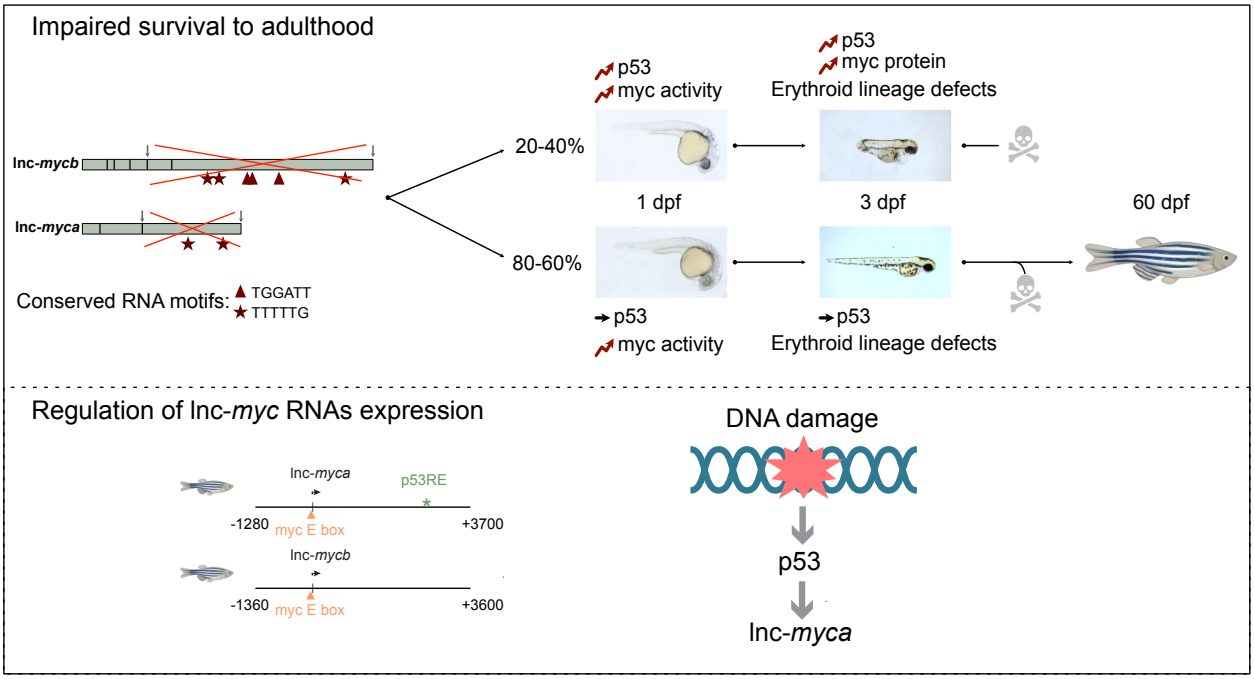


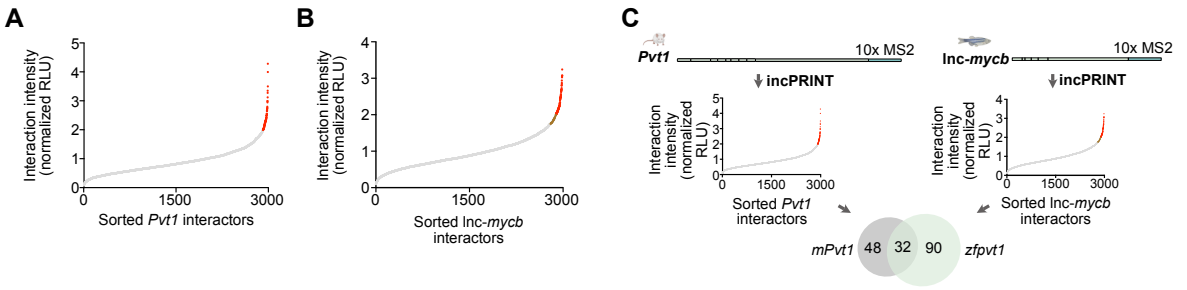
Figure 13: Schematic representing the results of the study

3. ADDITIONAL RESULTS

3.1 The deeply conserved RNA motifs drive the interaction of the *PVT1* lncRNA with key DNA repair factors

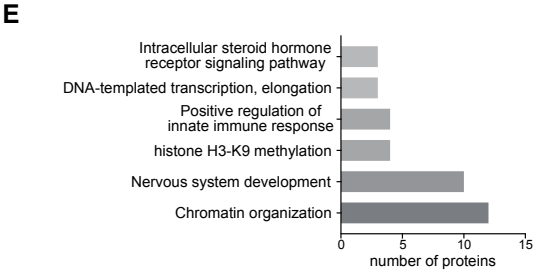
We showed that vertebrate *PVT1* homologous transcripts share deeply conserved k-mers. These short RNAs motifs recently emerged as an alternative to sequence conservation driving conserved molecular mechanisms of lncRNAs across species (Ross et al., 2021). LncRNAs act mostly through protein interactions and the identified k-mers on *PVT1* could correspond to protein binding sites. Thus, to analyze the *PVT1* protein interactome in vertebrates and the proteins that specifically interact with the identified k-mers, we used our newly published technology in cell protein-RNA interaction (incPRINT) (Graindorge et al., 2019).

incPRINT is a powerful tool that allows assessing the interaction between one RNA of interest and ~3000 proteins (Graindorge et al., 2019). We established the proteins interacting with the mouse *Pvt1* transcript (Fig 14A) and the zebrafish transcript, lnc-*mycb* (Fig 14B). Common putative binders were found between the two transcripts (Fig 14C), allowing the specific identification of conserved interactors. In total, 32 proteins were identified to interact with both transcripts (Fig 14C). These proteins were enriched for functions that are mainly nuclear (e.g., chromatin organization, H3K9 methylation) (Fig. 14D & E), in accordance with the previously established subcellular localization of the mammalian and zebrafish *PVT1* transcripts. In order to decrypt the specific interaction driven by the deeply conserved RNA motifs, we scrambled these motifs and analyzed the interaction score of the proteins identified (Fig 14F & G). The first motif, TGGATT was scrambled to GTATGT and the second from TTTTTG to TCGACG. Interestingly while both motifs are different, the highest drop in the interaction score upon the scrambling was associated with the same protein for both motifs: RAD51C (Fig 14F & G). Intriguingly, the decrease of the interaction score, for the first motif, was approximatively similar between the RAD51C and NUSAP1 proteins (Fig 14F). These two proteins are key regulators of the DNA repair pathway (Kotian et al., 2014; Somyajit et al., 2012). The scrambling of both motifs also identified the specific binding of the UPF3B protein (Fig 14F & G). However, this protein is involved in the Nonsense-Mediated Decay pathway, occurring in the cytoplasm (Cui et al., 1995; He et al., 1997).



D

Transcription factors	TOX2 / HMGB4 / ZNF85 / PRDM5
DNA repair	ERCC6 / RAD51C
Histone epigenetic regulators	SETD5 / JARID2 / RPS6KA5 / SCMH1 / SETDB1 / RING1
Epigenetic factor	SSRP1
RNA processing	UPF3B / MOV10 / NOL6 / XRN2 / AGO2
Transcriptional regulators	AFF1 / NR0B1 /MED1 / SMARCE1
Immunology response	MACIR / IFI16
Vesicle transport	SLC18A03
Microtubule organization	NUSAP1
Proteasome factor	PSMD12
Cytoplasmic receptor	NLRP11



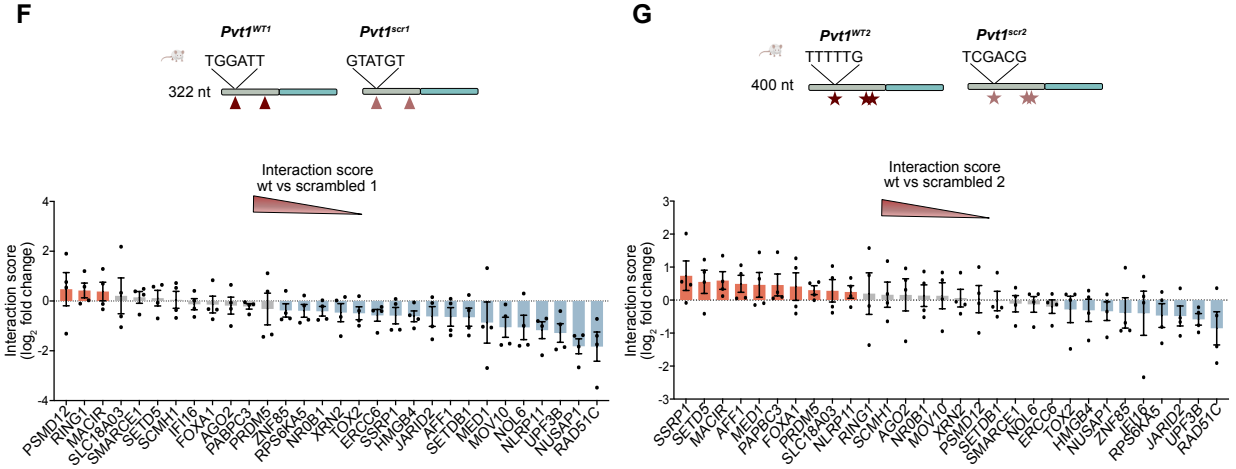


Figure 14: *PVT1* RNA motifs interact with key DNA repair factors

A-B. Schematic representation of the protein interactions with the mouse transcript (A) and the zebrafish transcript (B) determined by incPRINT. The interactors are represented with colored dots. **C.** Schematic representation of the intra-species interactome comparison between the mice and zebrafish transcripts. The interactors for each transcript are represented by the colored dots in the graph. The number of proteins is indicated within the Wienn diagram. **D.** List of common interactors ranked by their primary function. **E.** Pathway analysis, using the KEGG dataset, of the 32 common interactors between the *PVT1* transcripts. The pathways represented are significantly enriched over the background. **F-G.** Schematic representation of the transcripts used to investigate the specific interactor of the RNA motifs. The graphs show the ratio of interaction score between the WT and scrambled motif 1 (F) and motif 2 (G). In red are highlighted the proteins where the interaction score is higher upon scrambling of the motifs and in blue are the protein where the interaction score is decreased upon scrambling of the motifs.

PVT1 interactome analysis with incPRINT allowed the identification of protein interactors conserved from mice to zebrafish. Moreover, the specific examination of protein interaction driven by the RNA motifs showed that RAD51C and NUSAP1 interact specifically with the RNA motifs harbored on *PVT1*. Interestingly, since these two proteins are essential DNA repair factors (Kotian et al., 2014; Somyajit et al., 2012), it suggests a role for *PVT1* in this process. This is concordant with previously published reports and our findings showing that *PVT1* expression responds to DNA damages in a p53-dependent manner (Olivero et al., 2020). These interactions would need to be confirmed through orthogonal methods like RNA immunoprecipitation.

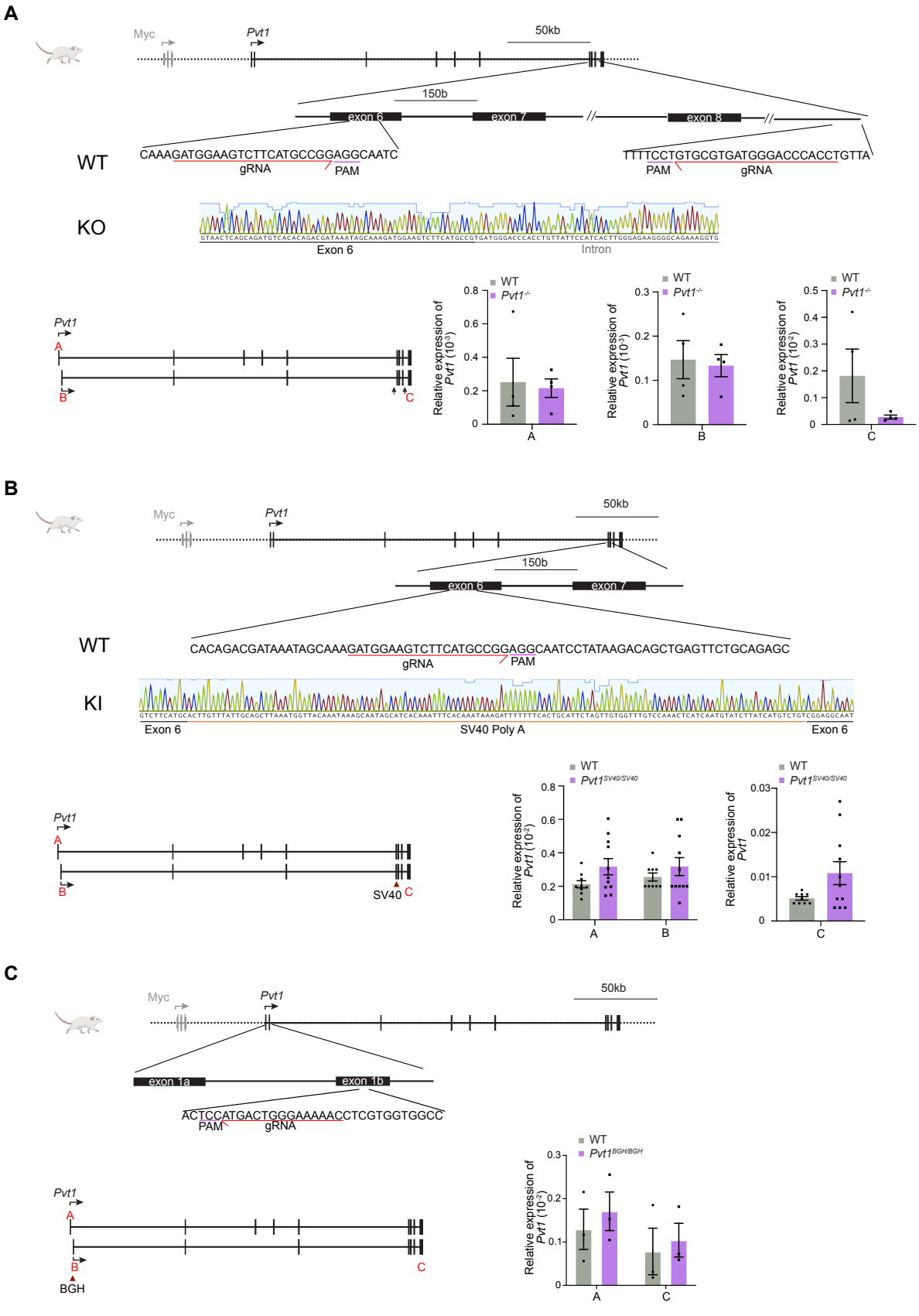
3.2 *PVT1* has a limited impact on pluripotency

The lncRNA *mPvt1* was previously described as highly expressed in stem cells (Tuck et al., 2018). Moreover, scRNA-seq of mouse embryonic stem cells (mESCs) revealed the high heterogeneity of *mPvt1* expression (Tuck et al., 2018). *mPvt1* is highly expressed in a subpart of the mESCs population, associated with high pluripotency (Tuck et al., 2018). Thereby, we analyzed the involvement of *mPvt1* in pluripotency and cell differentiation.

We used several approaches to inactivate *Pvt1* in mouse ESCs. To avoid impacting *Myc* expression, located upstream, we first performed deletion of the 3' part of the transcript, with CRISPR-Cas9. Interestingly, while we were able to retrieve homozygous mutants, *Pvt1* expression was not impacted, as the 5' part was still stably expressed (Fig 15A). Because we aimed for a full *Pvt1* inactivation, we then pursue with termination signal insertion (Anderson et al., 2016; Ballarino et al., 2018; Bond et al., 2009; Grote et al., 2013; Lavalou et al., 2019; Yamanishi et al., 2018). Interestingly, both BGH and SV40 polyA termination signal insertions did not impact the transcription as we observed readthrough (Fig 15B & C).

Figure 15: *Pvt1* genetic manipulation designs

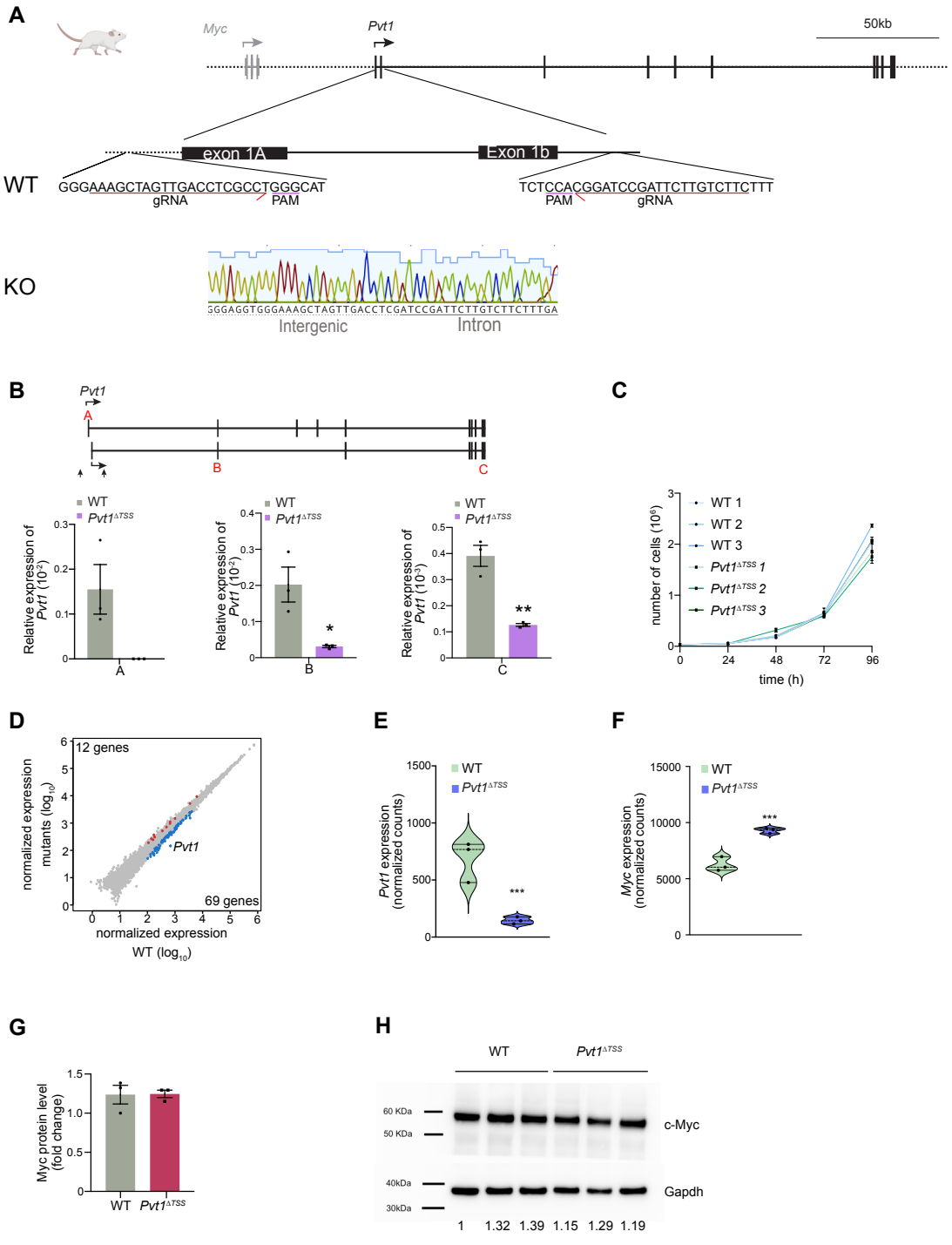
A-C. Schematic representation of the genetic manipulations performed within the mouse *Pvt1* locus with the gRNA sequences outlined. The sequencing results are a screenshot of Sanger sequencing quality file. The small *Pvt1* variants schematics represent the qRT-PCR strategy with the red letters indicating the position of the different qRT-PCR amplicons. The bar plots illustrate the *Pvt1* expression in mESCs of WT and manipulated cell line by qRT-PCR in the indicated region of the transcript. The mESCs where 3' deleted (A), knocked-in with a SV40 polyA sequence (B) or a with a BGH polyA sequence (C).


A
Myc
Pvt1
50kb
150b
exon 6
exon 7
exon 8
WT
CAAAGATGGAAGTCTTCATGCCGGAGGCAATC
gRNA
PAM
TTTTCCTGTGCGTGATGGGACCCACCTGTTA
PAM
gRNA
KO
Exon 6
Intron
Pvt1
A
B
C
WT
Pvt1-/-
Relative expression of Pvt1 (10-3)
B
KI
Exon 6
SV40 Poly A
Exon 6
SV40
Pvt1SV40/SV40
C
ACTCCATGACTGGGAAAAACCTCGTGGTGGCC
exon 1a
exon 1b
BGH
Pvt1BGH/BGH

At last, we performed a minimally invasive genetic removal of the TSS. Two alternative TSSs are observed in mice, hence we deleted both TSSs with a 2 kb deletion (Fig 16A). The TSS deletion did not completely abolish *Pvt1* expression but reduced its expression by at least 70% (Fig 16B). With this approach, we analyzed three independent $Pvt1^{\Delta TSS}$ cell lines. *Pvt1* TSS deletion did not impact in cell proliferation (Fig 16C) and had a limited effect on gene expression (Fig 16D). Indeed, RNA-seq of $Pvt1^{\Delta TSS}$ cells showed only 12 genes upregulated and 69 genes downregulated compared to WT cells. These misregulated genes were slightly impacted, with the most impacted gene being *Pvt1* (Fig 16D & E). We found that as in humans (Cho et al., 2018), the TSS mutation of *Pvt1* led to an upregulation of *Myc* expression at the RNA level (Fig 16F) but upregulation of Myc was not observed at the protein level (Fig 16G & H).

Figure 16: *Pvt1* mutation has a limited impact on mESCs pluripotency

A. Schematic representation of the TSS deletion with the gRNA sequences outlined. The sequencing results are a screenshot of Sanger sequencing quality file. **B.** Representation of the two expressed variants of the mouse *Pvt1* in mESCs. The vertical arrows represent the region targeted for TSS deletion. The red letters show the position of the different qRT-PCR amplicons. The bar plots show the *Pvt1* expression in mESCs in WT and TSS deleted cell line by qRT-PCR in the indicated region of the transcript. **C.** Number of cells in three independent WT and mutant cell lines. The number of cells was assessed every 24 hours for 4 consecutive days. **D.** Scatter plot of gene expression determined by RNA-seq. In blue are the significantly downregulated genes and in red are the significantly upregulated genes. **E-F.** Violin plot of *Pvt1* (D) and *Myc* (E) normalized count from the RNA-seq. **G.** Quantification of the western blot. **H.** Western blot of three independent WT and $Pvt1^{\Delta TSS}$ cell lines. The quantification of the Myc band, normalized to GAPDH intensity, is written underneath the respective lane.


A
Myc
Pvt1
50kb
exon 1A
Exon 1b
WT
GGGAAAGCTAGTTGACCTCGCCTGGGCAT
gRNA
PAM
TCTCCACGGATCCGATTCTTGTCTTCTTT
PAM
gRNA
KO
GGGAGGTGGGAAAGCTAGTTGACCTCGATCCGATTCTTGTCTTCTTTGA
Intergenic
Intron
B
Pvt1
A
B
C
WT
Pvt1ΔTSS
Relative expression of Pvt1 (10-2)
Relative expression of Pvt1 (10-2)
Relative expression of Pvt1 (10-2)
C
number of cells (10^6)
WT 1
WT 2
WT 3
Pvt1ΔTSS 1
Pvt1ΔTSS 2
Pvt1ΔTSS 3
time (h)
D
12 genes
69 genes
Pvt1
normalized expression mutants (log10)
normalized expression WT (log10)
E
Pvt1 expression (normalized counts)
F
Myc expression (normalized counts)
G
Myc protein level (fold change)
H
WT
Pvt1ΔTSS
60 KDa
50 KDa
c-Myc
40kDa
30kDa
Gapdh
1 1.32 1.39 1.15 1.29 1.19

Stem cells have the unique capacity to differentiate into different cell layers. Thus, we initiated neuronal differentiation to analyze the function of *Pvt1* upon differentiation. *Pvt1*$^{\Delta TSS}$ cells did not exhibit impairment in cell differentiation as we observed differentiation into neuronal progenitors, formation of rosettes, and neuronal networks (Fig 17A). Moreover, gene expression analysis by qRT-PCR of classical pluripotency and neuronal markers did not show any differences between *Pvt1*$^{\Delta TSS}$ and WT upon differentiation (Fig 17B-E).

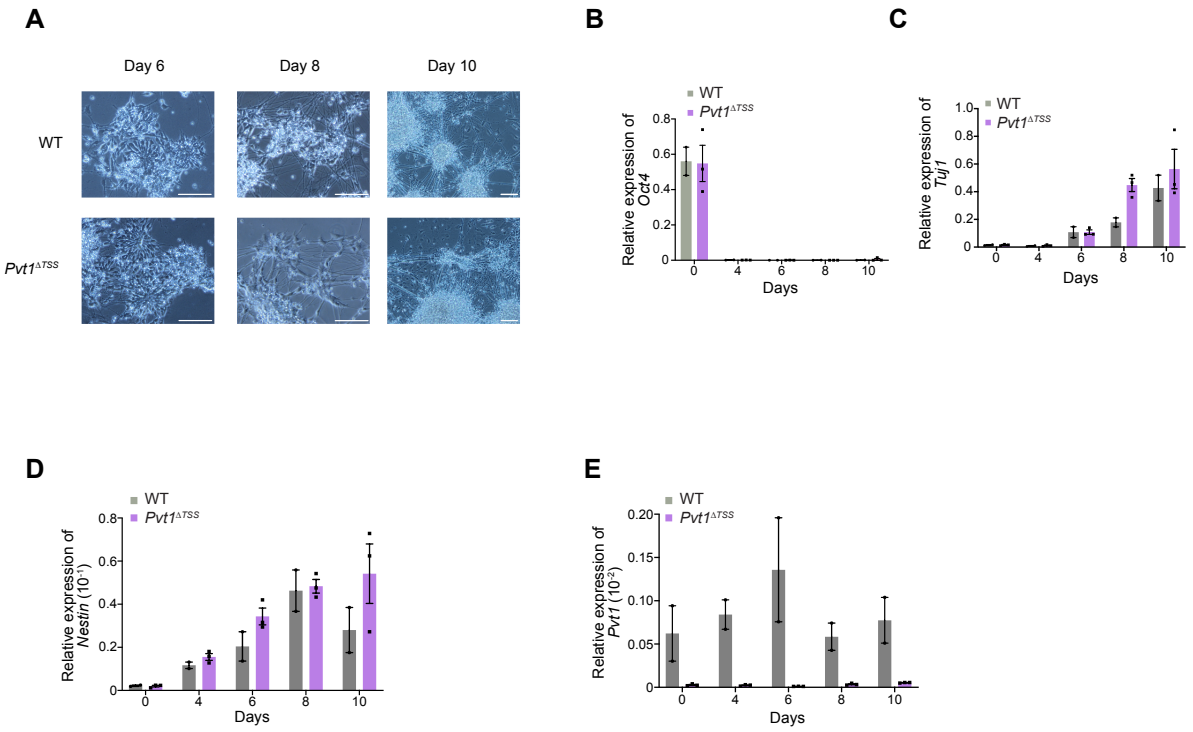


Figure 17: *Pvt1*$^{\Delta TSS}$ did not impact neuronal differentiation

A. Representative pictures of differentiated cells of the indicated cell line at days 6, 8, and 10 of differentiation. The scale bar is 100 microns. **B-E.** qRT-PCR of the expression of *Oct4* (B), *Tuj1* (C) *Nestin* (D) and *Pvt1* (E) across cell differentiation in the WT and mutant cell lines, n=3 independent mutant cell lines and n=2 cell lines for the WT.

Taken together, these results show that while being highly expressed in mESCs, *Pvt1* does not seem to play a role in pluripotency. Moreover, cell differentiation did not show any differences between the WT and mutant cell lines, highlighting the limited impact of *Pvt1* genetic manipulation on pluripotent stem cells. This limited impact can be due to the incomplete

ablation of the *Pvt1* expression. Alternative approaches to study the complete *Pvt1* suppression are limited, as the locus contains enhancers that regulate *myc* expression. In addition, knockdown experiments allowed only a downregulation of the *Pvt1* expression and not complete suppression of *Pvt1* expression.

4. DISCUSSION AND PERSPECTIVES

4.1 Zebrafish as a model organism to study lncRNA functions

LncRNAs have emerged as major regulators of a wide range of biological processes (Constanty & Shkumatava, 2021; Perry & Ulitsky, 2016; Statello et al., 2021; Ulitsky & Bartel, 2013). With the discovery of the first non-coding transcripts with physiological functions (Brockdorff et al., 1991, 1992; Brown et al., 1992; Sleutels et al., 2002), an enthusiastic interest followed in the systematic identification of lncRNAs in different vertebrate species (Hezroni et al., 2015; Pauli et al., 2012; Ulitsky et al., 2011), including mammals (Carninci et al., 2005; Katayama et al., 2004; The ENCODE Project Consortium, 2007). These studies notably revealed the presence of thousands of lncRNA loci, some of which are being conserved from humans to zebrafish (Hezroni et al., 2015).

Studies of lncRNAs in zebrafish were first driven by the urge to decipher their evolutionary trajectory (Bitetti et al., 2018; Lin et al., 2014; Ulitsky et al., 2011). Then, following annotation of lncRNAs in vertebrate species, it was noticed that intergenic lncRNAs were mostly located near genes involved in transcription and development (Amaral et al., 2018; Herriges et al., 2014; Necsulea et al., 2014; Ulitsky et al., 2011). This suggested that these lncRNAs might have a role in developmental processes since lncRNAs can regulate gene expression in *cis* (Gil & Ulitsky, 2020). This feature further put the spotlight on zebrafish as a relevant model to study lncRNA function. Indeed, due to their external fertilization and development, the zebrafish is an ideal model to study development (Goudarzi et al., 2019; Lavalou et al., 2019; Sarangdhar et al., 2018; Wu et al., 2019). Beyond being an appealing developmental model, the zebrafish can also be used to model human diseases (Bradford et al., 2017; Choi et al., 2021), for protein-coding genes misregulated in human diseases (Berger & Currie, 2012; Guella et al., 2011) but also for lncRNAs with potential pathological impact. Similarly, a recent study led to the characterization of novel zebrafish lncRNAs with potential relevance for Alzheimer's disease in humans (Banerjee et al., 2021). Therefore, the zebrafish studies considerably participate in lncRNA characterization in human diseases.

Developmental differences between zebrafish and mammals may also be of interest to addressing lncRNA function. In contrast to eutherians, zebrafish embryos develop externally, without extraembryonic tissues. While this may limit studying lncRNA functions in these specific tissues, it also allows overcoming spontaneous abortion that may result from essential lncRNA function in these tissues. It is therefore an attractive model to separate functional

requirements in extraembryonic and embryonic tissues. In mammals, the *PVT1* lncRNA has been shown to be essential for the viability and proliferation of the trophoblast, an extraembryonic structure that is vital for embryo implantation and the formation of a functional placenta (Latos & Hemberger, 2016). Accordingly, *PVT1* downregulation leads to spontaneous abortion and gestational diabetes (Qin et al., 2019; Wang et al., 2019). Thus, the zebrafish model allows studying *PVT1* functionality using a straightforward genetic manipulation, whereas its study in mammals would require conditional genetic manipulations. During my PhD, I characterized the lncRNA *pvt1* in zebrafish to study the evolutionarily conserved function of this lncRNA in a vertebrate specie. The absence of extraembryonic tissues was a notable advantage to generate and maintain mutant *pvt1* colonies.

Nevertheless, despite the advantages of the zebrafish model, very few lncRNAs have been characterized in zebrafish so far. This can be explained by the lower number of lncRNAs that are conserved from humans to zebrafish than from humans to mice. Moreover, more molecular tools are available for studying lncRNAs in mice. However, the growing interest for using zebrafish in lncRNA studies will certainly promote in the coming years the implementation of state-of-the-art tools applicable to this model.

4.2 Short RNA motifs to predict lncRNA function

The function of lncRNAs is nowadays mostly described through loss-of-function genetic approaches, which can be challenging. First, the tissue specificity and low expression of lncRNAs (Cabili et al., 2011; Derrien et al., 2012; Mercer et al., 2008) complexify lncRNA studies, especially in multicellular organisms. It is essential to study the tissues where a lncRNA of interest is expressed, but identification of the expressing tissue might be difficult, as RNA-seq or scRNA-seq may not detect lncRNAs due to their low expression. Second, it is challenging to accurately associate lncRNAs with a molecular function, as they can act at different molecular levels, in *cis*, in *trans*, transcriptionally, post-transcriptionally, and post-translationally (Constanty & Shkumatava, 2021; Marchese et al., 2017; Statello et al., 2021). For these reasons, efforts have been recently developed to try to identify specific molecular determinants that underlie the function of a lncRNA (Ross & Ulitsky, 2022) and to further perform function prediction of uncharacterized lncRNAs, as it is done in proteins (Loewenstein et al., 2009). However, we are still in the early days of this type of approach.

Two main molecular determinants can be exploited to predict lncRNA functions: the primary sequence (Constanty & Shkumatava, 2021; Kirk et al., 2018; Ross et al., 2021), or the secondary and tertiary RNA structure (Diederichs, 2014; Graf & Kretz, 2020; Smith et al.,

2013; Zampetaki et al., 2018). It is also included that the function of lncRNAs can be driven by both the sequence and the structure of the RNA in synergy. We have already reviewed in Constanty & Shkumatava the predominant role that the primary sequence can play in lncRNA functions. However, since lncRNA sequence is less conserved than mRNA, it is unlikely that large domains can be identified. Nonetheless, the majority of described lncRNAs interact with proteins (Constanty & Shkumatava, 2021; Statello et al., 2021), and this property should help define functional protein binding motifs.

Binding motifs of RNA binding proteins are often degenerated and short (Dominguez et al., 2018; Taliaferro et al., 2016). Thus, two main algorithms have been developed to identify short RNA motifs within lncRNAs: SEEKR (Kirk et al., 2018) and lncLOOM (Ross et al., 2021). The SEEKR algorithm identifies RNA motifs, or k-mers, on lncRNAs that are not evolutionarily conserved, from a user-defined group. SEEKR has previously allowed the identification of functional, degenerated RNA motifs, present in lncRNAs that are evolutionarily unrelated to *Xist* but harbor similar sequence motifs leading to similar functions (Kirk et al., 2018). In comparison, the lncLOOM algorithm identifies short k-mers on evolutionarily conserved lncRNAs from different species. This algorithm uses the syntenic conservation of short RNA motifs based on a graph representation of the input sequence. Thus, in this algorithm, the first uploaded sequence, called anchor, is essential as it will serve as a template for RNA motif discovery and evaluation. LncLOOM identifies perfectly conserved RNA motifs (Ross et al., 2021), even if proteins are known to bind degenerative sequences (Dominguez et al., 2018). This algorithm was successfully used on the *CHASERR* lncRNA, identifying two RNA motifs that were further shown to drive the interaction with the RNA helicase DHX36 (Ross et al., 2021). These two algorithms demonstrated that primary sequence analysis of lncRNAs can successfully lead to the identification of functional RNA motifs. Thus, with the identification of motifs present on lncRNAs, a functional classification of lncRNAs by their k-mers content can be performed, to predict the function of uncharacterized lncRNAs (Kirk et al., 2018). However, because the identified motifs are relatively short (between 6 to 12 nt), they probably appear on RNA transcripts by chance, making challenging the identification of functional motifs on a single transcript. Thus, to characterize a lncRNA of interest, it is of utmost importance to use these algorithms with a group of lncRNAs to identify motifs that can be functional. The implementation of these algorithms will greatly improve the identification of the molecular grammar of lncRNAs, which should be followed by experimental approaches to associate motifs with specific functions. With the identification of

more motifs, it may be possible to build a library to infer the function of uncharacterized lncRNAs.

In our study, I aimed to identify sequences that may underlie *PVT1* function in vertebrates. I applied the lncLOOM algorithm to *PVT1* transcripts from 11 vertebrate species and identified two deeply conserved RNA motifs. Interestingly, while the TGGATT motif is upstream of the TTTTTG motif in mammals, the synteny of these two motifs is inverted in other species, including the zebrafish. Since the lncLOOM algorithm is based on the synteny of motifs for their identification, these two motifs were identified using the human or the zebrafish lnc-*mycb* transcripts alternatively as anchors. The reasons for the inverted synteny among species are unknown. I found that these motifs are present multiple times in the *PVT1* transcripts, reinforcing the idea that they could drive *PVT1*/protein interactions. In the long-term, the identification of proteins interacting with these motifs will be essential for understanding the function of the *PVT1* lncRNA, but also of other lncRNAs harboring these motifs.

4.3 Genetic strategies to inactivate lncRNAs

Genetic manipulation of lncRNA genes is more challenging than protein-coding genes. While mutation of protein-coding genes can rely on small deletions or insertions (indels) to impair protein expression, inactivation of lncRNA expression requires invasive genetic manipulations (Gao et al., 2020; Lavalou et al., 2019). Moreover, different strategies to inactivate lncRNA expression can lead to divergent conclusions (Anderson et al., 2016; Han et al., 2019). One of the first *in vivo* lncRNA studies adopted a systematic deletion approach (Sauvageau et al., 2013). Although this was the best approach at the time, knowledge gained since then has shown the limitations of such deletions. Indeed, lncRNA loci harbor DNA regulatory elements like enhancers (Cho et al., 2018). In some cases, the transcription of the lncRNA, not the lncRNA transcript itself, carries out the function of the locus (Anderson et al., 2016; Furlan et al., 2018).

Several strategies have been developed, each with its advantages and pitfalls. Targeted deletion of the TSS can be used to abolish lncRNA transcription (Alexanian et al., 2017; Lavalou et al., 2019). This minimally invasive strategy might lead to complete lncRNA inactivation in specific cell types or tissues. However, in other tissues, alternative TSSs might restore lncRNA expression, limiting its application in multicellular organisms (Lavalou et al., 2019). To overcome this issue of alternative TSS activation, insertion of premature polyadenylation cassettes was implemented (Anderson et al., 2016; Ballarino et al., 2018; Grote et al., 2013; Isoda et al., 2017; Lavalou et al., 2019). Although this strategy minimally

impacts the lncRNA locus and does not disrupt enhancers, transcription of the locus is interrupted, leading to confusion about the importance of the transcript or the transcription. Knockdown strategies are good complementary strategies to confirm the function of a lncRNA transcript (Gao et al., 2020). But this technology is limited to cell lines as knockdown strategies are poorly amenable *in vivo*, especially in zebrafish. Furthermore, antisense-mediated knockdown has been shown to trigger premature termination of transcription leading to the same issue as polyA insertions (Lee & Mendell, 2020), although this can be avoided by specifically targeting the 3' part of the lncRNA transcript with antisense nucleotides (Lee & Mendell, 2020). Finally, the identification of lncRNA-associated small RNA motifs may allow for a more precise manipulation of lncRNA function, with minimal impact on the locus (Ross et al., 2021). However, it is important to keep in mind that targeted alterations of small RNA motifs allow characterizing the functions of these RNA motifs, not of the full-length lncRNA transcript. Overall, no approach is perfect and thorough characterization of a lncRNA function should systematically rely on independent and complementary strategies.

The identification of deeply conserved RNA motifs within the *PVT1* lncRNA allowed us to perform an educated genetic approach to try to decipher its function, by removing such conserved RNA motifs in zebrafish. Interestingly, as the motifs map to the 3' part of the lnc-*myc* transcripts, their removal left intact the enhancers present in the 5' part of the *pvt1* locus. Additionally, this deletion did not impact the transcription of the locus, with the lnc-*myc* truncated variants being expressed. As a whole, such strategy allowed the functional characterization of the regions of the zebrafish *pvt1* transcripts that contain the RNA motifs, specifically. However, complete inactivation of the zebrafish *pvt1* would be necessary to better understand *pvt1* function *in vivo*.

4.4 P53 upregulation leads to multiple developmental defects in zebrafish

The p53 pathway is a master regulator of cellular stress (Harris & Levine, 2005). Upon cellular stress, p53 is activated post-translationally and drives the expression of specific factors to reduce stress. If the stress fails to be reduced, p53 induces senescence or apoptosis (Harris & Levine, 2005). *P53* expression is essential in embryonic development. Its balance is primordial for correct tissue development as downregulation or upregulation of *P53* may lead to lethality or malformation (Choi & Donehower, 1999). Interestingly, exacerbated P53 activation led to developmental syndromes rescued by P53 loss-of-function in mice (Bowen & Attardi, 2019) and zebrafish (Chakraborty et al., 2009; Danilova et al., 2010; Plaster et al., 2006).

In *pvt1* zebrafish mutants, I observed multiple developmental defects in association with an increase in the p53 pathway and apoptosis. Importantly, double homozygous mutant embryos without developmental defects did not show such an increased p53 pathway. Moreover, I described that the p53 pathway was upregulated before the appearance of the first developmental defects at 1dpf. These can suggest that the developmental defects arose from the specific upregulation of the p53 pathway. A *p53* loss-of-function would be required to assert that p53 upregulation is responsible for the developmental defects, with for instance p53 morpholino injection. Unfortunately, this experiment was not performed due to the penetrance of embryos with developmental defects. Interestingly, since I observed *myc* being transcriptionally active as early as 1 dpf and upregulated at 3 dpf, it could be the cause of p53 upregulation. At 1 dpf, I observed increased ribosome biogenesis, a pathway regulated by MYC (Poortinga et al., 2004, 2011), which can moreover trigger the p53 pathway (Golomb et al., 2014). Thus, *myc* loss-of-function in zebrafish would be required to assess if myc proteins could be responsible for activating the p53 pathway in our mutants.

4.5 *PVT1* transcriptional regulation conservation in vertebrates

The gene expression program that establishes specific cell types and tissues in multicellular organisms is crucial for the correct formation of an individual. Misregulation of the transcriptional program can lead to diseases, cancer included (Lee & Young, 2013). The basic concepts of transcriptional regulation were described 61 years ago (Jacob & Monod, 1961), with DNA sequence elements bound by transcription factors (TFs), recruiting the transcription apparatus. Nowadays the notion of transcriptional regulation has become more complex with the involvement of transcription factors, cofactors, and chromatin regulators (Bannister & Kouzarides, 2011; Bonasio et al., 2010; Fuda et al., 2009; Ho & Crabtree, 2010; Lee & Young, 2013; Roeder, 2005; Taatjes, 2010).

PVT1 expression is regulated by various transcriptional factors (Olivero et al., 2020; Tseng & Bagchi, 2015; Xu et al., 2017). Interestingly, two transcription factors known to regulate *PVT1* expression in cancers, MYC (Tseng & Bagchi, 2015) and FOXM1 (Xu et al., 2017), interact with *PVT1* and are stabilized by this interaction. This leads to positive feedback between the lncRNA expression and the transcription factors (Tseng & Bagchi, 2015; Xu et al., 2017). In homeostasis and cancer conditions, P53 induces the expression of a specific *PVT1* variant through an alternative TSS (Olivero et al., 2020).

During my PhD, I analyzed TF binding motifs in the vicinity of the lnc-*myc* RNA promoters. I found, with high confidence, that an E box was downstream of the TSS of both

zebrafish genes. Concordantly, lnc-*myc* expression was correlated with *myc* expressions across zebrafish development. However, scRNA-seq would be required to verify that both genes are expressed in the same cells. Our data may also suggest that the zebrafish *pvt1* transcripts are involved in myc protein stabilization, as in humans. Indeed, I observed increased myc protein levels correlated with an upregulation of lnc-*myc* expression in embryos with developmental defects. To test this, complete inaction of *pvt1* expression in zebrafish would be required. However, inactivation of *pvt1* expression is challenging. Therefore, genetic manipulation of the E-box at lnc-*myc* promoters might inhibit the *pvt1* transcriptional regulation by myc. This manipulation could help to characterize the function of *pvt1* in the myc pathway.

Promoter analysis also identified the presence of p53 REs, only in the lnc-*myca* locus. Interestingly, p53 REs were identified 3 kb downstream of the TSS. This could suggest the presence of an alternative TSS, similar to the *PVT1b* variant in humans and mice (Olivero et al., 2020). Analysis of transcriptome and TSS-associated histone marks, such as H3K4me3, at different developmental stages, did not support the presence of alternative TSS at the lnc-*myca* locus. However, the alternative lnc-*myca* might be visible only in stress conditions since this putative TSS contains p53 REs, although the equivalent *Pvt1b* variant in mice is detectable in homeostatic conditions (Olivero et al., 2020). Therefore, it is more likely that the P53 REs is in the vicinity of the lnc-*myca* TSS by chromatin structuration. Interestingly, I showed that p53 REs may be functional: lnc-*myca* is upregulated upon DNA damage and its upregulation is inhibited in embryos injected with p53 morpholino.

These discoveries show that the transcriptional regulation of the lncRNA *PVT1* may be conserved from humans to zebrafish. I further described that one zebrafish *pvt1* homolog transcript might be involved in the stress pathway, with lnc-*myca* and not lnc-*mycb* expression potentially regulated by p53. This could enlighten a separation of functions of the paralogous genes. The lnc-*myca* transcript might be functionally equivalent to the *PVT11b* variant in mice and humans, and the lnc-*mycb* transcript might be equivalent to the *PVT1a* variant.

4.6 *PVT1* involvement in the erythroid lineage

Erythropoiesis is the process of producing red blood cells (RBCs), named erythrocytes or erythroid cells. Erythrocytes carry hemoglobin to supply all organs and tissues with oxygen. In vertebrates, erythropoiesis includes the synthesis of heme and hemoglobin, the clearance of organelles, and the remodeling of the plasma membrane. Interestingly, the molecular mechanisms involved in this process are highly conserved in zebrafish, positioning it as a model to study erythropoiesis (Carradice & Lieschke, 2008; Kafina & Paw, 2018; Kulkeaw &

Sugiyama, 2012). Erythropoiesis takes place in the mesodermal germ layer, with two sequential waves. The primitive wave generates erythrocytes and macrophages during early development. The definitive wave produces definitive hematopoietic stem cells that are able to differentiate in all blood cell types, maintaining homeostasis throughout life (Kulkeaw & Sugiyama, 2012). In mice, the primitive wave occurs from E7.5 to E10, and the definitive wave from E11.5 (Kulkeaw & Sugiyama, 2012). In zebrafish, the primitive wave takes place from 12 to 24 hours post-fertilization and the definitive wave starts around 2 dpf (Kulkeaw & Sugiyama, 2012). Unlike mammals, both waves of erythropoiesis depend on erythropoietin (EPO) signaling in zebrafish (Paffett-Lugassy et al., 2007).

Analysis of scRNA-seq across mouse development (from E6.5 to E13.5) showed that *mPvt1* is essentially expressed in the erythroid lineage (Cao et al., 2019; Pijuan-Sala et al., 2019). The function of *PVT1* in the erythroid lineage has not been elucidated yet, even if *PVT1* was previously shown to be expressed in *in vitro* differentiated erythroblasts (Ghetti et al., 2020). Accordingly, *mPvt1* expression was triggered after EPO stimulation in the murine immature erythroid J2E cell line (Gillinder et al., 2017). These suggest a specific function of the lncRNA *PVT1* in the erythroid lineage. In zebrafish, embryos deleted for the regions containing the RNA motifs had a lower number of active erythrocytes, a hallmark of anemia. Moreover, I described that embryos without developmental defects were more susceptible to oxidative stress, highlighting the impact of the deletion on the erythroid lineage. Indeed, acute sensitivity to oxidative stress could be due to decreased numbers of active erythroid cells, since erythroid cells are the main cell type involved in reducing oxidative stress (Kramer et al., 2017; Scott et al., 1991). Erythroid lineage impairment may be independent of p53 and myc upregulation, as embryos without developmental defects were impacted. These observations support the involvement of the lncRNA *PVT1* in the erythroid lineage in vertebrates. Further studies are required to understand the *PVT1* mode of action in erythroid development.

The zebrafish has become a model for the study of erythropoiesis and anemia (Avagyan & Zon, 2016; Kulkeaw & Sugiyama, 2012; Uechi & Kenmochi, 2019; Zhang et al., 2021). Thus, studying the involvement of *PVT1* in erythropoiesis could be furthered with our mutants. Different tools are available. Transgenic zebrafish alleles were developed to study erythropoiesis, such as the *gata1*:DsRed that specifically marks *gata1*-positive erythroblasts and facilitates *in vivo* study of erythropoiesis at different developmental stages (Kramer et al., 2017). Since erythropoiesis occurs in two waves (Kulkeaw & Sugiyama, 2012), it would be interesting to understand if *PVT1* is involved in both. Moreover, it might be interesting to investigate the dependence of *PVT1* action in erythroid lineage on EPO signaling.

4.7 Conserved *PVT1* interactome

LncRNAs are regulatory factors that have essential functions in a wide range of biological processes, regulating gene expression, protein stability, and/or their activity (Constanty & Shkumatava, 2021; Statello et al., 2021). LncRNAs act at different molecular levels by interacting with DNA, RNAs, or proteins. But lncRNAs are almost always bound by proteins (Constanty & Shkumatava, 2021; Statello et al., 2021). Therefore, the molecular function of lncRNAs is defined by their interactions with proteins. Thus, the identification of lncRNA/protein interactions has become a major goal in the field to identify lncRNA functions. Several methods have been developed, involving RNA-centric or protein-centric approaches (Gräwe et al., 2021; McHugh et al., 2014; Ramanathan et al., 2019).

There is a wide variety of RNA-centric methods, all requiring mass-spectrophotometry to identify novel proteins binding to the RNA of interest. However, mass spectrophotometry identifies only highly expressed peptides and is based on already known peptides annotated in databases. Despite the variety of existing methods, it is challenging to identify the whole interactome of a specific transcript (Gräwe et al., 2021). Protein-centric methods are based on the use of antibodies and sequencing of the transcripts that are bound to the protein of interest (Hafner et al., 2021; McHugh et al., 2014; Ramanathan et al., 2019). While antibodies might not be available for the studied proteins, protein tagging emerged as an alternative for straightforward protocols, as antibodies against specific tags are more specific. However, an important caveat of protein tagging is the possible obstruction of the native form of the protein and its activity.

Thus, despite the variety of available methods to identify RNA/protein interactions, each has caveats. Moreover, the use of sequencing or mass spectrophotometry requires time and complex downstream analysis. A new high throughput technology to identify RNA/protein interactions has been developed in my host lab, the incPRINT technology (Graindorge et al., 2019). incPRINT is based on a library of ~3000 human proteins, and identifies interactions in cells, limiting *in vitro* artifacts (Graindorge et al., 2019) . Although the identification of protein interactors is limited to the proteins present in the library, it includes all major RNA binding proteins, transcription factors, and chromatin-associated proteins (Baltz et al., 2012; Castello et al., 2012; Graindorge et al., 2019; Taipale et al., 2012). Thus, this technology allows the identification of new protein interactors in record time (one week). Interactions then need to be confirmed through orthogonal methods (Graindorge et al., 2019).

To identify the interactome of the *PVT1* transcripts, we relied on the incPRINT technology. To identify interactors that bind to mammalian *PVT1*, we used the mouse *Pvt1a*

variant, as the human transcript could not be fully amplified. To identify interactors that bind to zebrafish transcripts, we used the lnc-*mycb* transcript as it contains both RNA motifs. We found 80 proteins that potentially bound mouse *Pvt1* and 122 proteins that bound zebrafish *pvt1*. Of note, the protein MYC is not present in the incPRINT library. Comparing the proteins that bind to both zebrafish and mouse transcripts, we identified 32 common interactors, which represent putative conserved *PVT1*/protein interactions in vertebrates.

By scrambling the RNA motifs of the *PVT1* transcripts, decreased interaction was observed for two DNA repair factors, RAD51C and NUSAP1 (Kotian et al., 2014; Somyajit et al., 2012). This could be concordant with the role of the lncRNA *PVT1* in p53-dependent stress response. Consistently, other conserved interactors were involved in DNA repair. Among these, the histone chaperone Structure Specific Recognition Protein 1 (SSRP1) subunit of the Facilitates Chromatin Transcription (FACT) complex did not specifically bind to the RNA motifs but bound to the full-length *PVT1*. The FACT complex promotes deposition of new H2A.X histone at DNA damage sites (Piquet et al., 2018), being thus involved in DNA repair signaling (Piquet et al., 2018). These interactions need to be confirmed through orthogonal methods but may point to a role of *PVT1* in the DNA damage repair pathway. An hypothesis would be that *PVT1* may recruit repair factors, such as NUSAP1 and RAD51C, at DNA damage sites. Thus, it would be of high interest to assess if the *PVT1* transcript colocalizes with DNA damage sites on chromatin (using combined RNA FISH and immunofluorescence).

Interestingly, mutations in *RAD51C* were linked to Fanconi anemia (FA) (Somyajit et al., 2010; Vaz et al., 2010). FA is a rare chromosomal-instability disease associated with a variety of developmental defects, bone marrow failure, and predisposition to leukemia and other cancer. FA pathway has been shown to be the main pathway involved in DNA interstrand cross-link repair (Kim & D'Andrea, 2012). As such, sensitivity to cross-linking agents is an absolute diagnostic of FA. Bi-allelic mutations in one of the 23 FA genes lead to the accumulation of DNA damages in Hematopoietic stem cells (HSC). This drive bone marrow failure due to the exhaustion of HSC and lead to the development of anemia (Ceccaldi et al., 2012). Since I showed that (1) the lncRNA *PVT1* could potentially interact with RAD51C, (2) the zebrafish lncRNA is upregulated upon DNA damage in a p53-dependent manner, (3) *mPvt1* is mostly expressed in erythrocyte progenitors during development, and (4) the zebrafish mutants develop anemia, it could be interesting to test whether our mutants are more sensitive to DNA crosslinking agents. Several studies were conducted using the zebrafish to model FA-like phenotype (Botthof et al., 2017; Ramanagoudr-Bhojappa et al., 2018; Rodríguez-Marí & Postlethwait, 2011). Interestingly, these studies showed that mutations in FA genes such as

fanco/rad51c led to sex ratio bias between males and females (Botthof et al., 2017; Ramanagoudr-Bhojappa et al., 2018; Rodríguez-Marí & Postlethwait, 2011), similar to what I noticed in the zebrafish *pvt1* mutant. This highlights that our mutant could be a model to test the possible involvement of the *pvt1* lncRNAs in FA pathway.

It is often supposed that lowly expressed lncRNAs cannot regulate important biological processes, due to the stoichiometry imbalance. However, several models recently reported substoichiometric action of lncRNAs, notably through phase separation (Elguindy & Mendell, 2021; Unfried & Ulitsky, 2022). Accordingly, *mPvt1* was described, in a cell line, to form nuclear foci (Olivero et al., 2020) that might resemble membraneless structures. Therefore, *PVT1* could act through such mechanism to regulate the DNA repair pathway. A fine quantification of the number of *PVT1* transcripts per cell would be interesting to perform.

4.8 Inhibition of *PVT1*/protein interactions

The lncRNA *PVT1* is overexpressed in all major human cancers, highlighting that *PVT1* could be an interesting cancer biomarker, and maybe a therapeutical target. While most proteins associated with diseases are not druggable, RNA/protein interactions in diseases represent a promising set of therapeutical targets. Here, I showed that *PVT1* may be related to the DNA repair pathway. Interestingly, it was suggested that *PVT1* induces radioresistance by positively regulating the DNA repair capacity of cancer cells (He et al., 2018). In this study, downregulation of *PVT1* led to a higher apoptosis rate of tumor cells (He et al., 2018). This shows that *PVT1* might have a role in DNA repair in human cells. Therefore, the interactions between *PVT1* and DNA repair factors might be targeted to enhance the sensitivity of cancer cells to treatments. incPRINT has recently been successfully used to identify small molecules inhibiting RNA/protein interactions, using an FDR-approved small molecules library (Sabate-Cadenas et al., *in preparation*). Small molecules inhibiting *PVT1* interactions could be tested in zebrafish for toxicity and *in vivo* efficiency. The different developmental defects identified in this study could serve as a template to probe functional *PVT1*/protein interaction inhibition. However, the efficiency of such cancer treatment could be highly dependent on the cancer type: RNA motifs are present in specific *PVT1* exons and *PVT1* mode of action might vary depending on expressed *PVT1* variant.

4.9 Discordant *Pvt1* activity in mESCs

Although *Pvt1* is highly expressed in mouse ESCs (Tuck et al., 2018), our analysis did not show any impact of the *Pvt1* TSS deletion in mESCs. Moreover, despite $Pvt1^{\Delta TSS}$ ESC lines displaying a significant reduction of *Pvt1* levels (70%), Myc protein levels were not impacted. This observation is not concordant with previously proposed action of *PVT1* as a binder and stabilizer of MYC (Alessio et al., 2019; Lee et al., 2021; Tseng et al., 2014; Tseng & Bagchi, 2015). This suggests that *Pvt1*/Myc interaction may not exist in mESCs, and/or that redundant mechanisms are involved in Myc stabilization in this cell type, Myc being an important pluripotency factor (Chappell & Dalton, 2013).

The $Pvt1^{\Delta TSS}$ mESC model represents an opportunity to perform biochemistry experiments that are challenging to perform *in vivo*. Notably, $Pvt1^{\Delta TSS}$ cells could be used to confirm the involvement of *Pvt1* in DNA damage repair. However, I did not observe cell proliferation defects in $Pvt1^{\Delta TSS}$ ESCs, which may exclude an accumulation of DNA damage in these cells, at least in normal conditions. This should be retested upon induction of DNA damage. Moreover, it is possible that *Pvt1* might not act as a DNA repair factor in mESCs, as pluripotent stem cells rely on different DNA repair pathways than somatic cells (Luo et al., 2012; Wang et al., 2021).

Overall, I was not able to decipher the role of *Pvt1* in mESCs. Despite an important reduction of its expression, RNA-seq did not highlight important gene misregulation, and neuronal differentiation was also not impacted in $Pvt1^{\Delta TSS}$ ESCs. This could suggest that *Pvt1* is highly expressed in mESCs due to the high expression of its neighbor gene Myc or that stress induction is necessary to decipher *Pvt1* function in mESCs.

5. MATERIALS AND METHODS

Zebrafish maintenance and generation of mutant alleles

Deletion of the region containing the conserved motifs in the zebrafish was performed with co-injection of two single guide RNAs (sgRNAs) and cas9 protein (50 ng/µl, a gift from the Concordet Laboratory, Museum d'Histoire Naturelle, Paris) in AB zebrafish. sgRNAs were generated as described in Lavalou et al., 2019. Genomic DNA was extracted as described previously (Bitetti et al., 2018) and amplified by PCR to genotype the fish. All sequences of the relevant oligonucleotide sequences are listed in Supplementary Table 2.

All zebrafish were bred and maintained at Institut Curie, Paris. Animal care and use for this study were performed in accordance with the recommendations of the European Community (2010/63/UE) for the care and use of laboratory animals. Experimental procedures were specifically approved by the ethics committee of Institut Curie CEEA-IC #118 (project CEEA-IC 2017-017) in compliance with the international guidelines. Zebrafish were staged using standard procedures (Kimmel et al., 1995). Double homozygous adult fish were obtained from double heterozygous crosses. The embryos analyzed were from double homozygous crosses.

Culture of mESCs and generation of mutant alleles

E14 mESCs were cultured in gelatin coated plates in GMEM media (sigma) supplemented with 15% FBS serum (Gibco Life Tech), 100mM MEM non-essential amino acids (Sigma), 1mM sodium pyruvates (Gibco Life Tech), 1mM L-Glutamine (Gibco Life Tech), 0.1mM Beta-Mercaptoethanol (Gibco Life Tech) and LIF (Mitenyi Biotech).

To generate the *Pvt1*$^{\Delta TSS}$ allele in mESCs, gRNAs were cloned into PX459 plasmid and transfected into the cells using lipofectamine 2000. Transfected clones were selected with 1µg/ml of puromycin. Genomic DNA was extracted and used for genotyping by PCR and sequencing to ensure the correct mutation. All relevant oligonucleotide sequences are listed in Supplementary Table 2.

RNA motif identification

RNA motifs harbored within the *PVT1* transcripts were identified using the lncLOOM algorithm (Ross et al., 2021) with 10 000 iterations. 12 sequences from 11 species (see SuppTable 1) were used. The human *PVT1* and zebrafish lnc-*mycb* transcripts were

alternatively used as anchors. The sequences were generously provided by the Ulitsky laboratory, Weizmann Institute of Science, Israel.

RNA extraction and qRT-PCR

Total RNA from zebrafish embryos was extracted using the TRIzol/Chloroform (Invitrogen) according to the manufacturer's instructions, followed by DNAse treatment (ThermoFisher). For embryos up to 3 dpf, the chorion was removed by pronase treatment (1 mg/ml) (Roche). Each biological replicate was composed of 20 to 40 zebrafish embryos (unless indicated otherwise).

Total RNA from mESCs was extracted using TRIzol/Chloroform (Invitrogen) according to the manufacturer's instructions, followed by DNAse treatment (ThermoFisher).

Complementary DNA (cDNA) was produced from 500ng-1µg (250ng for sublocalisation experiment) of total RNA using oligo dT and random hexamer primers with the SuperScript™ IV kit (ThermoFisher). qRT-PCR reaction was prepared using *Power* SYBER™ Green PCR Master Mix (ThermoFisher). Gene expression was assessed using the Viia™ 7 Real-Time PCR System with 384-well Block (ThermoFisher). For each biological replicate, qRT-PCR reactions were performed in quadruplicates. The gene expression level was normalized to the *eukaryotic translation elongation factor 1 alpha, like 1 (eef1a1l1)* in zebrafish and to the *Glyceraldehyde 3-phosphate dehydrogenase (Gapdh)* in mouse, using the delta CT method.

RNA-seq

Following total RNA extraction, the quality of the RNA was assessed using the 4200 Tape Station (Agilent). Only samples with a RIN >7 were used for RNA-seq.

Following RNA extraction of single embryo at 1 dpf, 30-100ng of total RNA was used for RNA-seq library preparation using Stranded Total RNA Prep ligation With Ribo-Zero Plus-Illumina following manufacturer's recommended protocol.

Following RNA extraction of embryos at 3 dpf, 1µg of total RNA was used for RNA-seq library preparation using TrueSeq Stranded mRNA prep Ligation-Illumina following manufacturer's recommended protocol.

Following RNA extraction of mouse ESCs, 1µg of total RNA was used for RNA-seq library preparation using Stranded mRNA prep Ligation-Illumina according to the manufacturer's recommended protocol.

The libraries were sequenced on an Illumina NovaSeq. For analysis, reads were mapped to danRer11 or mm10. The sequencing was performed at the CurieCoreTech Next Generation Sequencing (ICGex) platform, Institut Curie, Paris, France.

The differentially expressed genes were analysed using DESeq2 (1.32.0) (Anders & Huber, 2010) with default settings. Lowly expressed transcripts were removed from the analysis (sum of the number of counts in all samples <10). Significantly differentially expressed genes were defined as genes with log2 fold-change superior or inferior to 0.5 with an adjusted p-value < 0.05. Enrichments analyses were performed with ClusterProfiler (4.0.5) (T. Wu et al., 2021) with an adjusted p-value, False Discovery Rate (FDR), at 0.05. The background used was all expressed genes of the dataset.

Promoter Motif Analysis

Promoter motifs enrichment was performed using HOMER (v4.11) (Heinz et al., 2010). The list of genes analyzed was obtained from the differentially expressed genes.

Scanning of motifs present on the zebrafish and mice *Pvt1* locus was performed using the FIMO algorithm (Grant et al., 2011), using the default settings. The sequences analyzed were downloaded from the danRer11 and mm10 genome assembly. The p53 and Myc binding motifs were downloaded from the JASPAR motifs database.

Cytoplasmic/nuclear fractionation

3 dpf embryos were harvested and cells were dissociated as described (Bresciani et al., 2018). After filtration, we followed recommendations of previously published protocol (Ron & Ulitsky, 2022). The cells were pelleted and resuspended in RLN buffer with 0.5% Triton X-100 and incubated on ice for 10 min. The composition of the cytoplasmic and nuclear fractions was assessed by qRT-PCR using *eef1a1l1* to normalize. *malat1* and *gapdh* were respectively used as nuclear and cytoplasmic control.

Western blot analyses

100 embryos at the desired developmental stage were collected and deyolked using the deyolking solution (55mM NaCl, 1.8mM KCl, 1.25mM NaHCO3). Embryos were washed and SDS sample buffer (63 mM Tris-HCl pH 6.8, 10% glycerol, 5% ß-mercaptoethanol, 3.5% sodium dodecyl sulfate) supplemented by 1x proteinase inhibitor (cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail, Roche) was added. The samples were boiled and spun down.

Mouse ESCs were spun down at 1000 rpm for 5 minutes to remove the supernatant. Cells were resuspended in Radioimmunoprecipitation assay buffer (RIPA) buffer supplemented with 1x protease inhibitor (cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail, Roche). The samples were boiled and spun down.

Total protein concentration was measured by the Bradford assay (Bio-Rad). 10µg of total protein was loaded into a NuPAGE 4-12% Bis-Tris Gel (Invitrogen). The gel was run using the NuPAGE MES SDS Running buffer (Invitrogen). The proteins were then transferred using wet transfer to a nitrocellulose membrane (Amersham Protram 0.45 um NC). The membrane was blocked using 5% milk 1h at room temperature, and the primary antibody was incubated overnight at 4°C in blocking buffer. Secondary antibodies were incubated 1h at room temperature and ECL revelation (GE Healthcare) was performed using the ChemiDoc MP (Bio-Rad).

To differentially visualize Myc and γ-Tubulin in the zebrafish embryo blot, the blot was incubated first with the Myc antibody and revealed. Then the blot was incubated 10 min with 10% sodium azide to quench the signal and then incubated overnight in 5% milk, 10% sodium azide with the anti-γ-Tubulin antibody before revelation.

Western blot quantification was performed using the Bio-Rad Image Lab 6. All quantifications were normalized to the loading control, γ-Tubulin in zebrafish and GAPDH in mouse. The original Western blots are available in Supplementary Data 1.

TUNEL assay and immunofluorescence

Embryos at 3 dpf were fixed for two hours at room temperature in 4% PFA, followed by overnight incubation at 4°C in 30% sucrose. Embryos were embedded in OCT prior to cryosection. 15µm sections were obtained.

TUNEL assay was performed using the in-Situ Cell Death Detection kit, TMR red (roche) following the manufacturer's protocol.

For immunofluorescence, cryosections were permeabilized with 0,2% Triton 5 min at RT and blocked using 20% goat serum and 0.5% BSA 1h at RT. Primary antibody was incubated overnight in 0.5% BSA at 4°C. Then the slides were washed, and the secondary antibodies were incubated 1h at room temperature. Following washes, the slides were mounted using vectashield with DAPI (Vector) and visualized using inverted scanning confocal LSM780 – Zeiss.

Antibodies

The antibodies used in this study for the Western blot analyses are anti-γ-Tubulin (T6557, Sigma), anti-MYC (for zebrafish) (ARP30292_P050, aviva system biology), anti-GAPDH (GA1R) (MA5-15738, Invitrogen), anti-MYC (Y69) (for mESCs) (ab32072, Abcam). The secondary antibodies were HRP conjugated: anti-Mouse (W402B, Promega) and anti-Rabbit (W401B, Promega). The immunofluorescence was performed with the anti-H2AXS139ph (GTX127342, GeneTex) and the secondary antibody Alexa-Fluor 546 (A11010, Invitrogen).

Image analysis

All images from TUNEL assay and immunofluorescence were analyzed using the FiJi software.

Zebrafish images

All images of zebrafish embryos were acquired using Zeiss Stereo Discovery.V20 with a numeric camera AxioCam MRc (Zeiss).

Morpholinos

10ng of tp53 morpholinos (Gene Tools) were co-injected with 80 ng of *H2B-RFP in vitro* transcribed RNA as a microinjection tracer into the one-cell stage embryos. The *in vitro* transcription of the *H2B-RFP* transcript was performed using the HiScribe SP6 RNA synthesis kit and cap analog nucleotides (NEB).

Hydrogen peroxide treatment

3 dpf embryos were treated for 10 min with 100mM of hydrogen peroxide added to embryo water as described (Reinardy et al., 2013). The survival of the embryos was assessed one day after incubation based on no heartbeat.

Camptothecin treatment

24 hpf embryos were incubated in embryo water with the indicated concentration of camptothecin (Sigma) at 28.5°C for 6h. The embryos were then collected. RNA was extracted as described above.

Hemoglobin staining

Embryos were incubated in the staining solution (0.6 mg/ml o-dianisidine (sigma), 0,01M sodium azide, 0,65% H2O2, 40% ethanol) for 7 min before being rinsed twice in embryo water. Erythroid staining used in this study is based on the reduction activity of the catalase of erythroid (O'brien, 1961). Embryos were then visualized using a stereo microscope and pictures were taken as previously described. The normal and abnormal staining was assessed visually.

incPRINT

Cloning of the constructs and full scale incPRINT was performed as described (Graindorge et al., 2019). The interactions between *XIST(A)* with eGFP and HNRNPC were used as negative and positive controls respectively (Graindorge et al., 2019). We established the threshold at two Relative Light Units (RLU) to select proteins putatively binding to the RNA of interest. For lnc-*mycb* screen, we noticed a decrease of the global intensity, including the controls, for the second half of the screen. Thus, we established the threshold at 1.75 RLU for the second half of the screen.

The small scale incPRINT, with the 32 common proteins, was performed using the full scale incPRINT protocol, using 150 ng/ul of plasmid expressing the RNA and 75 ng/ul of plasmid coding for the protein of interest. The plasmids expressing mouse *Pvt1* fragment with wt and scrambled motifs were synthetized from Genescript and cloned into a 6xMS2 plasmid. All sequences of the relevant oligonucleotides are listed in Supplementary Table 2.

Enrichments analyses were performed with ClusterProfiler (4.0.5) (T. Wu et al., 2021) with an adjusted p-value (FDR) at 0.05. The background used was all proteins present in the library.

Proliferation assay

30 000 cells were plated in a 6-well plate. At each time point the cells were gathered and proliferation was assessed by the number of cells. The cells were counted using the cell counter Vi-Cell XR (Beckman Coulter) according to manufacturer's protocol. The media was changed every day.

Neuronal differentiation

Prior to the neuronal differentiation, cell media was changed to DMEM/ F12 GlutaMax Neurobasal [1:1] (Gibco) media supplemented by 2i inhibitor (N2 and B27) (axon medchem) and LIF (Mitenyi Biotech). After several days, the cells were split at high density ($5.10^3/cm^2$)

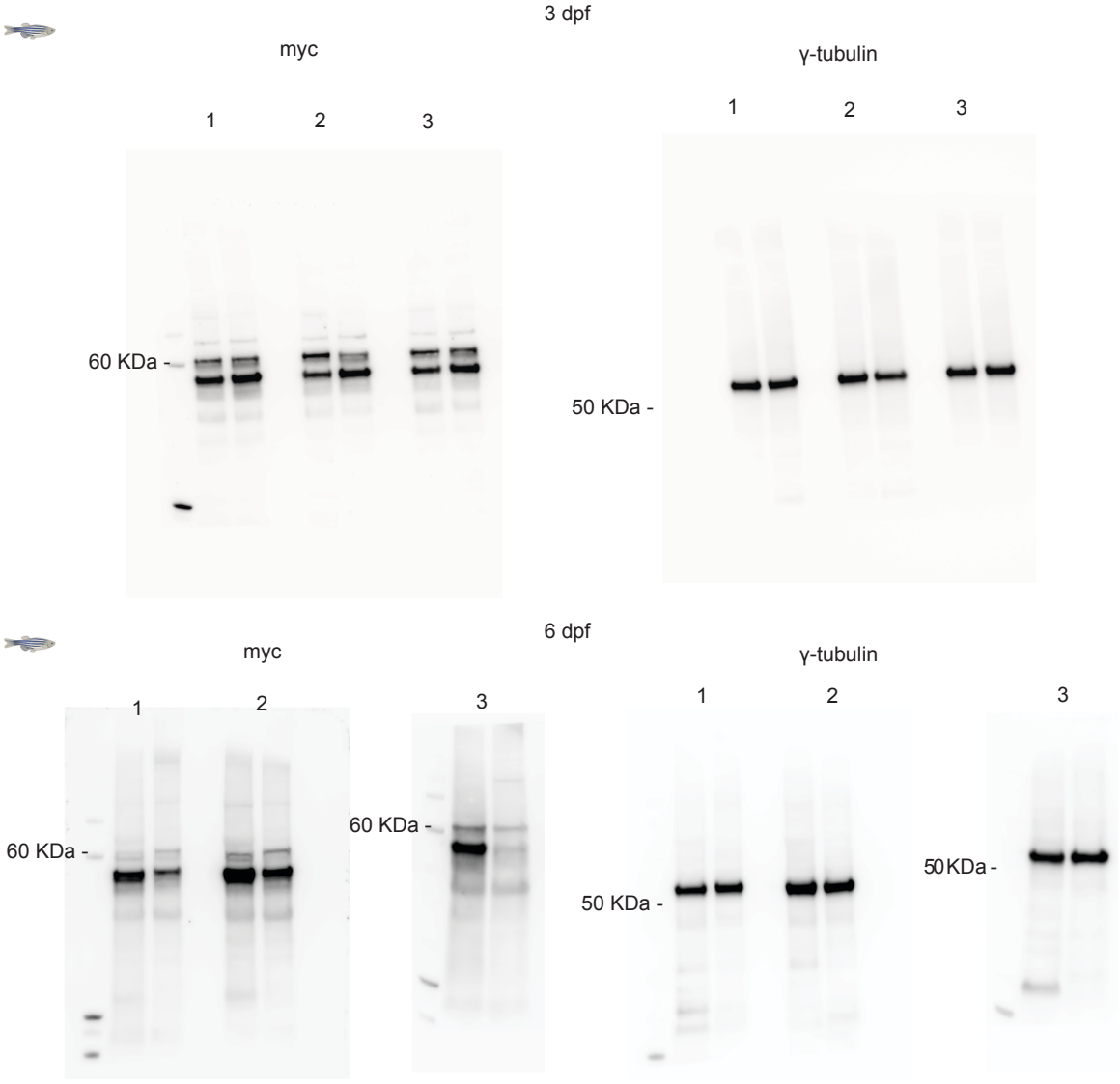
in a 6-well plate. 48 h later the media was changed, and retinoic acid was added (1uM). After 24 h of retinoic acid treatment, cells were washed, and the media was changed without retinoic acid. The media was changed every day and the cells were gathered at specific time point up to 10 days of differentiation. The differentiation efficiency was assessed by qRT-PCR. All sequences of the relevant nucleotides are listed in Supplementary Table 2.

Figures and statistical test

All Figures and statistical analysis were made using the GraphPad Prism 8. The RNA-seq figures including PCA and scatter plots and statistical tests on differentially expressed genes were generated with R. The data are represented as mean ± SEM. ns: not significant, *: pvalue<0.05, **: pvalue<0.01, ***: pvalue<0.001.

6. SUPPLEMENTARY DATA

6.1 supplementary data 1: Original Western blots



6.2 Supplementary data 2: Sequence of *PVT1* vertebrate homologous transcripts

>Zebrafish_lnc-*mycb*

CGATGCGCGTGCTGCTGTGAGCACATGGAGGAGAGAGGACGTTCACTTAGCAGCCTGGCTAATGTAGCCCGTGGCTCGCAAAATATTACG
AAATTAGCAACATTTTTCACCTGGAAAGTGAGCTGACTCGCCCAAGGCGTCCAAAAATGGATGAAAGATACTCATAACCACTTCAGGGAG
AGTAAAAGATAAGGGAGAAAGAAAAACATTGAAGATGTGAGTTGTCCCACAATGAGACTCGATCTTCAAAATGGAGCTTGTTGTTTCCTG
AGTCTGAGGACACAACTGAGTGGTGGATGGCTTGAGCTCCGTTGAAAGATTAAAAACAGGAAATGCTCCCCACCCAGATATGAAGAGTAA
GGAGAGGCAGAGATAGGTGAAGAGGTTGTGTATCAGATAACCGTACCCTCTGTTGGTGGACTCAAATTGGACTGAATTCCTCAGTAAAGT
GGTCTCTCATCTCTATCATGGGACGTGTTGTACTTCGTGGTTACTGGGTGGTCCTGTCACAGAGGGGCAGCAGGCCAGGCACACGTGTAC
CAGTGCTCTAAAGAAGGCCAGTTCAGTGGAAACTAGCAAGCAAACACTCCAAAAAGGACACGCGCTCCCTCAATGTCAGCAAATATTTGG
TCACAACTTCAAAGAGCAGTGACAGTCATTTGAAACTGCGGATCGGACTCATGTTTGTCACTGTAAGAGCAGGTGGAATGTGTTCATTTT
TCTTGCTTTATTTTTGTAGGCCACGTGGCAACATGTTTTAGAGGCAGTCTGTGATGTAACAGCACTGCTGTCACCTGCATGTCCAGCTGT
GCTGGTGGCTCAGAGAAGTGGACTAATGAGGGTCCTAAGGTGAATCTGATGGATTGAATGAAGTGTAATTAGAAAGATGGATTGTCACTC
AAGTGTGACAGCCAGCTCATTCTGTACCTGAAGTGTTCAGAAATCTGCCTACTAGTCTGTTAGGACTCTATGGACAAAAAATGTAATATC
ATTAACAAAGGGTAAAATATAGAGGTCTACAACTTGTTCCCTTTAAGGGTTTGGATTATGAAGAACATCACACCAAGTATGTTTATCTGA
GGGTAGTCGGGCAGTTTATACAAGTGTGTATTATTTGTCTGTCTTTCTGTATGATATATTCTGACTCTTTGACTTTGAATGAAATTTTA
CCCCAAAACTCCTGTACGCAGATGCATTGCAATGCATAATTTGTTCTGAGACAAAAGAGGTAAAGCACACAGCATATAATTATATCTCAC
AATGTCAGTGTCTAGTAATCTCATGAGTATTTGTATTCACATCAAATATGGTTCTACCAAATATATATATTTACTGTACTTATTAAAAAA
GTCTTAATACATTGACACAAACAGCACAATATTGTGATACTGACTATCTAAAATGAATTGTTTTTGTATGACCATGTCCATAAACTAATA
AAAAAATCTTGATTATGTAGTGTAATGTAATTTGATTAACCATTATTCATTTAATTATACTTTAGTTCACCATGGAACACACAATAGGCT
TGTTAAGTCAAACTGTGACACCAAAACTGCCAAATAAAAACACAAAAGCAATCAAATTCCAGTGAATGGACTCGCCACGCTCCACTGCCG
GCATTTACACATCTCTCACCATGCGCCTTCCCGTGGAGATTGGCTCACTTCGTTTTCTTCATGATGAATGTCTCTCTGAACCTCTTGTGG
CATTTGAGTGAACTGCCGCTATTTTACAGATTCCAGCCCAGACATGATTCATCCAATAGAAACTGCAGAGGAATTTGCTGCATCACAAAA
CACATACCCTGTTGAGGCTTTCATATGTACGCAGCATTCCTGTTTTGTGACCAACGCAGTCTTGAGAAATTCATTTATTTGTTCTCCTTT
AGCCTATTCTTCAGCTTCAACAGCCTCTCCGGCACTTTCATCTCTCTATAGAGACGGGTTTGTATGAATAATATATGCCACATTGCCTTC
CCATCATTAAAGGTGACCTTGGTTAATTATTGATCGGGTCAGATGATTACAGTCCTGACCTCTGTATGTTCTCATATGCAGCTTGGCTCG
CTTTTCATGTCCACGTTCTTCCTTTCGTTATTTCTGGAGGCCCTGAGATGATGAACTACTGTCTGTTATTGTTGTTGGGGAGAGTCGGCG
CCTTCACAGAGATGCAATGAAAATGCAACCCCCTTTAAGACTACTTGTGTTTTAATGAAATATGTGAGGGGTAATGTACAGTCAGCTCAC
TAAATAGACCCTGACAGGGTGATCAGGTTGAGCCTGAAGGGTTTTTTGTTGTTTCAATGACCGGCTGACTCTAAAATGTCCTGCTTATTA
CCCTGAAACTTCTCGAACGAATGTATGCGATGATGAATTATTGTTTAGTTGTTTTTATTATGCTTCATTTTAAAAAGAACCAAATCATAA
TACAGTTAAAGGTTAAGGTTCCAATCTTATTCAAGCGGAGAACTAACTAAAATCGGGTTTCATTAATTGTTGGAGTATTTTCTAGAATTT
GACTTGCATTCAATGGCCACTTTGGGCA

>Stickelack

CGTATCCCGAGCGCTCCCCGGAACGCGTGGTCCCGATCGCGAGCACGTGGAAGTGAGGTCGAGATGTTTCCGTGAGACGCGCGACTGATC
ATCAGGGATGTTGTGGGGGGGGGGAGGACCAAAAGGAATAGAAACGAGAGAAAACCCACCACGAAAAGCTGGAACCGAAGAGTTTAGACG
CCGGCTGGAAAACGGGTGTCCGCTCTTTAGACTGATCAGAAAGGTCACGACCCCCCCCCTTTTTTTCTTCTTCTTCTTTTCTCCAAGTGC
GGGAGAGGAACACGTACGGCTTCTTCATTTTGAGTCCTTCCTGAATGTGAAATCCTCACGTTGATGAAGACAAGATATTTCTTGGTTTTA
AAGAGCTTCAGCCGCTGTTACTTTTTGATGAATCAGATCTATCTGCAGCACGAGAACCGAGGAGCTGACGTCGTGGGAAAATGTCAACGC
TCATATTCTGGCGATGGCGTCGGGCTGAATTCAAAAAAGGGGAGCTCAGTTCCACTGAATGCGAAGAGTCCAGGGAGACTCGTACAGAAG
CTTTGGGTGCCACAGTTTGCGGCGCTTTGGAGGCGCTCGTAGTGGGAAGGGAAGGGATTACGCGTCCGTGACCTCTCGTTACGTTACACG
CATGCGCGTAAAGTATCCGTGTTCACAGGAAACGTCCCGTTGCTTTCCGATTAGTGAGCATTTGGTTCGTAGGCCGGTGCGTCTCCTGCG
GACTAGGCGGCGGTATTTGACTGAAAAGGGGTTGTGTTACTCCATAGTCGCGCTTTAAAACTAACAAACGTGGGAAAACTGCACGGTTGC
ATGTTATTTAATGCCAATGATCATCTAAGAAAATAAATATTGAATTGGGCCTTTGTGGTCACACACGGAACGGCTTAAAGGATGATTTCT
TTCTTTGTTCGGTATTTGTCTTCTTTTAACCTGTTTTATCAGATAAACTAACCAATACCCACCTGCACATTTGTCTTGTCATTCCAAACT
CGGCCTCCACAATAAACACAAATGTGTTTTTCTCTCTTCGCTGTTATCTGACTTTAGCAGCAAATCTACAACGAACGTTTCCACAACTCA
TACTGAAAGAAGAGGCTCATTTGCAAAGTTGAAAGTCATACTTGCATTATGAAACTAACGTTAGCTCACACTCGTATCGCTACCCGATCA
TCTTCTGTCTCTGTTTTGCAGGGGCAGTGCAGTTGCAGACCCCCCCCCCCCCGTGCAGTCCCGTGTTTTACGGCCGCCGCCGCACTTACG
ACCAGCAGAGATACCGATCAGAATGGGAAGGATGCGACGCTGAATGCAGTCGGCGTTGGATTCAGTTTCACCGAGTCCTAAAACACACAC
ACAAACACACAAACACCGCATGGGTTTTAATACTATCTACTTCCTCCTTTCTTCGAGGGGACACATACACCCTGAATGGGCCCCTGTGGT
GATCTGCGTTGTGGTCGAGAGTGTTACAACTGCCACGCTGGTCTGCTTAAAGGGGCCGTGCTGCTGATCCTCTTTCAGGGGGTCCGTACG
TCGCCGCAGCGTTTCAGGCTTCTGCACGGCTGCATGTTTCCGCCTGTTGTCAGTGAGCGAGTGCGTTCTTGTACTAATTGCATAAACTTT
AGGAATCTATAATAGCTGGGGGGGGGCGGGGGGGGCTTGAGAATTTGAGAATCTAAAGAGAATTATTGATCCGTCCACAGCGAAAGTGAA
AGTCGAATGGACTTTGTAAGTGTTTTAAGAAGCTGCACCCGTGTTTGAAATAAATATTTCACAGGT

>Tilapia

GTAGCACACTAATAAAAGTGTGGCACGGGAACTCAAACGGTCGTTTATACCTTAAAACCACAGATATTTAATATATAGATATAACGTATA
TTTATATCTTTATATATACATGTAAAATGAACTGTCTGAGTTCGAATAAATGAATAACGTGGTGAAGTTTATCTAGTTAGACCGTGGTAA
AACCCCACCTGTTCTACCGTTGTCTCGAGCGGCTTTCCAGACTCGACGTGACGTCACACACGCCGTATCCCGAGCGCTCCCCCCTCCACC
TCTGAGCGCGTGGTCCTGGTCGCGAGCACATGGAAGTGAGGTTTGAGATTTGTGTCTGAGAGACGCGCGAGCGACTGTCGGGAAAACGGA
ATATAAAGTGAATTAAAGAAGCAGCCCGAGCAGCCCGGTTCTGTCTCAGCTGTCAACACTGCCATAAGACGGCTTCTTCGTAACGATGAG
ACACGTGACAGCGGCACTGTGTGACGCTCCTGGATGAGAATGCATTTTTTCGGTATTCGCAGTTGCAGCTTGGCACATGAAAGATGGCCC
ATCACCTAAAGAAAAGCGGAAAAAAGTGAGTAGAGGTTATTGTGCAAGCAGCCTCAACCGGACTGTGTCATCTGTGGGGAGAAGCTAGTT
TGCACTTGGCGTTTTTTCGCGCCAACGGCGACGTCACAACCGCAGTTGCACGCCATGTTTTAGTGTTGCTTCAGCGAATCCTCCACCTGA
ACACAACCTGGACATCAGTCAAGGGAGGAGATGGAAGGCAGTTAAGGTAGAGGAAGCGCTCTCTCAGCCCGCTGGCTTTCCTCTGCAGCA
GCTGTCGGTCAGCGTCAGTTGGTGTATCTCGGAAACACAGATCGCTAATGAGGCTGGAGCTCTCGAGGACTGCCTGGCCCTGAACTCAGC
ATCACACGGTGCTAAGTGACGAGGATCCCCTGCTGCACCCAGGGAGGGTTATGCATCTAAAACGTGAATATTATATACGTCACGTCAAAT
CAGGAGACGGGTGCTCTCATCTGCAGCTTGAGCTGTCTGGCTTTGAGATCTCGCTGTGCTCGGGGTGGTCCTGCAAACTCAGGAAAGTCC
AGGTACGATTTGATTAATGTGGATTTTGGCTCGTCTGGACAGAAACATCAGCAAGGACACTGCGGACGTAAAATGCTTCTGCTGCTGCTG
CTGTGATGGTGGTGTTGCAGACCCCCCATGCAGGCCAGATCTTTCTGCAGCTTCTGCACCGACACATGATCTGACTAGGAAGGATGCAAA
GGTGGAAGCAGACCGTGCTGGATTTAGTTTCACAGCATCTTACTGGAGCAAACGCACACAACATACTGCAAGGTTTTCAGTAGTATCTAC
TTCCCCCTTTTTTTTTCTTTTTTTTGTTTGAGAGGACACGGATACACCCTGACTGGGCCCCTGTAGTGCTCTACATTGTGGTTGGGAAGT
ACCTAGTTTTACAACTGTCATGCTGGCCTGCTTAAAGGGACAGTGCTGATCAATTCGAAGCACATATTCAGGGGTCTTTTAGTCACCCTA
ACAAGGGGGGTCACATGGATGTTTGAGTCGTCATTCCTCATGTGAGGAATGTGCAGTCCGAAGAGCCGCTTTCTGCATCAAGCACTGCAT
AAAAATACTGATTGATATTTGTTAAAAGAATGTCTTTAATAAAAAGTTGAATTGAATTGGG

>Shark

GGTATTATTTTTTATATATATATGCGTTGCTCTCTCACCACCCGAAAGGCCGGTGTCGTTTTATTTTATGCTATTTATTCTCCTCCTCTC
CCCCCCCCCCCCCCTATTCTTTTTTTTAAAAAAAAATTAAAAATATGAATCCGATCGCACCAGAAACGTGCGACCGACCCTGATGGCGAT
CGTGATGGTGGGTCACTTGTGGGAAGGGATGGACGTCCGTCCGGGAGAGGATTTGCTGCTGTTTGTACCAAGTCAGAAATCAAGTGATGC
AGTCAAGGCCAATTGTAGCCCTCTGACTGCTACCCTGGATCAGATCAGCTAATTACAGATAGCCTCGCTAGAGCTCTTGAGTTTAAAAGG
AAAATATATCAGCCAGATTCCCACAGCTGATCACCACTCAGAGAGACCCCTGCTGGAAAGCACACTTGCGGATACCAGTCATGTCACTTA
CATAATTTCCTGCATGACTTAATCTCTCGCAGATACCTACCTCATGGGCACATTCAACTGTATGCCGTTGTCTCTTGAAAACACTACGGT
TTTATTATACCACACTCATTTTTACTTAACCTTTGCATGTTATATGCCATTTAAATTAGTTTCACCTTTAGATTCTGGAGACAGGCGAGA
GAATAACTACCTTGTTTTAACGAACTTGCCGTTGGAACCGTATCTCAAGACTGGAGGATCTCTCCATGCCCTCTGGTCACGGTCATCGGT

GATCAATTACACCTCCAGCTGCTCTCGCAATTCGGGGCACTGTTTGGTGGTCAAGATGAGGAACTAGAGGAAAGAAAAGCAACCAGGTTT
CATGAGGCTGAATTCCTAGGGGTCACAACTTTTTTTTGCCCCAGGATTGGTTGGACTTGGATGGTTAGCGCTTGTGGAGGAATCGAGGGG
TCCAGTCCCTCTGACCCATGAGACCATGGCGGGCGGTTATGTATGACTCAAATCTAGAAGGAACTTGAGGAGTTCCCCTTTCTTAAACCA
GATAGTGAAAATCACAAGGGAGGTGAAAGGGCAGTGGGAATGCACCACTCAAGCGTCCAGAACCAGTTTGCTCTAAAATATTTGAAGCTG
CTGGAACCATTAGTGGCTGAAGGAGAGAGATAAAGACTGCGATTCTCTGGCATTTTGCTGTCTGGTCAGTGCTCAAGGTGGCAATTGAAG
TTACTGGCCAAGAGAGTAATAGTACAGGACCAAATCCTGATTGTGGGGTCTGATAAGTGGACAAAGATTGAGCCAGAAGATGGGGAGCTA
TTCTCGGACACACCCTCCTTCCTGCCCGTCAGAAGTGCTTTAGTTCGAAGGACTTAAAAGGCGTTTGACTCAAGGCTTTGCCTGGAAACC
TCAACACGATTCGGAGATTTGCCTTTTCAACGCGAGGTGACCCTGCAACCCATGGTCCCAATTGGGAACTGATCATAGCCAGAACCAACT
ACAACCTCACATGGCACAAAGCCACCGTCGAACCGTTTAGGTTGAGTTTCAGACCAATGTCAGTGCCAGAACTGAGCTGTCACTCAGTAA
AAGAACAAGAAAGTCAGTTAAGGTTTAAGGTTTAGTCGTTAACTCTTGATGGGGGGAGACTTTTCAGACCAACATCTAAGAAGATTAAGT
TCCTGGAATGACTCCAGACACCAGTTGCAACACTGAAACAGCCGCACACGCCCAAATTGTCAACGAAAAAAGGTAATCAAATACAATTGT
CTATTATGTGTAGCCAGAGAAAACAAAAATCTAATTCCTATTCATAACTGTATACTGAGCTGTTGGAGAATTCAAAAAACTTCTTGCAGG
CAGCAATCAAGTGACTTGTGGACAGAAGACTTGAGTTCTGCCTTTTGGTGGATTGTTAGCCATGATAAATGTTAGGAAATTGTGACATTT
GTTTACAAAGTTGCTGCCAGCGGCAACTTCAAAGGTTTAAGATCAAAGTGGTCGGAGGCCATTCGCTGCACCTTTCTTGGTGCTAGTTTT
AGTGATTTAGATCTTTTTATTTTCAGGCTGTTCCTGAAAATCCCACATATGTGCGACGGAGGCACTGATGAGCGTTGTTGGGGTTTGGAA
ATGAAGGGGTAACATGTGTGGCTCATGTAATATTTAGAACATTGAATGGTCTGGAAGAGGAGCGTTCCATACCCCCTGCACTCCCCACCC
CTGAGGCCAGTCTCAATGGTGAAGAAACCAGGTCCGCAATATTGGATTTGCGATTTATGCCTTTGACTGCTAGGTATCAGATATCGTTAG
CTATTTAGTTCACAAAAAACTGGTAAGCACTCCGAGCCACTGCCCTATGTGGAGGACACTATATAAACACAAGTTGTTGTTGTTGGTATT
CCTGTTGAAGGCTTGAGATGAAAGTACAAACCCCTGGACTCCACTGACACTTCACTGCTGGCTGCTGCCTCAGATTATTAGTGAGAAACA
AGGAGGGAGATTAAATGCTTACCTAGCTTTCTCCTTTATATGCTATTCGACTTACATTCATGCCTTTTTTTATCATTACTGTATCAACA
ACTCTCTTTGATTTTGTAACTAGCCTCCAGCCTACTTGGTGTCGGTGGGTAGAGATGGGGGAGGGGTTATTGAAGTGGGGAATCTGTGTT
CCATATCTGTTACCATCCTGTAAAACCATTTAATTGTCTCATTCTTCAATCCACTAAAATAAGTTTCCATTGTACTGAGAAATGTAAAAT
TTGAATAAAACACTACTTGTTATAACTGCAGTTTACTTTCTACCACATCTTGGTATTATTTGGAACATGTTATATGCTCTTTGGTGAGTC
AACTCTCTGTGGGCGTCTCAGTAGTCTGGATCCTAAGGCAGGGTCCACAGTATTCCAGGCACTGCTGATGACAGCACAGCCAGTCTGGAT
CTTCAGCAGGACGTGAAAGGGCAGACCAGTCACTGGGGCAGGGTTAATCCTTCAAGCCAAAGCCCTAATGGGTAGGCTGGTCCTAGCTGG
ACAGGCAGTGGACAAGTTTAATAATAAAAATAATAAACCTTGCAATTGATATGAAACTTGCCAGTTCGTGCTCATAATGACTCTTCGCTC
TTGTCAATCTGCGCCTAATACTCACCAGAATTAGGGAATTGGGAAGCATCTTTCTATGAAGAGAACGTGTGATAGCATGGTTACTTTAGA
TGGCCAGTCCACTGGTAGCATGTGGTTACATCAAGCTATGGAGGAAAAATTCAGCCTGGGCTCTGATCCCTATAAATTGACACAGACCGC
AAAACAGTATTAATGCGTGCACAGCATGTAAGGAAAACACTATGTCTCTTCCCCCCACCTAAACTCCGCACAGTATTGAATGATTGTAGC
TGCAACACAAGGGGCTGCCTAGGCCCTTGAGGCATTATAAAGGAGAGCAACAAGGACAATTCTCCACATAAGAGGCATGAGCTATGAGCA
ATGATTAGACAAGCTAGGGCTTTTCAGTCTTGAGTGACAACGACTCAGAGGGGACCTTATAAAGGTACATACAATTTTAAATGCGGTTGT
GAACTCGCAAGCAAACTCAGAACAATAGCTTCGGACGTACAAAGACAAGATTGTTCAGGATAATTTTAGGTCAGACACCAGGAAGTTTCT
TCTTTCCCCAGAAGGTAATTAACAATTGGAACAAGTTACCAGGAAGGGTGTTTGAGGCCAGCAGGCTGGACTCGTTTAAAAACAGCTTGT
TGCTGTGATGGAAAGGCAGCAGGATGGGTATGCTGGAGGTATAAGATCAAATGGACTGAAGGGATTTTCTCATCCAAACTCTCATCTCTT
GAGACCTGCTGACACTCACCAACTATATCTGCAAAGGATACCCACATACATGAGGCTACCTCTCAGAGAGCAGCCAATGTCTGCAGAGCT
GTGCCCCAGGACAAGCCAATTCCTTATGAGAAGGGTGGAGAACCCCCTCGCACACACACACACACACAAGAGTTGGATTTCACCACCACC
ACAGCAGTTGCAGGTTCTGTCATACTGAAACCAGCCATGATCTAAAAATAAATAAATGAAGTGTCAGGCATCAAACTAGTTCATTTCCAG
ACAACATCAATGCAATTTTCAGGGAGTCCAGTTTTCATCGGTTCTTCACAAATGTGTGAATCTGAAATGTGCAAATGAATGCAAGAATTA
TGAATGACTGAGGGAGCCACTCCGTGCCCAAGCAGCCTCTTTTAGCTATGACAACGGGATGGTTTGAATCATCGAATACTCTGTGCTACC
TCGTGCTCCGACAAAGTACTGACAGCCCAAACTGTGCCAGGTGAACTTGAACCATGGGTAGCACTGGGACTCTAACTCACAACTGTTTGG
TAGGAAGGCAGCCGCACCACAAAATATACTATTTTAGCAGCTTATATTTTCATTTACCGTATACTTCAGCATATGTAACAATAGCTGCAC
GCAATACTGTCCGTCTGAAAGGTTGCTGTAGAGGAATTTATCTGGGTATTCGTTCCCGGTATCTTGTACGATAATGGTGTGGTTCAAGTG
AAATAAAGACTGCAACGTGCACAAGCTTGCAAAACCTCTATGAAACAATCCAGTCGTCTCTGGATGAAGGTTCTGTGTGGACTAAGAGTT
TGTTGCTGGCTATCGTTGGCATATGACTGTGTGCTGAACAACAGCAAACCACAAATGTCAAAGGTTTTGAAGTAGGGGCTTGAGCACCCT
ACCAGAATACTGACCCAGCCATCCTACACTTTGGAACAAAGAATGGTGTACCACAGGAACAGGACACACCCCAAAGTAACACTCAACAGC
TCCACTTAAAGCAAACATTCAGCTCTGAAATGTGTAATACGGGAAGTGCAGTCTTATTGCAGAAGCACTAATAAAATACTAAGCATGGGT
GAGGAAGACCAATGTTGGGATGGGGTGGAGGAGAATGATGACATGCCAATGGACCTTGATGGAGGTCCAGCAAGCTGAGTGCCTCTACTT
GGTTAAAAAGGAATCACAAGATCACACCAACTTGCATTTATATAGCGGCTCTCATGCAGGATTGCATTCTCGAGCGCTTGACAGAAGCGG
AACACAACACCGAGCCTGGTGGAGGGGGAGGTGACCGTAGGTACAGTCGAATAGGAAGGATTCAGGCGCGACCTGAAAATGGGGAAGGAA
AGGTCAAGGCAGAGAGATAGAGGAAGAGTGTTCCAGAGTTTCGGGGCCTAGTGGCTGAAACCACAGCCACCAATGGTGGAACGGAGGGCA
TGAGAGACATGCAGAAGGCCGGAGCTAGTCAAGCAAAGGTAGTGGGTTGTGGGTTGTGGGGGTAAGCAGGTCGGAGAGGTAGATTGGTTT
GCATGAAGAAGGGCCTTTGACCGATTTGAACTCATTCCACCATTCTCCCATCACAGCATTGATCTCTTTCTTCAATAAGTCCAGCATCCT
GGCTTCAGCCACTATCGGCCTGGCTCCGTTGATAGCACTCTTGCCTCTGTGTCAAAAGATTCAACCTATTGTGATACATAGTATAGTGCT
GGAGCACCCCGTCGCTGTCAAAATCCTTCATACTAGACATTAGACCAAGGTCCTGTCTGCTTGTTTAGGTAGAAATTAAAGAATCCATGA
TATTATCTGAAAATAGAAGGGAATTTCTCTTGGTGTCCTGGTCAAAAATCCACCAGATAAATTGAAGGTTGTTTGTGGGACCTTGTTATG
CATTTGACTGTAAAACAGTACTATATGTTGACAGTAATTACTTGTTAAGTACTTTTTGATGCAATGAGGCACAATAAAAAATGTGACATG
GCAAGTTATGTTATAAAGTTTAAGGTCCGTTTTCTTAACATATTCTATGTAACATTTAATGGTAGTTATTGTACTACTGAGAAATGTTCA
ACATGATCTAAGATGCTAATAAAGTGATTGGTTTTCTGGAGGATATTTTTCCAAACCCTCTAAAATTGTCAAGTAAATTATTGCCAGAGT
AAGTCAATCTTTTGTGAAGTCAGTTGTGAAGTGCAAACACTGGCTGTATTATACAGATTTTAAATTGTTCAAATGCAGCTCCATCCACAC
GCAATAAAGGGCACTCTACTGTTTGCCTGTTTTTTATTGATTCTTTGCGTTGGTCCTTGCTGTACCAGCTACCATGGATCTCATAAGTCT
TTGACTTTGGGCACATCAGACGTGAGCGATCCTGTAAAGACTTCAAAGTATTGGCCAAGATAGCAGCACAGAACAACAAACCCATTGAAC
TCAGCGGCACGACAAACAGTTATCTCGGTGATAAAATCATAAAGATTACAACTTGATGTCGCAACTGCCATGGTTGAGTACTAACTGGTG
TAGAGCAGAGCCAGTAAACGTGGGAAGTCGGTGTACTGTGAACAGATGACCTCAGCTGAGTGGAGAAGGAAAAGTATTTGGATACAATCA
TTAATCATTCTCTGAAGATTTCCAGCCACTGTGATGCTAATCATATATTGCGCAAGACAGCATCTTGAGGGCATTTAAATATTGGACAAC
TTTCTGAGGGCATTTCCGATCTCTTGCCATCAACACTTTTCTTTGGGGGAGATTTTGTGAGTGTTATTTTTCTAGTCAACTGATGGGCTG
CAGAGGACTTGGCACTCTAGTGCAACTCTGTGGTATGCGTGGCTGCCTCTCTCATTGCCTTCATTAAAGTCTTGTTGCAGGAAGTGACAG
ATTAATTTGATCTGGTCTTATCAACAGTTGATCTTTGAGAAGCTTTTAAAAAGTTGAATAAATTGAAAAAAACATCTGTGGCTCAGCCAG
TAAACAGCATTTGCGCTATCAGAAATATTGGCTGTTTTTCTGATTTTCTTTTAGGTTCCTGAGCACAAGTGCGCGTATTCCCACAAGAAA
ACTATCAGCAGAATTCTTAATAGGGCTTGTACAGAGCGTACTGCAAGTCACTTAACACGTGTACGAGTCATCTGGTTCATGAGCCCTCTT
ATTATCTCGGCTGCCATTCAAGAATTCCAGTACATTTCAGGGCAGACTATATTCCATTGACTTTCCCTCATTTTGACTGGTAATTGATAC
ATGGTGGACATGTGGGCTTAGTGCCCAGCCCTGCAAGCACTCATTAAATATATGTACATTCTGTGTTCAAGTCTAGGGAAACTTTACTAC
CAACAAAAGCTGCTTTATTTAAATGTAACCAGTCTGACTATCATTTCATTCGAGTTAGTATTTCGGGGATTTGGTTGCATTTGAAAGGTG
ACCTATTTTAAATAAAATCATGTGCCTCA

>Human

CTCCGGGCAGAGCGCGTGTGGCGGCCGAGCACATGGGCCCGCGGGCCGGGCGGGCTCGGGGCGGCCGGGACGAGGAGGGGCGACGACGAG
CTGCGAGCAAAGATGTGCCCCGGGACCCCCGGCACCTTCCAGTGGATTTCCTTGCGGAAAGGATGTTGGCGGTCCCTGTGACCTGTGGAG
ACACGGCCAGATCTGCCCTCCAGCCTGATCTTTTGGCCAGAAGGAGATTAAAAAGATGCCCCTCAAGATGGCTGTGCCTGTCAGCTGCAT
GGAGCTTCGTTCAAGTATTTTCTGAGCCTGATGGATTTACAGTGATCTTCAGTGGTCTGGGGAATAACGCTGGTGGAACCATGCACTGGA
ATGACACACGCCCGGCACATTTCAGGATACTAAAAGTGGTTTTAAGGGAGGCTGTGGCTGAATGCCTCATGGATTCTTACAGCTTGGATG
TCCATGGGGGACGAAGGACTGCAGCTGGCTGAGAGGGTTGAGATCTCTGTTTACTTAGATCTCTGCCAACTTCCTTTGGGTCTCCCTATG
GAATGTAAGACCCCGACTCTTCCTGGTGAAGCATCTGATGCACGTTCCATCCGGCGCTCAGCTGGGCTTGAGCTGACCATACTCCCTGGA
GCCTTCTCCCGAGGTGCGCGGGTGACCTTGGCACATACAGCCATCATGATGGTACTTTAAGTGGAGGCTGAATCATCTCCCCTTTGAGCT
GCTTGGCACGTGGCTCCCTTGGTGTTCCCCTTTTACTGCCAGGACACTGAGATTTGGAGAGAGTCTCACTCTGTGGTCCAGGCTGAAGTA
CAGTGGCATGATCCCAGGTCACTGCAACCCCCACCTCCCGGGTTCAAGTGATCCTCCTGCCTCAGCCTCCCGAGTAGCTGGTATTACAGG
CGTGTGCCACAAAGCCTGGCTAAGTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGTTGGTCTCGAACTCCTGACCTC
AAGTGATCCACTCACTTTGGCCTTTCAACGTGCTGGGATTACAGGCGAGAGTCACCGCACCCGGACGACTCTGACATTTTTGAAGAGTCC
AGAATCCTGTTACACCTGGGATTTAGGCACTTTCAATCTGAAAAAATACATATCCTTTCAGCACTCTGGACGGACTTGAGAACTGTCCTT
ACGTGACCTAAAGCTGGAGTATTTTGAGATTGGAGAATTAAGAGCCAGTCTTGGTGCTCTGTGTTCACCTGGTTCATCTGAGGAGCTGCA
TCTACCCTGCCCATGCCATAGATCCTGCCCTGTTTGCTTCTCCTGTTGCTGCTAGTGGACATGAGAAGGACAGAATAACGGGCTCCCAGA
TTCACAAGCCCCACCAAGAGGATCACCCCAGGAACGCTTGGAGGCTGAGGAGTTCACTGAGGCTACTGCATCTTGAGACTCAGGATGAAG

ACCCAGCTTGGGGCTGTCAAAGAGGCCTGAAGAGGCAGAACACCCCAGAGGAGCCTGGGGCCACCACCCAGCATCACTGTGGGAAAACGG
CAGCAGGAAATGTCCTCTCGCCTGCGTGCTCCACCTCGGTCCACGCCTTCCCTCCTTCTGGAAGCCTTGCCTGACCACTGGCCTGCCCCT
TCTATGGGAATCACTACTGACCTTGCAGCTTATTATAGACTTATATGTTTTTTGCATGTCTGACACCCATGACTCCACCTGGACCTTATG
GCTCCACCCAGAAGCAATTCAGCCCAACAGGAGGACAGCTTCAACCCATTACGATTTCATCTCTGCCCCAACCACTCCAGCAGCAAGCACC
TGTTACCTGTCCACCCCCACCCCTTCCCCCAAACTGCCTTTGAAAAATCCCTAACCTATGAGCTTTGAATAAGATGAGTACGAACTTCAT
CGCCCACGTGGCGTGGCCGGCCTCGTGTCTATTAAATTCTTTTTCTACTA

>Mouse

CGTGACGTCACGGGCCACCCGCCAGCCCCGCGCTCTACCAGGCAGAGCGCGTGTGGCCGCCGAGCACATGGACCCACTGGTCAAGCGGGC
TCGGCAAGGCCAGGATGAAGAGAAGCTACGAACTTAGCATTCCCAGAGCCCCCCGAGTGGATATCCGCGTGGAAAGGATGTTGGCAGCCC
CGGTGACCTCGGAGTCACGGCTGGATCCACCCTCTTGTCTGATTTTCTTAACTCTACTACGGTAAGAAGATGCCCTCCATCAGCAGGCCG
TGTCTGCTGGGACCTCTCTGGAGACTTCCTGAACCCTGTGGATTTACAGTAAGATGCAGTTTCCTGGGGAGTAATTCTGGTGGATACAAA
TTCCGGAGTATTACACCTGGCTTGCCTGGAGATGTTACAAGTAGCTGGATGGATTCTTCAGCCTGGATGCCCACTGAAAACAAGGACCGA
AACTAAGAGGATTGTATCTCTACCCTTGACTCAATGAGAATGTTTGTTGTCATCTCTCGGGCTACCTGGAGCTCCCTCTAAAATAGAGCC
GAGTTCTGTCACCACAGACATTGGCAGTGGCACTTGCATGCCCCTGGCAAGCAGTCTATCCTGAGGTGGAGGAAGTTGCCCTTGAGTTCT
GGCTGGTAACCAGCACAGGGGCTGGGACACTGCCTGATTGAGGTGGCCATTGAGAATTCAAGTGTACCTCTTGGTCCCTGATGCAGAAAT
GAAGTCCCCAGCATGCTCAAGAAGAGCTTCCTTGGAGGATGAGAGGCATGGGAATCTAAGTATACCCTTTAAGCGTTCCAGAAGGATTTT
GAGATCCTTTCCTAAATCAAAGGTGGAATATTTGGGGATTTGGAAAATTGAGATGTGAAGCGTTGACTTAAGAGATGCCAAGTAACTCAG
CAGATGTCACACAGACGATAAATAGCAAAGATGGAAGTCTTCATGCCGGAGGCAATCCTATAAGACAGCTGAGTTCTGCAGAGCTGGTAG
GAGACAGACTTGCTCAGGTGATAGATCCAGCCATGATACTGACCCTAAGAGAATGAGACGCTCTGCAGAAGACAGAAGATTCCTGAAACT
GGGAAAGGTGCCTAGAAATCCTGATAAGAGTGAAGAAGGAGCTGGCAGAGCAGCCTTCCTCCGCACTATGAAAGACATCCAACAGAGAGC
AAGTTCCCACCTCACTTCCAGCTTGACTCCTTTTTATCTCCTAACTGTGGTCTGTGATGTCTTCAGCAATGGTATTAACTAGTTCTGGGT
ACCTGGAGGCACAGCAGCCATGTTGAACCTATGAGGTGATGAAAAACCATGCAACTAGAAACCAGAATGGAAGAAAAGAAGAAGCTGGAT
CCTTGACTGAATTGTCGATGGGGGATAGTCCATCATCCCAGGGCTAAGACCACCTCCAGATTTCTTGTTTGGTGAAACAAGTCCTGATTG
TTTACATTCCCTCGTGTTGAGCTGTCTCTTGCTTGTAGCTGAAAGCATTTCTAATAGATATTCAGATTATCAGCCTTAGCTCCAAGCTAG
AATACTCTATGGGACCTTAGCTTCTCCATGAATTCCTTTGCTCATCTGTGTAATGCTCAATAAATGTAAATTGAACCCTGAATGAGGGAC
AATGACACTATTGAAGTCATGATTGTCAATCTTATTAACGGACTAACTACTTAACAAAAAGGTTGGAACACTGCCTGACGCACAGTAAGC
GGCTGGGGAAAGGGTTTGCCAAATGGAGTGACAAGAAAACAGTCATGTTTATAGGCTTTCTTTTACATCTTGATCCATCTTCAAGCCTCT
GTGGTCTTTAAGCAGAACCTTGACCCCTTCAGTACTGGGGCTCCTGGAGCATGACACGCCCCCTAAACTCTTGTCTCAAAACACCCACTC
TAGCAGTTTCTTTTTGACAAAAAAAAAGCCTTCTTAGTGAATGCTGGCTTGTGAAAAAGAAGGGGCTGTGATCATCAAATCGTGTTAGTA
GATGACCAAAATACCCAAAGCAATTCATGCAAATAGATTCAAGTCTCTCTGGGCAAGCCTTCCTGAAGGCTCTAGAATTTCTGAAGCTCA
TTTTTGAAAGGGTCTTTCCATATGGTGGCCTTTTCTTTTTTGCACCTCCTCCCTTCTTTCTCATCAGAGCTAAAGGGATGCTCAGACTC
TCTGTCCTTTCTGTGTAGTGAGGGACCAGGCTGCTGGGCGAGTGTGAAGGAGCGAGTGTGATTTGGGATATTCTTTTCTGGTATGGCAAT
GGCCAGCCTTTAGAGAGGTCATTTGAAGTTGGCCAACAAGTAGACATTATTTCAAAATAGTTCTTGATCATTAAAGAACCAACCAGTATG
TCTTAAAAACAAACAAACCAATCTTTTGGAAAGTTTCTGCTGAATGTGTGCCCTTCCGACTCACTTCTTATCCTTTCTGTCAGAAACACC
CCAGAAAAGGGGGGTCTTGCCTTTGAACCACTCTGCTACTTTAATCTTCCTGCTTCTGTCTGACCTACCTTCCTTCTTTCAGTAGCCAGGA
TCCTTTCAGAGTGAGCTCCAAAAGGTGGGGATTCTCCAGCCACGAGTGCCCTCTGAAATTCTATCTTACTTAGAATAAAATATCTAGATT
TATACAGTGTCCCCTTGTATGGGTGGTGGCTTTTTTTTTTTTTCTATCATTTTCCTGCTTGTTCACACCACTCTCTCCCTGGCCTTCTTA
TGTCACAGATTTCCTGAGTACACTCTGGCCTCAGTCTTTCCATGAAGTAACTCTCCTTTCTAGAACATTCTTGCCTGCATAACTATCTGC
TTCTGTCCTCACTGTATTCTGATCTGTATTTCTATGTCAATTTCCAGACAGGCATTCTCTGCCCATCCTCTAAAGTGTGCCCCCTGTCTT
CTTTGCTGTTACCTGTATGCCCTAAGTCTAAGTCACTGACCTTATTGTAGACTAAATTTGTTAGCATTTCTTTCCAGGATGCAAGCCTAG
TGATGAAGGAGTCATTGTTCCATTGTCAGAATAGCTACAGAATCCAGCCTGTTCTGCCCTGGATGCTTTATTAAGTAATGAAGCTCAATA
AAACAATTTGACCAAG

>Rabbit

GCTCGCCGGGCAGAGCGCGTGCGGCCGCTGAGCACATGGGCCCGCGGGTCGGGCAGGTTCGGGGCGGCCGGGACGAGGAGGAGCCCCGAG
GAGCAGCTAGCAGGCATGAGCCTGGGCTCTCCCGGCACCCGCGAGTGGATTTCCACGCGGACAGGAGGTGGGCGCCCGGGTGACCCTGGA
GACACGGCTGGAGCCACCTTCCAGGTGCATGGTCCATAAAGTGGTTATCTGAGCGAATGAGCATTCTCAAGACAGTCGTCAAGACAGTCA
AGAAGAGGTTAGATCCACATACATTTGGATGAAGCTGACCGGAAGCTGAGAAGGATGCGCTCCCTCCCTTGCCCCACGCTGGCTGCGCCT
GTTGGTCGCCTGGGACCCTGCTCGAGGATTTTCGGAGCCTGTTGGATTCACAGTGACACCTGGCCCCCTGGGGAGGACACCATGGAAGCA
CGCTGTGGAACCACGCCTGGATGTGTGTGATGACTGGGAGGAAGGACCTGACTGAGTGCATTGTGGATTCTTCAGCCCGGGTGTCCGTGG
GAGACAGAGGACCTCAGCTGCCTACCAGGGCTGAGGTTTCCATCCTCGTGTAGGGCAGGAATGACTGGTTGGGTCTCCACCGGTTCCGCT
GGGGCTCCCCTGTGACACGGGTGAGGACACCGTGGACTTGCTCCGAGATGCTGAGCTCTTGTGCCTGTTGGACTGTGGCTTACGTAGCTG
ACACTGCCTCCAGCTGCACCCTCCCGTCCTCTGCCTCTGACCTGCTGTTCCTTGGCAGTGGTGGCTTCAGTCTGCAAGTGCTGTGCCCAG
TGATGTGTCGGAAAGCTAGCCCTTCTACAACATCACCATAGCCGAGGTACTCGGAGATCCAGCGTGAATACTGTGTTCTGGAGCTGGAGG
AAGGGTGAATGAGGGAGAGTACGATGCGCCAGCCCTAGAGGTGAGGACGACCAAGAGACAGTTGAAGACTTTAACCAAGAGACAAGAAGA
CACATGAAACTGTGAAAAAGGAGCCTGGAAAATTGCCACCATTTCCCCCTACACAGCCCTGTCTCTCATAGCACTTAATACTTTGTAGCA
GCGATTTGGTCACTGCTGAGCTATCTTCGTTTAATGTGTTTTTGACAGAATAGTTAGTCTGGAAATTTTATTTTCAATTAAACTTCCTTG
TTGAATGGCCTTGGTTTTGAAAGGATTCCATTAGGGTGGCTAATTACATAGCAGTTGCTACAAGGAGATATTTGCACTCAATTTCAGCTC
TTAGGTAATTCAACTCAAATGCTTATATTTTCCTCCCTGGAGCTGTGTGTGGCCATCTTCCTCTCATACTGCAGTCTTCTGAGTTATTTT
TAGAAACCCAAGTCGAAATAGCAGTTTTGGTGGAGGCCTTTTCTAGTCTAGCCAAGTAAGTCATATGTACTGATAAAGACAGCGGATGCT
CCCGCGGCTTTCAGAAGCAGCAACACTAACTCCTGTGCGAAGGGGCTCCATGTCTGTGAGCTGTCCTTTCTCCTATGACATTCCTTTGTG
AACTCCAGCTGTCACCAAAGGCAACGGAAATTCCATTACAACTGGGAAAGATTTGCCCATCGGCAAAGGTGGCAGATCACTCGGAATGCC
CCGGGGACCCTGTTGTGGCCGGTGCAGCCCCCGTGAGATGTGGACGTCACCGTCTCTGTTCTCTGCCTGGGCTCTGTTTTCTTCTCTGAG
CAGAGCACTTGTCATGTCCCTCAGCACGCTTGTGGAGTTGGGATGCTTTCTGCTGCAGTAACAGAATTCCCGAGACTGGGACATTTGTGG
GGAGTAGGATTGATTTAGGCTCATTCCGGAGCCACTGATGCCTTTGTGGGAGCCCCACCCTCATGACCTCATCTCAACCTAATCACCTCC
CAAAAGCTTCACCTCCAGATACCATCAGCATGTGACTTTGGGTACTGAGTTTCTAACATATGAACTTCTGGGGGCCGCAAGCCACAGCAC
TTACTAAGCAAATAAACCTCTCTCACAACAAAGTATTGTTGGTGGAGCCACGTTTTTCACCTTGTTCCATGCCTGGCACATAGTAGGTGC
TCAGATAGGGAAGGAATAACCATGCACCATTACAATCTAAGTTTTTCTGTTTTCTCTGTATTTTTTTAGTAGTTAATATTTTATGACAT
TCAACCTGAAAAAAATAAACTTTCCAAATAATAAAGAAGTCCCTTTTGGTTACCATCCCATA

>Dog

GCGGGGGGCCGCGCGCAGCCTCGGGCCCCCGCGCCCCCCGCGCCCCCGGGCGCATTTCCACGCCGAGAGCGCGTCGGCAGCCCCGGCGAC
CTGCACACGCACGGCGGGACCCACCTTCCAGCCTGATCTTTTGACCAGAAAGGGATTAAAAAGTTGCCCATCCTGATGGCTGTGCCTGCC
GGTTGCATGGAGCCTCATTCAGGATTTCCTGAGCCTCTTGGATTTACAGTGATACTCAGCTGCGGGGAGTGGAGGAAGCATTCTCTGGAA
TCACAGACGGCACACGTCTGGATGCTAATGATGGTTTTACGGATGGCCCCGGCTGAGTGCTTGGCAGATCCCTGCAGCCTGGATGCCTGT
GGGGTCAGAGGACCGCGGCTTCCTGGGACGGCGGAGGTCTCTGCCCTTATCTTCATGCAAATGTTCAGTTGGGTTGCTGCTAGCCTCACT
GAGGGCGCCCCGTGAAACAATCCTGTCATACCTCGGATTTTGGTGCCTTCAGTCTGAAAAATACACACATCCATTTGGTACCCTAGAGGA
ACTGGAGAATTGTCTTCCCCTGAATGAAGCTGGACCATTCTGGGATTTGGAGAAAGGGCCAGGCAGGCAACTCGTGAGCGAGTGATGTGG
ATCTGGTGGGCAGACCGCAGAGGTTGGTGGGGAACTCGAAGGAACCAGGACGCCCGCTGTGATGTACTGGCCTGGGTGCCCCGCTCAGAA
CAGCCCCCGGGGGATGTGCCTGGTCAGTGGGTGGCATCTCCTCTGCAGCGATTGGAAGACTCAAGTGTTCCATGCTCGTCGTCTGCCA
TCCTTTCAACGACCTTGAAGTCTACCCAGCTGGCAGAGGAGGGACAAGGAGGCTCCCCGAAGCCCCCAGGAACCCCCGCAGCCTGGCTCT
CGGCCGACCCTCTTCGGGACACCGTGTCATGTTGCGTTCACTGATCCGTATGATGGCATGTCCGGCACCACCAGTGACGGCAGGTTCGCT
GCTCCGTGTGATGACGCCTTCCCCGCTCCACGGAGTGCTAGCCTGCGTGTTCGTCGTGCTGAGAGGTCACAGAGGGAAAGGGAGCCTTGG
AGCAATGCCCCGGGGGATAGAGGCCAGAATGTTCCAGGTGGAAAAAATCTGTGCTCCAGGTGGTGACAGCAGTGAAATCTGGCCACACCA
GGGAGTGTCTGCTGCGGGGAGCCTCGGCCAGGCCACCGTCGGGCCTCCCTCCAGCCTGCTGGGCAGTAGCCCCCCTTCCCACGTCCACGC
TCCGTGGGCCGCGGTGAGTGTGTTAGCCAGGACGTTACACATGAGAAAGCTTGACGTCGGCTGGGCGACGAGTTGGTGTTGCTGCCTGCA
GTTCAGCCCCCCACCCCCACCCCTCCGCCCCCTGAACATTCTCTGGCTGTGACGGCACAGGGGGAAGGGCCGGCGGGCCCCGGAGACGCT
CTCCTAGTTGCCTGACATTCCGAGTATGATTCATAATGATTGTAGCAGATGAGGCATGAGGGCCGAGGCTCAGACTGCAGCTGATGGAGG
TCTTCTCCCAAGGAAGCGTGCCCAGCCTCCCCGGGGGCCTCCGCCGTCTGGCTGGCCTGGGCCGCCCGGTGCCCAGCAGCGTGGTCAGTG

ATGAGCGGCCCGTTATGGGGGGGCCACTGTGCGTAGGGGATGATGCTGCAGCCGCGAGAGAGCCCTGAGAATGGGGATCCCAGCAAGCTG
GGGTGCGTTTTACGTTCATTACTTTTTAAATATCTGGAAAATAATGAAGAACTATGAAAAAGGCTGTAATTGCACCAATTTTAAAAATTA
ACAAACGGGGGAGTGAAAAGTGGCCAACCCCACTGTCAAACACCTCCCACCTCTTCCTGAGATAATCCCTATTATCAGTTCGGTCATTTA
TGCAAACCCAGTGGTGGGCTGTTCAGACTCTTATGAAGGAAAGAAAGCCCCGGATTGTAGGTTTGCCGACATCTGAGGGGTGAGATCCCA
CCATGGCGATTAAAGATTTAAAATCCATTTTGCACAATTCTTGAATATTTACCAGTTGGCTTCTGTTAAACCCCTGCACTGTCACAAGCA
ATCACTTTCTTTTTGGCGGCACGAGGGGCGGAAATTGGAAACCCATTACGCATAGTGGTCTGCTTGTCATTTTCATTTCCTGTGTCGCAT
ACCTCTTCCTCAACCAGCGCATGCAGATTCACTTAGCTCATTTCCCTTGCATGGTGGGGTTTCTTCCCTGAAGGTGCACTTTGTTGCTCT
TGGAAGGTTCTTGCAGAATCGTCCCACCGGGGGAAGGGAGGGTGACAGCGGATCTTCGCGTTCAGGGAGGCTGGCTCCTGCTGCCCACAT
CTTAGACCAACTTCGTTGGAGTTCATTGTAGCTCCTGGCCTCTCCCGAACGTGTCATTCTGAGTCTGGAGGCTAAGCACCTCTGATGGTG
ACACTTGCATTCACTTTCACTCATCTTTAGGTTCCATTGGATAATCTCATTTTTCTTCTGGGGAGTCTTTCCTGATATTCCTCTCTTTCC
CCACACCTGCTCTCTGAGGCTGGGCAGTGGGCCTCTTCCCGGTTTCTTAGAACCCAGGGCTCCCATCTATGATGATGGTCCCCACACTTT
ATTATTTATTTTACAGGTTTCCTGTAGAGAAGCAGAGAAGCTCTGCCAGCCAGAATAGTACACTCCCCACCCCCCAATGCCTACACCCAA
ATCTCTGGGGCTGTGCCTATGAAATCGTTACTTTACCTGACAGAGGGGCCTTTGCAGATGGCATTAAGGTGACTAAGCAGTGGGCCCATG
CGGAGATTATTCTGGATGATGCAGACAACTTTCTTTGATTGGGATCAGAGAGATCGGCTCCCCCCTTCCCCCAAGAGGAATCAGCCTCCA
GTGTGAGAAGGGTTTAGTGTGCTGCTGGTTCCCACATGTAGGGACCCAGGCATGAGGACCGGAGGCAGAAAGGTGGCCCCTGCTGACTGC
CAGCAAGGAAAAGGACCCTCAATCCTCAACCACAGGAGACCAAATTCATCCGAGGGCGTGGATGAGTGAACTTGGAAGATGGCCCTCCCC
ACAGCCTCCCAGTTAAGAGCCAACACCTTGAATCCAGCCCACAGAAGCTGGGATGGAGGCCCCGGCTGTGGCACGTCTGACCTGGAGACT
GAGACAGTGCATGTGTGCTATTTGGAGCCACCGTGTTGGGGCATCTGGGTGGCTCAGTGGCTGAGCGTCTGCCCTCCGCTCAGGGCGTGA
TCCCGGGGTCCTGGGATCGAGCCCCACATCGGGTTCCCCACGGGAAGGCTGCTTCTCCCTCTGCCTATGTCTCTGCCTCTCTCTGTGTGT
CTCTCATGAATAAATAAATAAAATCTTAAAAAAAGAAATGAAGC

>Opossum

AGGAGATGAGGGGCCTTTGCTGCCTGCCTACCCCAGGGTCACCCTCCCAGGACCGGAACCCGCCGCCTTGCTCTGGGCTCCCCGGGATGC
ACACAGGCAGGCCTCTATGGGCGCGCGCGCAGGACAGCGGGGAGCTGGAGCCGCCGCCGCCGCGCCGCGCTGTTCCGAGGTTTCGCGAGC
AGTCGTGACGTCAGAGGGCCGCTCGCCAGCACTACGTTCCATTGAGGCGAGCGCGTGCGGCGGCCCGAGCACATGGGGGAAAACGGGCTT
GGGCGAGCTGCATCGCAACTGGATACAGGGGACTGTGCTCCTTGGGCTCCGTTGTTGCACCCCTTTTGGACCTCCGCCCTGGAAGGGATT
TTAGCCGCACGATGACCTTTGGATACAGGGCCGGATTCACCCTCCAGTTTTGATCCTGATCTTTTGACCAGAAGAAAAATTAAAAGAGTC
TCCTGACAGCTAACTCGATGGAACCACTAGGATTTTTACCAATTTACAGAGATACTCAGCCATCTGGGGAATCCTGATGAAAGAATGCCC
TGTAATGCCATGGTGCCATCATTTGCATTGTTTTGTTTTTGTTTTTTTTTCAAAGCTGGGTTGATGAAGAATTCACCTTGAATTTTTTCTG
GGAATCTTAAGTTGGATTCTTTTGAATGTTGATTGAGAAGAGGAAGCCTTTTTCCATTGCAGGAAAAGCTTTTGGCTCATCCTTTTCTTT
TTAAACAATTCTGATCTTTATAAGGCTGTAACCTCAGAAGACCCTTCATCCTGGGAGCTCACGGCAGCCCTGGAACCTCACATGTCTGAA
GTACCCTTCATCTCTGCGTCCTCTCCCTCACATCAGATGGCTTCTCCTAGGACTTCCCTTCAAGCCTCCCAGCTGAAATTCCATCATGCT
CTGCAAGGAACCCCTGATTGGGGAAGATGGGAAAGCTTGTTCTAAACAAGTGAAAACACTACCACCTTGACCGCCGTCTTCCTGCACACA
CTGATGAAGACAACCCAGGAGACTTGGGCATAGCAGTGCTTGTGACCAGTGTGCACAGCCGTCACCCTGAGAGGTGCCACCTGGCAATTG
TTCTTGCAGCAGCTACAGATGCAGCCATCAAAGACACCAAACGGAAAGTTCTTGGGACTGCTAAATGATTCAACATCACCAATGTTTACT
ATTTTATCAGTGCAATTGATGTTAAAGACAAATCCTTCCTTCCTTCCTTCCTTCATAAATGTGTTTTCAAACTCGGGAATTTTTTATATT
CATATATTTTTATATTTTTATATAATTTTTATATTTATATATTTATTCCATGTTGTTTTCTGTTCTATTAGATACATATCTTTAGAAAAA
CTATTTCACTCTTATTAAAATATTATTAAATTTTTCATTATTGGCCTGTAGCATAACTGAAACATTGTCCTTTCCAAACCAAATAACTCA
TTGAAATGCCTTTGCTTAACGGTTGTGTTACATGACTATGGGATTTTTCCAAATGGCTTGCAACTGAAATCTCCCAGATATGTTGTCTCT
CCAATTAGAGTGTGAGCTCCTGGAGGGCAAGAACCTTTTTACCTCTCTATTTGTGTCCCCATTAGCACAATGCTTTGTAGATAGTAAGCA
CTTAATAAATAACTTTTTTCATTTGTGCATTCATTCAAATAAATGTTTTTACACAAATTTTCCTTGACTTCTCATATCTTCTTTTCTACC
TTTTTGTCATTTTTGTCCCTTATGGTAGATTCTTGTTTTTGGAACTAATAGAAGAAAGAACTCAAACTCAAAGGTAATCTCAAGCAACAA
TTGCAGAATTAAGACTTGCTCAGCCTTTCTGCAGGACCAGTGGGTCCTTTCCAAGATAATAAAATACAAAACCAAGAAAAAAGTATTCAT
GTTCAAATTAAAAAAGTCATCTATCACTTGTGTCCAAAACTCCAGCCTTGCACCCTTCTAACAGATTGTGATTTTTCCTTTATCTGATTT
TCTGGTGCCTTCTAGTTACTGTA

>Chicken

GCTGACCCCGCGCCCCGTCCCGCTGCACTCAGGGGCGGGATCCGCGCTCTGCCCCGGGCTGCGGCCGGCGGCGGCGGCGGCGGCGGC
GGGGGGCGGTGACGTCACAGAGCCGCTCGCCAGCACAACGTTCCATTGAGGCGAGCGCGTGCTCCGGCGAGCACATGGGCAGAGCGGCCC
GCACCGCTCCGCTCCGCGCCGCTCCGCTCGGCCCCGCCGCCGCCGGACCCCCGGCCGGGGATTGGGAGGGGGGGGGCGCGGTCCCAGCCA
TGTGCCTTCAACCACTCTGGAGAAATGTTGGGAGGAGCTCTCAGTGCGATCCTTTCCATACCTTCCAAGGATTTTTAGGATCGTGGGATT
TATGGACCTGCAAAATCCTCTGGAGAACATCAGCCACCGACAAAACCATCTGGAAATTCCTGCTGTCTTTTTGCCCTGTCTTTCCTGAGT
CACCGAGTCTTGGGAAGGAAGCATCTCTGCTCTGAGGATTACAGCAAATCAGTTGCTTGCCGTCCTTACTGGAATACCTTCCCTTTCAGT
TGCTCAGAACATAGCAATTACTGAAGAAGGCAGAAGGATATTTTTTCCCTTTCACCGAGCTTTTGCAGCGCAGCAGTGGCTGTGTCCGGC
TATGTAGATGGTTTCCACATTCTCTACCCAATTTCTTTTCCTGATGTTTAGGAACTGAATCATTCTGTTTCCCTACTAGCCTCTGCTAAA
GGTCTCGCTTGACTGCAAAAAGGAGAAAGGAGGAGTGTGTGCAGGGAGTTCAGGTCCTGCTGTGGCTTCTTGGGAGGGTGAGTACTCAAAC
TTGCTTTATACAGAAGACAGCAGAAAGCCACAAAATTGGAGCCTCAGTAATTTCTTTACACCCACCATTTGCCAGTGTTCTGGATTGGAA
ACCAACTGCAACAGCTACTGAGAGTCTCATTCCTCCTCCACGTGAACCTCCTTCCAGGGAACTGTGAAAATTCCCACTTCCATCTCCTTC
GCATACTGGGGTTTCATGACCATTAAAACCCGTTCCTGGAACCATGGGATGGCCCCACAGCATGCTTTTGATCCAGCTCAGATGTGGGCT
TCTCCGTGGGATGAGTGCCCAGCAAGTCCAGCCATCAAAACCTACCTAGACCCTATGGAGCTGCATGCTCTGACTCATCCCTACTGAAAT
GTCCCCACACCCTGCCCCAGGGCTGTGCATTCCTGGGCACTGGTTGCCTGTGCCTCATGAGGTCTTTCTACCAAACTCATCACTTTGGCT
TGGTTCTGCATCACCACACTGAAACTGTGCAAACTGCAAGCCACCAAACTGTACCTGTGAGCCACCTCACGGGGCACATGTCCCACAGTG
GCTTGGCTGCTGGACCACTCAGGGTGGGATGCGCAACACACTGCAAGGAATTTAGTTCAGCACTGGTGCTGGTAGAGCACGGAGCTGGGA
GAGCAGCATCCCTAATTCAGGTTGATGCGTAGCTGTGTTGGAGCCCTTTCTGTAAACCACTTCAAAGGGCGTTTTGCTGTTTTGTCTCTT
CATTCACTGTTCTTTCCCACTTGTGCATTTTGTGCAATTTTAAAAACTTCTTGCATGTATTCCATTCCTCTTATCCTAGACCAAGTGGCT
ATCTTCTCTTCTTTCTTAGCTGGCTTGTTCATGCAGAGGCAAGGAGTAGCAAAACATGTTTTTCCTTCTGAAAAGAGAGAAAATAAATGT
TTGATACCTGATAATAGTTTTTGTTGTTTGTCTGAGGCAGCTTGAAATCCTTAGGGAAAATCAGAATTGTTCCATTAGGATGAGCTGCTT
GCAAGCATCCTGCCAAACTGGAACAGCCCTTCAAGCTGTACTGAAATGAATTCAAACCCTCCACAGATAAGCACAACATAGGGTATAATT
TTACTTCAACAGAAGGAAAAAGTCCAGTACATCATCTCCTTTTATTTCAGTGTCTGGTTGTCTTGACACCCATGTACATATAGAAATATG
CTTTCTTCACATGTATTTACTGAGGCACTCAAGCTAAATGCAATTGAGCCTGGTGGTTGCTGGGATGGTTCATCGTGCACGTCCAGGTGC
CATCTGTACACAGGGAAGCTGGCAATGCAGTAAATGGCTGTTTTCCCTTCTTTAAGAGATGCATGTTGGGGTTGACTTTTGATGATCACT
TAGGATCTTAAGCTTTTGTGCTTTTGCATTTACCTTTGTGTTGGCTAAATTTCAGTTCACTTACTTCCCTGTGAACAATTCTTTTTTTGT
TGAATAAACTATTCCTTGCAAACCATAACACTGATGTCATTGTTGTTGTGGGGGTTTTGTTTGGGTTGGGTTGGGTTTTAA

>Lizard

GGCGAGCGGGGAGCGCGTGCTCTGCAGAGCACATGGGCAGCAAAGCAGCCAAGGCAGCAACTCCATTGCTCCCTCCGCCTTCCCCTCGCT
GCACAGCCCAGGCCGTGGAAGCCTTTCCTAGACCCTTGTTTTTTTTTTTTTTTAATTTAAAGGATCCATATATATATATATATATTTT
GCAACCGACTGCAACTTGGATGTTTCTGGAGAACCTCCTCCTCCTCCTCCCCCTTTATCCGGCCTTGTTGTAGGGGATGATGGGTGG
TCTCTGGGCTTCAGTTCCCAACCAATCCGGGGGTTGCAGCAACATCTCCAAGGTTGTGGGATTTATTGAAACACTCAGAAGACTCCTGAG
AATGTCAGCTGCAGACTAAATGCCTTCAGGGGGAGGAATTCATTCTTTTGTACATTTCACGATTTGGTGCACATGTAGCACTGAAGATGG
ATTATTCTCTTATTATCACTGGAATATATTTTTTTGGAAAAGGAAACACAACAAAAGCACGAATCATTCGAATATAAGGCTGCCCAATCT
TCTATGTCATGCAAGCATGTTGCTGCCCAGATAATTGCAGAAGCAGTGGCCGAGGTTTGGAATCAGATCTGACACGGTGAAATGAAAGAC
GGCAGCAGATGCTGAGCGTAACAACGAAAGCGTATGCACAGGAGAAACACAACAGGATTTTTCTTTCTCCTTTCTTTCTTTCAAGCCGAT
TTCTCCTTTCAGCAGCACCCTCTGAGCTATCAGCTCACAACATGATGCTTACCTCATCCAACAGCTTCAACCCAAATCAAGCCAAAATAC
CAAGGAAATGCACAGAAACATCTTGACAGAGGGACGCAAGGAACAATTATGCCAGATACAGCTTTCCCCACTATTGCATTCATGTAATGG
GGACAGTTATACCTTCCAGAATATTTCCTTGGAGTTAGAAATTATAAATAATACCAGGGAGGAGCGGATGTATGTGCTTTCTATCTGTCC
ATCGCTGTAAGCTATTTTAAATAACAACAAAAACAAGATTGTGTTAGGGGTTTTCTTTAAGCCTTAAATTCCCCCAGATCGTATATACG
CAATTGGTCTTCACTACTAGAGTTATTTGCTTTCTCCATTGGAATACCTTGCTTTGGTGATACAAGGACTTCTACTATATACATCTCTGA
AAGTGGGCACCTAACCTCCATAACTGAAGCTCTGTGGAGGACATCATAACTGTGTTCTTTCCTGAACAGCCTCCCTTTCATCTAGCAAGC
CTTGCATTCCCCCAGTTTGCATTACCAAAGAAAAGAAGCAAAAGGGAGTTTGGTGTTAATTATATTCCAGAAGGTGACGATGGCTTAGCG

AGAGACCTGCCCTTACTGCTGGCTCCAAAGATCTGCCTAACATAAGACTTTCACAGATTCCTCCTAACACAAGTCTTCTTTTCACTTGGG
ATTTTTTTTTATTGCTTCTAAACACCTCTCCCCCTTCTTATGGGACTACATGGTGCTTCGTTTCAGATGGGCCCCGTGAATTAACATATT
GTACCTCCTTGCGGCTCAGGTAAATTGCATGCTCAGGTTGACCGCTGCTCAGACACAAAGCATCCTTCACCAGTTTA

>Zebrafish_2_lnc-*myca*

AGCGCGTGCCGCGGAGCGCGTGCTCTGAGCACATGGAAGAGGGCTCGAGACGAAGTATCCAAAGAATAACTCCACACCTGAAAGGGAATA
TCTGCAGAAACCGAATCAATTCTGGAACAAAAGAGGGACCTTCAGAACAGCTGCCGGCCTGCCTGTCGTTCTCTTGCTGGAAGCCTGAAA
CCAGAGCTAGGGGTGAATGCCAAACTTTTCAGGAAAAGACCTCCACATTCTGGCCCCCCGTGAGCTGACAGCTCTTTGCCATCAAAGTCA
ACAATACTGCTTTGCTGCAAGGAGATGAAGATCGGCTGTTGACTGCGGACAGCTGGGTTTCACACAATGGCATAAAGATGCGGAAAAGGG
TAAGGAATGCTGTGGTATTTGGAGTCGGACAGAAACCGGTATATAGGCAGCATATTTGTATATGTATTATAAACTTCAGCTTATACTGGC
GTTTTTGCATTATTTGTAATTGTTTATTAGTAATTTTAGCCCATAATAACCTATTTCTCCTTATTGGAAAGTTATGAGGCATATAATACA
GACCGAATGGACCACATATAACACGTTTCGGCAGTTGTTTTTGAAATGCAGCCCCTCTTTGTGAAAGATCTGTAATGTTATTAGGGATGT
AACAATTCAACTGACTTTCGATTCGATAGGATTCACGATACTAATCTCACGATTTATTCACAATCATTTTTAACACAATTATTTTTAAAT
AAAATGTTTTAAAACTCATTTAAGATGAAAAGCCCTCGATTTTTTTTCTTAACTGCTGCACATTTTTTGTTAAATAATATTTTCTGCCAC
AACCAAATTGCAGTCTGCATCCTCCGAAAGATGCGGACTTCAACGGTGAGTCCTTCGAAGTCCACACAGGCTGAACGGAGCGTTTTCTAC
CCAAGAGAGGACAGATTTCAGCACCAAGCAACCATAACAGCTGCTGATAATAAGCGGAAATAAAAATTGTAATGCATTTTTTTTTCAAAG
TGATTTTAAAACTTATTTTCAGCGATCTTAAGGCAAAGCAAATTTCCTTTGATCCATGAATTTACACATAAAAATGTAAACTGTCGACTC
TTTTTGTAATTTAACATGTGTTATTGGAAGGGGAATAAACTTAATTTATACATT

6.3 Supplementary Table 1: Relevant sequence of oligonucleotides

gRNAs		
Zebrafish		
	Zebrafish lnc-mycb 5'	GCTACACAGTTAAGGGGTGG
	Zebrafish lnc-mycb 3'	CAGTGGACAACTGGCTAATC
	Zebrafish lnc-myca 5'	GGACCAGCTGTCAGCTTGAG
	Zebrafish lnc-myca 3'	GGGGTATGGAAGTCTGGTGT
Mouse	mPVT1 TSS 5'	AAAGCTAGTTGACCTCGCCT
	mPVT1 TSS 3'	GAAGACAAGAATCGGATCCG
Primers qRT-PCR		
Zebrafish	ZF lnc-myca 5' F	GAAGAGGGCTCGAGACGAAG
	ZF lnc-myca 5' R	GCATTCACCCCTAGCTCTGG
	ZF lnc-myca 3' F	AGCGGAAATAAAAATTGCAATGC
	ZF lnc-myca 3' R	TTCCCCTTCCAATAACACATG
	ZF lnc-mycb 5' F	CCCAAGGCGTCCAAAAATGA
	ZF lnc-mycb 5' R	TCAGGAAACAACAAGCTCCA
	ZF lnc-mycb 3' F	CCTCTGTTGGCGGACTCAAA
	ZF lnc-mycb 3' R	AGAGCACTGGTACACGTGTG
	ZF EIF1a F	CTGGAGGCCAGCTCAAACAT
	ZF EIF1a R	ATCAAGAAGAGTAGTACCGCTAGCATTAC
	ZF myca F	CATGCTGGTCCTGGACACTC
	ZF myca R	CCTCCTCTTCTTCCTCCTCCTC
	ZF mycb F	CTCACGCTGACATCTGACCA
	ZF mycb R	AGGGCTGGTAGGAGTCGTAG
	ZF tp53 F	CTTGCCCCGTTCAAATGGTG
	ZF tp53 R	GCTGATTGCCCTCCACTCTT
	ZF p21 F	ACACTTGGCGACGTTCATGA
	ZF p21 R	TGTGCACGTAAACTTTAAGCCT
	ZF Malat1 F	CGTCACCTGAATGCAAGTGC
	ZF Malat1 R	CCCCAAAATCCCCAGTCGAA
	ZF GAPDH F	GTGGAGTCTACTGGTGTCTTC
	ZF GAPDH R	GTGCAGGAGGCATTGCTTACA
Mouse	PVT1 A F	CGAGCACATGGACCCACTG
	PVT1 A F	AACATCCTTTCCACGCGGATAT
	PVT1 B F	AGTAAGATGCAGTTTCCTGGGG
	PVT1 B R	GTCCTTGTTTTCAGTGGGCATC

	PVT1 C F	TAAGCAGAACCTTGACCCCTTC
	PVT1 C R	CCCCTTCTTTTTCACAAGCCAG
	GAPDH F	ATCAAGAAGGTGGTGAAGCGGCA
	GAPDH R	TGGAAGAGTGGGAGTTGCTGTTGA
	Myc F	CTGTTTGAAGGCTGGATTTCCTTTG
	MYC R	CTGGTGATAGAAATTCTCTTCCTCGTC
	Tuj1 F	CATAGACCCCAGCGGCAA
	Tuj1 R	AAAGGCGCCAGACCGAAC
	Nestin F	CTCCTGTGACAGCCTTTCTGAAG
	Nestin R	AGGATAGGGAGCCTCAGACATAGG
	Oct4 F	CCCCAATGCCGTGAAGTTG
	Oct4 R	TCAGCAGCTTGGCAAACTGTT
Primers Genotyping		
Zebrafish	lnc-myca F	AATCTTGTCTGTCCCTATAAACAACAGAAACT
	lnc-myca intern R	ATGGTTGCTTGGTGCTGAAATCTGTCCTCTC
	lnc-myca R	CGATATGTACAATGAATAGGTGTCCACATAGT
	lnc-mycb F	CGGTGAGTTCATCTAAAAATCAATGACAATGT
	lnc-mycb intern R	TGAGTTGTGGGAAAGTTGGCATTACAGGAGCG
	lnc-mycb R	ATGAGTAAATGGAAATGAACATACTGTGAGCC
Mouse	mPVT1 TSS F	CTTCTCTCCAGAACCACAGATGATGGATGGTAC
	mPVT1 TSS R	CTAGATTGGGCAGGCAGACAGAAGGTCCCGGACATGG
	mPVT1 TSS intern F	TGCAGGTCTCTGTTGAAGTGCATGTGGTAGGATGTTG
IncPRINT construct	mPVT1 F	CGTGACGTCACGGGCCACCCGCCAGCCCCGCGCTCTAC
	mPVT1 R	TGGTCAAATTGTTTTATTGAGCTTCATTACTTAATAAAGC
	lnc-mycb F	TGCTGTGAGCACATGGAGGAGAGAGGACGTTCAC
	lnc-mycb R	TTGGCAGTTTTGGTGTCACAGTTTGACTTAAC
	miniPVT1 1	Genscript
	miniPVT1 scr 1	Genscript
	minipvt1 2	Genscript
	miniPVT1 scr 2	Genscript

7. FRENCH SUMMARY

7.1 Introduction

7.1.1 Le flux de l'information génétique

Watson et Crick publiaient pour la première fois, le 25 avril 1953, la structure moléculaire de l'ADN (Watson & Crick, 1953), en s'appuyant sur les travaux de Franklin. Cette révolution en génétique ouvrait la voie sur la compréhension de la circulation de l'information génétique dans une unité biologique. Quatre ans plus tard, en septembre 1957, Francis Crick, lors de sa conférence, exposait les principes de ce qui deviendra la théorie fondamentale de la biologie moléculaire, en poussant plus loin son modèle de flux de l'information génétique (Cobb, 2017 ; Crick, 1970). Ce théorème peut être résumé par : « l'ADN dirige sa propre réplication en ADN identique, ainsi que sa transcription en ARN, pouvant ou non être traduit en protéines ». Ce dernier permit de comprendre le transfert de l'information de l'ADN vers la protéine, que l'on pensait être la seule unité fonctionnelle principale de la cellule. Ainsi, l'ADN est le support de l'information génétique qui est ensuite traduite en ARN. La molécule d'ARN est utilisée par la suite comme modèle pour la synthèse des protéines. Il est intéressant de noter que, contrairement aux connaissances générales, Crick avait déjà suggéré la possibilité que l'ARN se réplique, dirige la polymérisation de l'ADN et que l'ADN soit directement le modèle de la synthèse des protéines, les qualifiant de "transferts spéciaux" (Crick, 1970).

Au cours des années suivantes, de nombreux efforts ont été déployés pour élucider la composition des types de molécules d'ARN impliquées dans le flux d'information. Compte-tenu de leur rôle central, les ARN messagers (ARNm) ont été largement étudiés depuis leur découverte en 1961 (Jacob & Monod, 1961). Parallèlement, des ARN non codants (ARNnc) ont été découverts et leurs rôles cellulaires précisés, tels que les ARN ribosomiques (ARNr) (Palade, 1955), les ARN de transfert (ARNt) (Hoagland et al., 1958), et les petits ARN nucléaires (snRNA) (Hadjiolov et al., 1966 ; Weinberg & Penman, 1968). Les parties restantes du génome étaient alors appelées ADN « poubelle » (Ohno, 1972), car on pensait que ces fragments d'ADN ne jouaient aucun rôle dans l'espèce.

Ce n'est que dans les années 80/90 qu'une vision plus complexe du transcriptome est apparue. Les nombreuses découvertes comme le premier ARNnc régulateur chez E. coli (Coleman et al., 1984), la première identification d'un locus non codant, le locus *PVT1* (Graham & Adams,

1986), le premier long ARN non codant (lncARN) humain fonctionnel H19 (Brannan et al., 1990), ainsi que le lncARN Xist (Brockdorff et al., 1991, 1992 ; Brown et al., 1992) et les microARN (miARN) chez C. elegans (Lee et al., 1993), ont ouvert la voie à un rôle plus complexe de la molécule d'ARN.

Les deux dernières décennies ont marqué le début d'une expansion dans la compréhension des fonctions biologiques de ces ARNnc accompagnée d'une amélioration considérable de la technologie. Le séquençage complet du génome humain a permis de mieux comprendre la structure de l'ADN chez l'homme (Lander et al., 2001 ; Venter et al., 2001, Nurk et al., 2022). Enfin, avec l'émergence du séquençage de l'ARN, des grands consortiums (ENCODE et FANTOM) ont multiplié leurs efforts pour permettre une perception globale du génome et du transcriptome des mammifères (Carninci et al., 2005 ; Katayama et al., 2004 ; The ENCODE Project Consortium, 2007). Ces découvertes ont révélé que 80% du génome est transcrit en ARN alors que seulement 1% du génome est transcrit en ARNm (Djebali et al., 2012). L'étude du transcriptome humain a identifié entre 20 000 et 60 000 lncARN potentiellement fonctionnels (Hon et al., 2017 ; Iyer et al., 2015), sans inclure les autres classes d'ARNnc, alors que 20 000 à 25 000 gènes codants pour des protéines ont été identifiées (Lander et al., 2001). Ces analyses ont révélé la haute importance des ARNnc dans le transcriptome humain.

Le transcriptome est donc composé d'ARN codants et non-codants. Les ARN non-codants représentent la majorité des ARN présents dans le transcriptome (Djebali et al., 2012). Parmi les ARNnc, on peut trouver plusieurs sous-classes : les ARN domestiques et les ARN régulateurs. La sous-classe des ARN domestiques est composée d'ARN essentiels au maintien des fonctions cellulaires de base. Ils sont généralement exprimés de manière constitutive (comme les ARNr, les ARNt, les snoARN, les snARN). En revanche, l'expression des ARN régulateurs est contrôlée dans l'espace et dans le temps. Ces ARN sont essentiels à la finesse du réglage de l'expression des gènes et des fonctions des protéines (comme les miARN, piARN et lncARN).

Les ARN domestiques

- Les petits ARN nucléaires (snRNA) sont des composants essentiels dans l'épissage des pré-ARNm des transcrits de la polymérase II.

- Les ARNr sont la classe d'ARN la plus représentée dans le transcriptome, 80% des ARN sont des ARNr. Ces ARN sont les constituants principaux des ribosomes, composés à 60% d'ARNr et complétés avec des protéines. Le ribosome est la machinerie essentielle pour la traduction des ARNm en protéines.

- Les petits ARN nucléolaires (snoARN) sont une famille d'ARN jouant un rôle central dans la maturation des ARNr. Ces snoARN sont impliqués dans la maturation, la modification de base azotée et la stabilisation des ARNr.

- Les ARNt font le lien entre les ARNm et les acides aminée constituant les protéines. Chaque ARNt est couplé à un acide aminé et s'apparie, dans le ribosome, sur l'ARN, par complémentarité de base. Les ARNt représentent la plus grande famille de gènes dans le génome humain, avec environ 500 gènes codants pour des ARNt.

Ces exemples non exhaustifs d'ARNnc domestiques mettent en évidence le rôle essentiel des ARNnc dans les fonctions cellulaires. Ces ARN sont indispensables au bon fonctionnement de la cellule, ils agissent principalement dans la maturation des pré-ARNm et la synthèse des protéines. Ils sont donc essentiels pour faciliter le transfert de l'information génétique des gènes aux protéines.

Les ARN régulateurs

Des efforts considérables ont été déployés pour la caractérisation et la compréhension des ARNnc régulateurs. En effet, ces ARN sont d'un grand intérêt notamment en santé humaine, comme le montre le nombre élevé de rapports publiés associant les micro-ARN, et les lncARN à des maladies humaines tel que le cancer (Bhan et al., 2017 ; Calin & Croce, 2006 ; Carlevaro-Fita et al., 2020). Les ARN régulateurs sont classés en fonction de leur taille, en deux catégories : les petits ARNnc (<50 nt) et les lncARN (>200 nt).

- Les micro-ARN (miARN) sont des ARN endogènes et conservés dans toutes les espèces. Ces ARN régulent l'expression des gènes en interagissant avec eux par complémentarité de base et en induisant la dégradation de l'ARN ciblé.

- Les ARN interagissant avec les protéines PIWI (ARNpi) sont une classe d'ARN avec un mode d'action similaire au miRNA. Mais, ces ARN sont spécialement exprimés dans les gamètes. Ils sont donc les gardiens du génome des gamètes utilisées pour la reproduction.

- Les ARN longs (>200nt) régulateurs sont transcrits par l'ARN polymérase II. Les lncARNs présentent un grand intérêt en raison de leurs fonctions. Alors que les petits ARNnc sont spécialisés dans la régulation de l'expression des gènes c'est-à-dire dans la régulation fine du flux de l'information génétique (Bartel, 2009), les lncARN ont des fonctions moléculaires diverses et sont des effecteurs directs de cette information génétique (Constanty & Shkumatava, 2021 ; Hansen et al., 2013 ; Kopp & Mendell, 2018 ; Kristensen et al., 2019 ; Marchese et al., 2017 ; Memczak et al., 2013 ; Statello et al., 2021). De plus, ce dernier groupe est le plus large car il est composé de différents types de lncARN. On peut trouver deux formes distinctes de lncARN, les ARN circulaires, refermés sur eux même, et les ARN linéaires.

Au fil des années, la vision du transcriptome humain, et plus généralement des métazoaires, est devenue plus complexe. La découverte des fonctions canoniques et non-canoniques des ARNnc domestiques et la diversité des ARNnc régulateurs montrent l'importance des ARNnc en tant que molécules dans le bon fonctionnement de la cellule et de l'organisme.

7.1.2 Les long ARN linéaires

Les lncARN linéaires (ci-après dénommés lncARN) constituent une classe complexe d'ARNnc. La principale caractéristique des lncARN est leur incapacité à coder une protéine ou une protéine fonctionnelle. Cependant, il a été récemment démontré que certains transcrits, précédemment décrits comme des lncARN, codent des protéines fonctionnelles. Ces observations ouvrent la porte à un débat sur la nomenclature des lncARN.

<u>L'origine évolutive des lncARN</u>

L'évolution des lncARN est particulièrement remarquable. Alors que la majorité des lncARN humains sont considérés comme non conservés (Ruan et al., 2020), il a été déterminé, par conservation de la position génomique, que des centaines de lncARN ont des homologues chez les vertébrés. Seulement 18 lncARN ont été trouvés conservés de l'homme à l'oursin (Hezroni

et al., 2015). L'observation que la plupart des lncARN sont peu conservés en dehors du sous-phylum des vertébrés montre que les lncARN ont une origine dynamique. Plusieurs mécanismes sont à l'origine de l'apparition des lncARN. La perte du potentiel codant d'un gène codant peut conduire à l'apparition de lncARN. Les mutations au sein des gènes codants peuvent entraîner la perte de la protéine. Si le transcrit n'est pas déstabilisé par la mutation et que le gène reste exprimé, alors cela peut conduire à de nouveaux lncARN. Le lncARN Xist a évolué en tant qu'ARNnc par ce mécanisme (Duret et al., 2006). Les éléments transposables sont également à l'origine de la formation de nouveaux lncARN. Ils sont fortement impliqués dans les innovations du génome chez les mammifères (Cordaux & Batzer, 2009). Les insertions d'éléments transposables contenant des promoteurs fonctionnels peuvent conduire à l'expression de nouveaux lncARN. Ceci est supporté car 80% des lncRNAs chevauchent au moins un élément transposable (Kapusta et al., 2013). Les insertions d'éléments transposables peuvent également conduire à de nouveaux lncARNs sans insertion de promoteurs fonctionnels. Des études ont montré que la transcription se produit souvent dans les directions divergentes des gènes codants et à partir des enhancers (Seila et al., 2008). Cependant, les produits de ces transcriptions sont rapidement dégradés. Les insertions d'éléments transposables peuvent stabiliser ces transcriptions cryptiques conduisant à l'apparition de nouveaux transcrits stabilisés codants ou non codants (Gotea et al., 2013 ; Wu & Sharp, 2013). Il est remarquable que l'annotation des ARN codants ou non codants change au cours de l'évolution. Les transcrits peuvent apparaître en tant que lncARN chez les vertébrés basaux et évoluer en transcrits codants des protéines chez les mammifères ; comme le transcrit libra/Nrep, non codant chez le poisson zèbre, et codant chez les mammifères (Bitetti et al., 2018).

Conservation des lncARN chez les vertébrés

Peu après l'identification des lncARN chez différentes espèces, des études de comparaison génomique ont vu le jour pour atteindre trois objectifs : (1) annoter et identifier les lncARN chez les vertébrés, (2) élucider l'évolution de ces transcrits, et (3) utiliser les lncARN conservés comme indicateurs de leur importance fonctionnelle pour les distinguer des gènes transcrits de manière systématique.

La position chromosomique, également connue sous le nom de synténie, est la meilleure approche pour identifier les lncARN conservés chez les vertébrés. Elle peut être utilisée comme première couche pour identifier les lncARN conservés par comparaison génomique avant de validations plus poussées. Il est intéressant de noter qu'en utilisant cette méthode, des centaines

de lncARN ont été trouvés conservés parmi les vertébrés (Hezroni et al., 2015), dont certains de l'homme au poisson zèbre (Ulitsky et al., 2011). Les génomes du poisson zèbre et des mammifères ont été largement réagencés au cours des 400 millions d'années de séparation, cependant, la conservation de la synténie des lncARN (33,9 %) est significativement plus élevée que la conservation des paires de gènes codants pour les protéines (14,7 %) (Ulitsky et al., 2011). Par conséquent, seule la position chromosomique d'un génome annoté permet d'identifier les lncARN conservés à travers les espèces grâce à la synténie.

Fonctions moléculaires des lncARN

Les loci de lncARN peuvent agir selon différents mécanismes moléculaires. Ils peuvent avoir (1) une fonction par la présence d'enhancers régulant l'expression des gènes voisins, (2) une transcription, et non le transcrit, peut être impliquée dans la régulation de l'expression des gènes environnant, et (3) la fonction moléculaire peut être portée par le transcrit lui-même. Il est intéressant de noter que les loci de lncARN peuvent avoir une ou plusieurs de ces fonctions. Cependant, les loci les plus intrigants sont ceux dont la fonction est portée par le transcrit. En effet, ces transcrits peuvent eux-mêmes agir à différents niveaux moléculaires (Constanty & Shkumatava, 2021 ; Statello et al., 2021). Ils peuvent interagir avec l'ADN, l'ARN ou les protéines. La grande majorité des lncARN interagissent avec des protéines. Ce sont ces protéines qui confèrent la fonction moléculaire des lncARN. Par conséquent, le déchiffrage des déterminants fonctionnels des lncARN guidant l'interaction ARN/protéine est étudié afin de permettre la prédiction de l'activité des lncARN. On a d'abord supposé que la structure de l'ARN porte la fonction du lncARN (Diederichs, 2014 ; Guttman & Rinn, 2012 ; Wutz et al., 2002), alors que cela est le cas que chez quelques lncARN (Holmes et al., 2020 ; Uroda et al., 2019 ; Zhang et al., 2010). Des études non biaisées sur les sites de liaisons des protéines ont montré des motifs d'ARN souvent peu complexes et bipartites (Dominguez et al., 2018). La spécificité des liaisons de ces protéines est cependant soutenue par la structure locale de l'ARN (Dominguez et al., 2018). Ces résultats montrent que la fonction moléculaire des lncARN est portée par la séquence primaire et leur structure locale plutôt que par la structure globale de l'ARN. Plusieurs motifs d'ARN courts ont été découverts pour guider la fonction d'un grand nombre de lncARN agissant dans un large éventail de processus biologiques. Ces motifs se sont révélés essentiels pour le traitement de l'extrémité 3' des lncARN, les formations de R-loop, la localisation subcellulaire des lncARN, la séquestration des protéines par les lncARN, la régulation des interactions lncARN-protéines et la régulation des miARN (Constanty & Shkumatava, 2021).

7.1.3 Le lncARN *PVT1* (*Plasmacytoma Variant Translocation 1*)

Le lncARN *PVT1* a été nommé d'après l'implication du locus génomique dans des translocations génomiques, dans le plasmocytome et du le lymphome de Burkitt (Cory et al., 1985 ; Graham & Adams, 1986 ; Webb et al., 1984). Ce lncARN est d'un grand intérêt car son expression est associée à de multiples maladies humaines (Ghetti et al., 2020 ; Guo et al., 2022 ; Hanson et al., 2007) dont des cancers (Onagoruwa et al., 2020 ; Tseng et al., 2014 ; Tseng & Bagchi, 2015). Néanmoins, sa fonction moléculaire reste mal comprise. Chez l'homme, *PVT1* est localisé sur le chromosome 8, en aval du gène codant pour la protéine MYC. Il est intéressant de noter que cette caractéristique génomique spécifique est conservée chez tous les vertébrés (Hezroni et al., 2015).

Le locus *PVT1* est d'une grande complexité. D'une taille d'environ 400kb, ce locus abrite plusieurs enhancers régulant l'expression de MYC et de *PVT1*. Ce locus code pour 5 miRNAs, et code pour de multiples variants linéaires du transcrit *PVT1* et un variant circulaire. Cette grande complexité montre la difficulté d'établir de façon précise la fonction de ce locus. En effet, chaque variant peut avoir une fonction différente. Typiquement cet lncARN est un très bon exemple de locus fonctionnel dû à la présence d'enhancer et de transcrits fonctionnels. Au fil des années, plusieurs études ont été consacrées à la fonction du transcrit *PVT1*, principalement dans le cancer (Onagoruwa et al., 2020). Il est intéressant de voir qu'une grande variété de mécanismes moléculaires a été trouvée, en fonction du modèle étudié, conduisant à une généralisation complexe de la fonction de ce lncARN. Pour des raisons de clarté, les parties suivantes se concentreront sur les variants linéaires du lncARN PVT1.

L'ARN *PVT1* peut agir comme un oncogène. Il a été montré que l'expression de *PVT1* pouvait être induite par les protéines FOXM1 (Xu et al., 2017) et MYC (Tseng et al., 2014) dans différents types de cancer. Ces deux protéines peuvent également interagirent avec *PVT1* ce qui les stabilisent et inhibent leur dégradation. Ceci montre un effet de boucle de régulation positive entrainant le développement de tumeur (Tseng et al., 2014, Xu et al., 2017).
De façon étonnante, il a aussi été montré que l'expression de *PVT1* peut être induite pour la protéine suppresseur de tumeur p53 (Olivero et al., 2020). P53 induit l'expression d'un variant spécifique de *PVT1* appelé *PVT1b* et son induction limite le développement tumoral (Olivero et al., 2020). Ceci montre deux effets contraires de l'ARN *PVT1*, agissant comme un oncogène

et comme un suppresseur de tumeurs. Cette double fonction de *PVT1* est dépendante du variant exprimé mais s'ajoute à la complexité du locus montrant la difficulté de l'étude de cet ARN.

7.1.4 Objectif de la thèse

Bien que *PVT1* soit l'un des premiers lncARN identifiés dans le génome humain et qu'un nombre important d'études aient tenté de déterminer sa fonction, la plupart des études se concentrent sur l'implication de *PVT1* dans les cancers humains. La caractérisation du mode d'action de *PVT1* dans le cancer est un défi. Ainsi, déchiffrer le mode d'action de *PVT1* dans d'autres modèles aidera considérablement à mieux caractériser son mode d'action. *PVT1* représente un cas fascinant d'évolution des lncARN. Il est conservé chez les vertébrés par conservation de la position génomique, situé en aval du gène codant pour la protéine MYC (Hezroni et al., 2015). Il est intéressant de noter que les transcrits homologues de *PVT1* chez les mammifères partagent une conservation de séquence alors qu'en dehors de cette classe, la conservation de séquences n'est pas présente. Ainsi, comprendre la fonction conservée du lncARN *PVT1* chez les vertébrés améliorera grandement la caractérisation de la fonction de *PVT1* et fera la lumière sur les mécanismes conservés des lncARN sans conservation de séquences linéaires. C'est pour cela que nous étudions l'implication de *PVT1* chez le poisson zèbre.

7.2 Résultats

7.2.1 Conservation de petits motifs ARN

Chez le poisson zèbre, dû à une duplication complète du génome pendant l'évolution des poissons téléostéens, deux transcrits homologues au transcrit *PVT1* humain ont été identifiés par conservation de position génomique, nommé lnc-*myca* et lnc-*mycb*. Parce que la localisation subcellulaire des lncARN est primordiale pour leurs fonctions (Chen, 2016 ; Guo et al., 2020), nous avons dans un premier temps, établi que les ARN du poisson zèbre sont tous deux enrichis au niveau nucléaire, comme le transcrit humain. Ainsi, la fonction moléculaire de *PVT1* pourrait être conservée chez les vertébrés. La conservation de séquences linéaires supporte généralement la conservation fonctionnelle des lncARN. Cependant, les transcrits homologues des mammifères et du poisson zèbre ne partagent pas de conservation de séquences (Hezroni et al., 2015). Néanmoins, les motifs ARN sont apparus comme une

alternative pour décoder les fonctions moléculaires conservées entre les transcrits à évolution rapide, agissant comme des sites de liaison aux protéines (Kirk et al., 2018 ; Ross et al., 2021). Ainsi, nous avons utilisé le nouvel algorithme lncLOOM (Ross et al., 2021) et identifier des motifs ARN conservés sur les transcrits de *PVT1* chez les vertébrés. Nous avons découvert que deux motifs restent conservés sur tous les transcrits des espèces testés (11) TGGATT et TTTTTG. Afin de comprendre l'implication de ces motifs chez le poisson zèbre, nous avons utilisé la technologie CRISPR-Cas9, appelée également ciseaux moléculaires, pour retirer les régions entourant ces motifs du génome du poisson zèbre.

7.2.2 La manipulation génétique des motifs ARN entraîne de nombreux défauts de développement

Afin d'éviter tout mécanisme de compensation des deux transcrits du poisson zèbre, nous avons généré des mutants doubles homozygotes pour les ARN lnc-*myca* et lnc-*mycb*. Pour obtenir ces mutants, nous avons croisé des poissons doubles hétérozygotes. Lors d'un tel croisement, nous nous attendons à obtenir 1/16 de poissons de type sauvage (WT) et 1/16 de doubles mutants homozygotes. Cependant, nous avons obtenu 1/50 de poissons doubles mutants homozygotes à l'âge adulte (au moins 2 mois post-fertilisation) ; alors qu'à 2 jours post-fertilisation (jpf), nous avons retrouvé le bon nombre de doubles mutants homozygotes. Ceci montre que la manipulation génétique des motifs ARN entraine une faible survie jusqu'à l'âge adulte. Une certaine fraction d'embryon double mutant homozygote présente de sévères défauts de développement à 3 jpf. Ces défauts sont exacerbés à 6 jpf. Les embryons présentent un œdème péricardique et autour du vitellus, une malformation de la tête, des yeux et de la mâchoire. Enfin, ces embryons ne présentaient pas d'inflation de la vessie natatoire indispensable pour leur mobilité dans leur environnement en trois dimensions. Ceci montre que les motifs ARN présents sur les transcrits du poisson zèbre homologues au *PVT1* humain peuvent être impliqués dans le développement embryonnaire. Les embryons mutants doubles homozygotes ne présentent pas tous des défauts de développement, environ 30% des doubles homozygotes mutants présentent ces défauts, visibles à partir de 3 jpf.

7.2.3 Les défauts de développement sont associés à une augmentation de la voie régulant le stress

En vue de déchiffrer l'impact de la délétion des motifs ARN, nous avons étudié l'expression des gènes chez les WT et les doubles mutants homozygotes à 3 jpf. Nous avons remarqué que les mutants avaient un taux d'expression de la voie de régulation du stress plus élevé que chez les WT. La voie de régulation du stress est importante lors du développement embryonnaire afin d'assurer le bon développement des tissus sans erreurs de divisions cellulaires. Cette surexpression spécifique des embryons présentant des défauts de développement montre que ces défauts sont associés à l'activation de la voie de régulation du stress. En cas de stress cellulaire telles que des cassures de l'ADN ou stress ribosomique, cette voie est induite afin de le réguler. Lorsque le stress persiste, la voie de régulation du stress entraine l'apoptose. Nous avons, en effet, observé un plus grand nombre de cellules apoptotiques chez les embryons avec des défauts de développement. Ceci démontre que la délétion des motifs ARN des transcrits lnc-*myca* et lnc-*mycb* entraine une susceptibilité accrue au stress entrainant la mort cellulaire. C'est certainement l'accumulation de cellules apoptotiques qui entraine les défauts de développement.

7.2.4 La protéine myc est surexprimée lors de l'apparition des défauts de développement

Le gène *myc* est situé en amont de *PVT1* chez les vertébrés. Parce que les lncARN peuvent réguler l'expression des gènes en *cis* (Gil & Ulitsky, 2020), nous avons étudié l'impact de la délétion des motifs ARN sur l'expression de *myc* chez le poisson zèbre. L'expression de *myc* n'était pas impactée, au niveau de l'ARN, à 3 jpf. En revanche, au niveau protéique, myc était surexprimé chez les doubles mutants homozygotes. La surexpression de la protéine myc était corrélée avec une surexpression des transcrits tronqués de *pvt1* chez le poisson zèbre, à 3 jpf. Ceci peut suggérer que la stabilisation de myc par les ARN *pvt1* se produit également chez le poisson zèbre et suppose donc que la délétion des régions autour des motifs ARN n'impacte pas l'interaction *PVT1/*MYC. L'étude des motifs de liaisons protéiques présents au niveau des promoteurs des gènes *pvt1* chez le poisson zèbre a montré la présence de E-box aux promoteurs de ces gènes, site de reconnaissance de la protéine myc. Myc pourrait donc induire leurs expressions. Ainsi, comme chez l'homme (Tseng et al., 2014), les ARN *pvt1* et la protéine myc

chez le poisson zèbre pourraient être impliqués dans une boucle de régulation positive stimulant la stabilisation de la protéine myc et son activité.

7.2.5 L'expression de *pvt1* est induite par la voie de régulation du stress chez le poisson zèbre

L'étude des motifs de liaisons protéiques au niveau des promoteurs des ARN *pvt1* chez le poisson zèbre a également révélé la présence d'un site de reconnaissance de la protéine p53 uniquement autour du promoteur de lnc-*myca*. P53 est une protéine clé de la voie de régulation du stress. Son activité est induite par la détection de stress cellulaire. P53 induit l'expression de gènes essentiels à la régulation du stress ou à la mort cellulaire (Williams & Schumacher, 2016). Afin de tester la régulation de lnc-*myca* en condition de stress, nous avons placé les embryons de poisson zèbre en présence de camptothécine (CPT), un inhibiteur de topoisomérase, ce qui induit des cassures d'ADN. Les embryons traités avec le CPT présentent une surexpression de p53 et de lnc-*myca* mais pas de lnc-*mycb*, montrant la spécificité de l'expression lnc-*myca* qui pourrait avoir une fonction dans la voie de régulation du stress. Afin de comprendre si p53 induit directement l'expression de lnc-*myca* dans des conditions de stress, nous avons exposé des embryons préalablement injectés avec des morpholino inhibant l'expression de p53 au CPT. L'inhibition de l'expression de p53 proscrit l'activation de la voie de régulation du stress même en présence de CPT, ce qui inhibe l'activation de l'expression de lnc-*myca*. Ceci montre que l'expression de lnc-*myca* est spécifiquement induite par p53 en condition de stress chez le poisson zèbre. Par ailleurs, il a déjà été démontré qu'un variant spécifique de *PVT1* chez la souris et chez l'homme est induit directement par p53 (Olivero et al., 2020), démontrant une conservation de fonction des transcrits homologues de *PVT1* chez les poissons-zèbres dans la régulation du stress cellulaire via la voie p53.

7.2.6 La délétion partielle de *pvt1* chez le poisson zèbre impacte les érythrocytes

L'expression des lncARN est tissu-spécifique. Nous avons donc analysé les tissus dans lesquels est exprimé *Pvt1* pendant le développement de la souris avec des séquençages ARN de cellules uniques déjà établies (Cao et al., 2019; Pijuan-Sala et al., 2019). Les deux séquençages d'ARN étudiés montrent que *Pvt1* est spécifiquement exprimé dans les progéniteurs des érythrocytes. Afin de comprendre l'impact de la délétion des motifs ARN des transcrits de *pvt1* chez le poisson zèbre dans ce tissu, nous avons marqué les érythrocytes actifs chez les embryons WT

et les doubles mutants homozygotes à 2 jpf. Les doubles mutants homozygotes présentaient un taux plus faible d'érythrocytes actifs que les WT. Les érythrocytes sont essentiels dans l'acheminement de l'oxygène vers les organes. Une baisse du taux d'érythrocytes actifs peut être engendrer de l'anémie. Les érythrocytes sont le type cellulaire majeur dans la prise en charge du stress oxydatif dans l'organisme, en raison de leur fonction principale de transporteur d'oxygène (Kramer et al., 2017 ; Scott et al., 1991). Nous avons testé si la résistance oxydative chez les mutants doubles homozygotes était altérée. Cependant, pour éviter tout biais dû aux défauts de développement, nous avons sélectionné uniquement les embryons sans défaut de développement à 3 jpf. Les embryons homozygotes doubles mutants étaient plus sensibles au stress oxydatif, avec environ deux fois plus d'embryons qui mouraient après un traitement au peroxyde d'oxygène que les embryons WT. Ces résultats montrent que les mutants, bien que ne présentant pas de défaut de développement visible, sont plus sensibles au stress oxydatif. Ces résultats montrent qu'indépendamment des défauts de développement, les embryons doubles mutants homozygotes ont un taux d'érythrocytes actifs plus faible que chez les WT. Comme chez la souris, *pvt1* chez le poisson zèbre joue un rôle dans le développement des érythrocytes.

7.2.7 Les transcrits de *PVT1* interagissent avec des protéines essentielles dans la réparation des cassures d'ADN

Nous avons montré que les transcrits homologues de *PVT1* chez les vertébrés partagent des motifs ARN conservés. Ces motifs ARN sont récemment apparus comme une alternative à la conservation de séquences pour déterminer les mécanismes moléculaires conservés des lncARN entre les espèces (Ross et al., 2021). Les lncARN agissent principalement par le biais d'interactions avec les protéines et les motifs ARN identifiés sur *PVT1* et pourraient correspondre à des sites de liaisons protéiques. Ainsi, nous avons analysé les motifs identifiés sur les transcrits de *PVT1* afin de déterminer s'ils correspondent à des sites de liaisons protéiques. Pour analyser l'interactome de *PVT1* chez les vertébrés, nous avons utilisé notre technologie incPRINT récemment publiée (Graindorge et al., 2019).

IncPRINT a permis d'identifier des protéines qui se lient aux transcrits de souris et de poisson zèbre (lnc-*mycb*). 32 protéines interagissent de manière spécifique et similaire avec les transcrits des deux espèces. Ces protéines sont enrichies pour des fonctions nucléaires (organisation de la chromatine, méthylation de H3K9), conformément à la localisation subcellulaire précédemment établie des transcrits *PVT1* de mammifères et de poissons-zèbres.

Dans le but de décrypter l'interaction spécifique induite par les motifs ARN, nous avons crypté ces motifs et analysé le score d'interaction des protéines identifiées (Fig 13F & G). Le premier motif, TGGATT, a été remplacé par GTATGT et le second, TTTTTG, par TCGACG. Il est intéressant de noter que, bien que les deux motifs soient différents, la plus forte baisse du score d'interaction lors du remplacement des motifs est associée à la même protéine pour les deux motifs : RAD51C. De manière intrigante, la diminution du score d'interaction, pour le premier motif, était approximativement similaire entre les protéines RAD51C et NUSAP1, deux régulateurs clés de la voie de réparation de l'ADN (Kotian et al., 2014 ; Somyajit et al., 2012). De façon intéressante, ces résultats sont cohérents avec l'implication de *PVT1* dans la voie de régulation p53. En effet, p53 est la protéine essentielle dans la voie de réparation de l'ADN.

7.3 Conclusion

Cette étude représente la première analyse in vivo de la fonction physiologique du lncARN *PVT1* associée au cancer. La délétion des motifs ARN entraîne des défauts de développement sévères, montrant que *PVT1* est impliqué dans le développement. Nous avons montré que les caractéristiques et la fonction du lncARN PVT1 pourraient être conservées chez l'homme et chez le poisson zèbre: (1) par la présence de E-box aux promoteurs des gènes *pvt1* chez le poisson zèbre, montrant que myc peut réguler l'expression de ces gènes ; (2) ainsi que par l'implication du transcrit lnc-*myca* dans la voie p53, (3) et la délétion partielle des ARN lnc-*myc* ayant un impact sur la lignée érythroïde ; lignée dans laquelle *Pvt1* a été trouvé principalement exprimé pendant le développement des souris ; (4) enfin, par l'identification de deux protéines qui semblent se lier spécifiquement aux motifs ARN présents chez tous les transcrits *PVT1* de vertébrés. Ces deux protéines sont impliquées dans la réparation de l'ADN. Ceci pourrait justifier la fonction moléculaire conservée de *PVT1* chez les vertébrés, régulant le développement embryonnaire via la réparation des cassures d'ADN.

BIBLIOGRAPHY

Abdelmohsen, K., Panda, A. C., Munk, R., Grammatikakis, I., Dudekula, D. B., De, S., Kim, J., Noh, J. H., Kim, K. M., Martindale, J. L., & Gorospe, M. (2017). Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by *CircPABPN1*. *RNA Biology*, *14*(3), 361–369. https://doi.org/10.1080/15476286.2017.1279788

Abe, T., Inokuchi, H., Yamada, Y., Muto, A., Iwasaki, Y., & Ikemura, T. (2014). tRNADB-CE: tRNA gene database well-timed in the era of big sequence data. *Frontiers in Genetics*, *5*. https://doi.org/10.3389/fgene.2014.00114

Akinyi, M. v., & Frilander, M. J. (2021). At the Intersection of Major and Minor Spliceosomes: Crosstalk Mechanisms and Their Impact on Gene Expression. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.700744

Akiyama, Y., Kharel, P., Abe, T., Anderson, P., & Ivanov, P. (2020). Isolation and initial structure-functional characterization of endogenous tRNA-derived stress-induced RNAs. *RNA Biology*, *17*(8), 1116–1124. https://doi.org/10.1080/15476286.2020.1732702

Alberti, C., & Cochella, L. (2017). A framework for understanding the roles of miRNAs in animal development. *Development*, *144*(14), 2548–2559. https://doi.org/10.1242/dev.146613

Alessio, E., Buson, L., Chemello, F., Peggion, C., Grespi, F., Martini, P., Massimino, M. L., Pacchioni, B., Millino, C., Romualdi, C., Bertoli, A., Scorrano, L., Lanfranchi, G., & Cagnin, S. (2019). Single cell analysis reveals the involvement of the long non-coding RNA Pvt1 in the modulation of muscle atrophy and mitochondrial network. *Nucleic Acids Research*, *47*(4), 1653–1670. https://doi.org/10.1093/nar/gkz007

Alexanian, M., Maric, D., Jenkinson, S. P., Mina, M., Friedman, C. E., Ting, C.-C., Micheletti, R., Plaisance, I., Nemir, M., Maison, D., Kernen, J., Pezzuto, I., Villeneuve, D., Burdet, F., Ibberson, M., Leib, S. L., Palpant, N. J., Hernandez, N., Ounzain, S., & Pedrazzini, T. (2017). A transcribed enhancer dictates mesendoderm specification in pluripotency. *Nature Communications*, *8*(1), 1806. https://doi.org/10.1038/s41467-017-01804-w

Alvarez, M. L., & DiStefano, J. K. (2011). Functional Characterization of the Plasmacytoma Variant Translocation 1 Gene (PVT1) in Diabetic Nephropathy. *PLoS ONE*, *6*(4), e18671. https://doi.org/10.1371/journal.pone.0018671

Amaral, P. P., Leonardi, T., Han, N., Viré, E., Gascoigne, D. K., Arias-Carrasco, R., Büscher, M., Pandolfini, L., Zhang, A., Pluchino, S., Maracaja-Coutinho, V., Nakaya, H. I., Hemberg, M., Shiekhattar, R., Enright, A. J., & Kouzarides, T. (2018). Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. *Genome Biology*, *19*(1), 32. https://doi.org/10.1186/s13059-018-1405-5

Ameres, S. L., Horwich, M. D., Hung, J.-H., Xu, J., Ghildiyal, M., Weng, Z., & Zamore, P. D. (2010). Target RNA–Directed Trimming and Tailing of Small Silencing RNAs. *Science*, *328*(5985), 1534–1539. https://doi.org/10.1126/science.1187058

Andergassen, D., & Rinn, J. L. (2022). From genotype to phenotype: genetics of mammalian long non-coding RNAs in vivo. *Nature Reviews Genetics*, *23*(4), 229–243. https://doi.org/10.1038/s41576-021-00427-8

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10). https://doi.org/10.1186/gb-2010-11-10-r106

Anderson, K. M., Anderson, D. M., McAnally, J. R., Shelton, J. M., Bassel-Duby, R., & Olson, E. N. (2016). Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature*, *539*(7629), 433–436. https://doi.org/10.1038/nature20128

Aravin, A. A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., & Tuschl, T. (2003). The Small RNA Profile during Drosophila melanogaster Development. *Developmental Cell*, *5*(2), 337–350. https://doi.org/10.1016/S1534-5807(03)00228-4

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J. J., Ju, J., Sheridan, R., Sander, C., Zavolan, M., & Tuschl, T. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, *442*(7099), 203–207. https://doi.org/10.1038/nature04916

Ashwal-Fluss, R., Meyer, M., Pamudurti, N. R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., & Kadener, S. (2014). circRNA Biogenesis Competes with Pre-mRNA Splicing. *Molecular Cell*, *56*(1), 55–66. https://doi.org/10.1016/j.molcel.2014.08.019

Avagyan, S., & Zon, L. I. (2016). Fish to Learn: Insights into Blood Development and Blood Disorders from Zebrafish Hematopoiesis. *Human Gene Therapy*, *27*(4), 287–294. https://doi.org/10.1089/hum.2016.024

Azam, S., Hou, S., Zhu, B., Wang, W., Hao, T., Bu, X., Khan, M., & Lei, H. (2019). Nuclear retention element recruits U1 snRNP components to restrain spliced lncRNAs in the nucleus. *RNA Biology*, *16*(8), 1001–1009. https://doi.org/10.1080/15476286.2019.1620061

Baker, B. S., Gorman, M., & Marín, I. (1994). DOSAGE COMPENSATION IN *DROSOPHILA*. *Annual Review of Genetics*, *28*(1), 491–521. https://doi.org/10.1146/annurev.ge.28.120194.002423

Balakin, A. G., Smith, L., & Fournier, M. J. (1996). The RNA World of the Nucleolus: Two Major Families of Small RNAs Defined by Different Box Elements with Related Functions. *Cell*, *86*(5), 823–834. https://doi.org/10.1016/S0092-8674(00)80156-7

Ballarino, M., Cipriano, A., Tita, R., Santini, T., Desideri, F., Morlando, M., Colantoni, A., Carrieri, C., Nicoletti, C., Musarò, A., Carroll, D. O., & Bozzoni, I. (2018). Deficiency in the nuclear long noncoding <scp>RNA</scp> *Charme* causes myogenic defects and heart remodeling in mice. *The EMBO Journal*, *37*(18). https://doi.org/10.15252/embj.201899697

Baltz, A. G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C., & Landthaler, M. (2012). The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Molecular Cell*, *46*(5), 674–690. https://doi.org/10.1016/j.molcel.2012.05.021

Banerjee, B., Koner, D., Karasik, D., & Saha, N. (2021). Genome-wide identification of novel long non-coding RNAs and their possible roles in hypoxic zebrafish brain. *Genomics*, *113*(1), 29–43. https://doi.org/10.1016/j.ygeno.2020.11.023

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. https://doi.org/10.1038/cr.2011.22

BARR, M. L., & BERTRAM, E. G. (1949). A Morphological Distinction between Neurones of the Male and Female, and the Behaviour of the Nucleolar Satellite during Accelerated Nucleoprotein Synthesis. *Nature*, *163*(4148), 676–677. https://doi.org/10.1038/163676a0

Barsotti, A. M., Beckerman, R., Laptenko, O., Huppi, K., Caplen, N. J., & Prives, C. (2012). p53-dependent Induction of PVT1 and miR-1204. *Journal of Biological Chemistry*, *287*(4), 2509–2519. https://doi.org/10.1074/jbc.M111.322875

Bartel, D. P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, *136*(2), 215–233. https://doi.org/10.1016/j.cell.2009.01.002

Bartel, D. P. (2018). Metazoan MicroRNAs. *Cell*, *173*(1), 20–51. https://doi.org/10.1016/j.cell.2018.03.006

Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., Vejnar, C. E., Lee, M. T., Rajewsky, N., Walther, T. C., & Giraldez, A. J. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *The EMBO Journal*, *33*(9), 981–993. https://doi.org/10.1002/embj.201488411

Bensaad, K., Tsuruta, A., Selak, M. A., Vidal, M. N. C., Nakano, K., Bartrons, R., Gottlieb, E., & Vousden, K. H. (2006). TIGAR, a p53-Inducible Regulator of Glycolysis and Apoptosis. *Cell*, *126*(1), 107–120. https://doi.org/10.1016/j.cell.2006.05.036

Berger, J., & Currie, P. D. (2012). Zebrafish models flex their muscles to shed light on muscular dystrophies. *Disease Models & Mechanisms*, *5*(6), 726–732. https://doi.org/10.1242/dmm.010082

Bhan, A., Soleimani, M., & Mandal, S. S. (2017). Long Noncoding RNA and Cancer: A New Paradigm. *Cancer Research*, *77*(15), 3965–3981. https://doi.org/10.1158/0008-5472.CAN-16-2634

Bi, Y., Ji, J., & Zhou, Y. (2021). LncRNA-PVT1 indicates a poor prognosis and promotes angiogenesis via activating the HNF1B/EMT axis in glioma. *Journal of Cancer*, *12*(19), 5732–5744. https://doi.org/10.7150/jca.60257

Bitetti, A., Mallory, A. C., Golini, E., Carrieri, C., Carreño Gutiérrez, H., Perlas, E., Pérez-Rico, Y. A., Tocchini-Valentini, G. P., Enright, A. J., Norton, W. H. J., Mandillo, S., O'Carroll, D., & Shkumatava, A. (2018). MicroRNA degradation by a conserved target RNA regulates animal behavior. *Nature Structural & Molecular Biology*, *25*(3), 244–251. https://doi.org/10.1038/s41594-018-0032-x

Bonasio, R., Tu, S., & Reinberg, D. (2010). Molecular Signals of Epigenetic States. *Science*, *330*(6004), 612–616. https://doi.org/10.1126/science.1191078

Bond, A. M., VanGompel, M. J. W., Sametsky, E. A., Clark, M. F., Savage, J. C., Disterhoft, J. F., & Kohtz, J. D. (2009). Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nature Neuroscience*, *12*(8), 1020–1027. https://doi.org/10.1038/nn.2371

Botthof, J. G., Bielczyk-Maczyńska, E., Ferreira, L., & Cvejic, A. (2017). Loss of the homologous recombination gene *rad51* leads to Fanconi anemia-like symptoms in zebrafish. *Proceedings of the National Academy of Sciences*, *114*(22). https://doi.org/10.1073/pnas.1620631114

Bowen, M. E., & Attardi, L. D. (2019). The role of p53 in developmental syndromes. *Journal of Molecular Cell Biology*, *11*(3), 200–211. https://doi.org/10.1093/jmcb/mjy087

Bradford, Y. M., Toro, S., Ramachandran, S., Ruzicka, L., Howe, D. G., Eagle, A., Kalita, P., Martin, R., Taylor Moxon, S. A., Schaper, K., & Westerfield, M. (2017). Zebrafish Models of Human Disease: Gaining Insight into Human Disease at ZFIN. *ILAR Journal*, *58*(1), 4–16. https://doi.org/10.1093/ilar/ilw040

Brannan, C. I., Dees, E. C., Ingram, R. S., & Tilghman, S. M. (1990). The product of the H19 gene may function as an RNA. *Molecular and Cellular Biology*, *10*(1), 28–36. https://doi.org/10.1128/mcb.10.1.28-36.1990

Bresciani, E., Broadbridge, E., & Liu, P. P. (2018). An efficient dissociation protocol for generation of single cell suspension from zebrafish embryos and larvae. *MethodsX*, *5*, 1287–1290. https://doi.org/10.1016/j.mex.2018.10.009

Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., Norris, D. P., Penny, G. D., Patel, D., & Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature*, *351*(6324), 329–331. https://doi.org/10.1038/351329a0

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S., & Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*, *71*(3), 515–526. https://doi.org/10.1016/0092-8674(92)90519-I

Broughton, J. P., Lovci, M. T., Huang, J. L., Yeo, G. W., & Pasquinelli, A. E. (2016). Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Molecular Cell*, *64*(2), 320–333. https://doi.org/10.1016/j.molcel.2016.09.004

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, R. G., Xing, Y., Lawrence, J., & Willard, H. F. (1992). The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, *71*(3), 527–542. https://doi.org/10.1016/0092-8674(92)90520-M

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*, *25*(18), 1915–1927. https://doi.org/10.1101/gad.17446611

Calin, G. A., & Croce, C. M. (2006). MicroRNA signatures in human cancers. *Nature Reviews Cancer*, *6*(11), 857–866. https://doi.org/10.1038/nrc1997

Candeias, M. M., Malbert-Colas, L., Powell, D. J., Daskalogianni, C., Maslon, M. M., Naski, N., Bourougaa, K., Calvo, F., & Fåhraeus, R. (2008). p53 mRNA controls p53 activity by managing Mdm2 functions. *Nature Cell Biology*, *10*(9), 1098–1105. https://doi.org/10.1038/ncb1770

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J., Trapnell, C., & Shendure, J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, *566*(7745), 496–502. https://doi.org/10.1038/s41586-019-0969-x

Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P., & Lovell-Badge, R. (1993). Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell*, *73*(5), 1019–1030. https://doi.org/10.1016/0092-8674(93)90279-Y

Carlevaro-Fita, J., Lanzós, A., Feuerbach, L., Hong, C., Mas-Ponte, D., Pedersen, J. S., & Johnson, R. (2020). Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Communications Biology*, *3*(1), 56. https://doi.org/10.1038/s42003-019-0741-7

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V. B., Brenner, S. E., Batalov, S., Forrest, A. R. R., Zavolan, M., Davis, M. J., Wilming, L. G., … Hayashizaki, Y. (2005). The Transcriptional Landscape of the Mammalian Genome. *Science*, *309*(5740), 1559–1563. https://doi.org/10.1126/science.1112014

Carradice, D., & Lieschke, G. J. (2008). Zebrafish in hematology: sushi or science? *Blood*, *111*(7), 3331–3342. https://doi.org/10.1182/blood-2007-10-052761

Carramusa, L., Contino, F., Ferro, A., Minafra, L., Perconti, G., Giallongo, A., & Feo, S. (2007). The PVT-1 oncogene is a Myc protein target that is overexpressed in transformed cells. *Journal of Cellular Physiology*, *213*(2), 511–518. https://doi.org/10.1002/jcp.21133

Cassar, S., Adatto, I., Freeman, J. L., Gamse, J. T., Iturria, I., Lawrence, C., Muriana, A., Peterson, R. T., van Cruchten, S., & Zon, L. I. (2020). Use of Zebrafish in Drug Discovery Toxicology. *Chemical Research in Toxicology*, *33*(1), 95–118. https://doi.org/10.1021/acs.chemrestox.9b00335

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsveld, J., & Hentze, M. W.

(2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell*, *149*(6), 1393–1406. https://doi.org/10.1016/j.cell.2012.04.031

Ceccaldi, R., Parmar, K., Mouly, E., Delord, M., Kim, J. M., Regairaz, M., Pla, M., Vasquez, N., Zhang, Q.-S., Pondarre, C., Peffault de Latour, R., Gluckman, E., Cavazzana-Calvo, M., Leblanc, T., Larghero, J., Grompe, M., Socié, G., D'Andrea, A. D., & Soulier, J. (2012). Bone Marrow Failure in Fanconi Anemia Is Triggered by an Exacerbated p53/p21 DNA Damage Response that Impairs Hematopoietic Stem and Progenitor Cells. *Cell Stem Cell*, *11*(1), 36–49. https://doi.org/10.1016/j.stem.2012.05.013

Chakraborty, A., Uechi, T., Higa, S., Torihara, H., & Kenmochi, N. (2009). Loss of Ribosomal Protein L11 Affects Zebrafish Embryonic Development through a p53-Dependent Apoptotic Response. *PLoS ONE*, *4*(1), e4152. https://doi.org/10.1371/journal.pone.0004152

Chandradoss, S. D., Schirle, N. T., Szczepaniak, M., MacRae, I. J., & Joo, C. (2015). A Dynamic Search Process Underlies MicroRNA Targeting. *Cell*, *162*(1), 96–107. https://doi.org/10.1016/j.cell.2015.06.032

Chappell, J., & Dalton, S. (2013). Roles for MYC in the Establishment and Maintenance of Pluripotency. *Cold Spring Harbor Perspectives in Medicine*, *3*(12), a014381–a014381. https://doi.org/10.1101/cshperspect.a014381

Chen, L., Lü, M.-H., Zhang, D., Hao, N.-B., Fan, Y.-H., Wu, Y.-Y., Wang, S.-M., Xie, R., Fang, D.-C., Zhang, H., Hu, C.-J., & Yang, S.-M. (2014). miR-1207-5p and miR-1266 suppress gastric cancer growth and invasion by targeting telomerase reverse transcriptase. *Cell Death & Disease*, *5*(1), e1034–e1034. https://doi.org/10.1038/cddis.2013.553

Chen, L.-L. (2016). Linking Long Noncoding RNA Localization and Function. *Trends in Biochemical Sciences*, *41*(9), 761–772. https://doi.org/10.1016/j.tibs.2016.07.003

Chen, N., Zhao, G., Yan, X., Lv, Z., Yin, H., Zhang, S., Song, W., Li, X., Li, L., Du, Z., Jia, L., Zhou, L., Li, W., Hoffman, A. R., Hu, J.-F., & Cui, J. (2018). A novel FLI1 exonic circular RNA promotes metastasis in breast cancer by coordinately regulating TET1 and DNMT1. *Genome Biology*, *19*(1), 218. https://doi.org/10.1186/s13059-018-1594-y

Chen, Q., Yan, M., Cao, Z., Li, X., Zhang, Y., Shi, J., Feng, G., Peng, H., Zhang, X., Zhang, Y., Qian, J., Duan, E., Zhai, Q., & Zhou, Q. (2016). Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science*, *351*(6271), 397–400. https://doi.org/10.1126/science.aad7977

Chen, Y., Mao, Z.-D., Shi, Y.-J., Qian, Y., Liu, Z.-G., Yin, X.-W., & Zhang, Q. (2019). Comprehensive analysis of miRNA–mRNA–lncRNA networks in severe asthma. *Epigenomics*, *11*(2), 115–131. https://doi.org/10.2217/epi-2018-0132

Chew, G.-L., Pauli, A., Rinn, J. L., Regev, A., Schier, A. F., & Valen, E. (2013). Ribosome profiling reveals resemblance between long non-coding RNAs and 5′ leaders of coding RNAs. *Development*, *140*(13), 2828–2834. https://doi.org/10.1242/dev.098343

Cho, S. W., Xu, J., Sun, R., Mumbach, M. R., Carter, A. C., Chen, Y. G., Yost, K. E., Kim, J., He, J., Nevins, S. A., Chin, S.-F., Caldas, C., Liu, S. J., Horlbeck, M. A., Lim, D. A., Weissman, J. S., Curtis, C., & Chang, H. Y. (2018). Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell*, *173*(6), 1398-1412.e22. https://doi.org/10.1016/j.cell.2018.03.068

Chodroff, R. A., Goodstadt, L., Sirey, T. M., Oliver, P. L., Davies, K. E., Green, E. D., Molnár, Z., & Ponting, C. P. (2010). Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biology*, *11*(7), R72. https://doi.org/10.1186/gb-2010-11-7-r72

Choi, J., & Donehower, L. A. (1999). p53 in embryonic development: maintaining a fine balance. In *CMLS, Cell. Mol. Life Sci* (Vol. 55).
Choi, T.-Y., Choi, T.-I., Lee, Y.-R., Choe, S.-K., & Kim, C.-H. (2021). Zebrafish as an animal model for biomedical research. *Experimental & Molecular Medicine*, *53*(3), 310–317. https://doi.org/10.1038/s12276-021-00571-5
Chooniedass-Kothari, S., Emberley, E., Hamedani, M. K., Troup, S., Wang, X., Czosnek, A., Hube, F., Mutawe, M., Watson, P. H., & Leygue, E. (2004). The steroid receptor RNA activator is the first functional RNA encoding a protein. *FEBS Letters*, *566*(1–3), 43–47. https://doi.org/10.1016/j.febslet.2004.03.104
Ciganda, M., & Williams, N. (2011). Eukaryotic 5S rRNA biogenesis. *Wiley Interdisciplinary Reviews: RNA*, *2*(4), 523–533. https://doi.org/10.1002/wrna.74
Claude, A. (1939). Chemical Composition of the Tumor-Producing Fraction of Chicken Tumor I. *Science*, *90*(2331), 213–214. https://doi.org/10.1126/science.90.2331.213
Claycomb, J. M. (2014). Ancient Endo-siRNA Pathways Reveal New Tricks. *Current Biology*, *24*(15), R703–R715. https://doi.org/10.1016/j.cub.2014.06.009

Cobb, M. (2017). 60 years ago, Francis Crick changed the logic of biology. *PLOS Biology*, *15*(9). https://doi.org/10.1371/journal.pbio.2003243
Coleman, J., Green, P. J., & Inouye, M. (1984). The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes. *Cell*, *37*(2), 429–436. https://doi.org/10.1016/0092-8674(84)90373-8
Constanty, F., & Shkumatava, A. (2021). lncRNAs in development and differentiation: from sequence motifs to functional characterization. *Development*, *148*(1). https://doi.org/10.1242/dev.182741
Conte, F., Fiscon, G., Chiara, M., Colombo, T., Farina, L., & Paci, P. (2017). Role of the long non-coding RNA PVT1 in the dysregulation of the ceRNA-ceRNA network in human breast cancer. *PLOS ONE*, *12*(2), e0171661. https://doi.org/10.1371/journal.pone.0171661
Cordaux, R., & Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics*, *10*(10), 691–703. https://doi.org/10.1038/nrg2640
Cory, S., Graham, M., Webb, E., Corcoran, L., & Adams, J. M. (1985). Variant (6;15) translocations in murine plasmacytomas involve a chromosome 15 locus at least 72 kb from the c-myc oncogene. *The EMBO Journal*, *4*(3), 675–681. https://doi.org/10.1002/j.1460-2075.1985.tb03682.x
Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, *227*(5258), 561–563. https://doi.org/10.1038/227561a0
Cui, Y., Hagan, K. W., Zhang, S., & Peltz, S. W. (1995). Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. *Genes & Development*, *9*(4), 423–436. https://doi.org/10.1101/gad.9.4.423
Czech, B., & Hannon, G. J. (2016). One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends in Biochemical Sciences*, *41*(4), 324–337. https://doi.org/10.1016/j.tibs.2015.12.008
Dang, C. V. (2012). MYC on the Path to Cancer. *Cell*, *149*(1), 22–35. https://doi.org/10.1016/j.cell.2012.03.003
Danilova, N., Kumagai, A., & Lin, J. (2010). p53 Upregulation Is a Frequent Response to Deficiency of Cell-Essential Genes. *PLoS ONE*, *5*(12), e15938. https://doi.org/10.1371/journal.pone.0015938

Davis, M., Malcolm, S., & Rabbitts, T. H. (1984). Chromosome translocation can occur on either side of the c-myc oncogene in Burkitt lymphoma cells. *Nature*, *308*(5956), 286–288. https://doi.org/10.1038/308286a0

de Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., Muller, H., Ragoussis, J., Wei, C.-L., & Natoli, G. (2010). A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. *PLoS Biology*, *8*(5), e1000384. https://doi.org/10.1371/journal.pbio.1000384

Derderian, C., Orunmuyi, A. T., Olapade-Olaopa, E. O., & Ogunwobi, O. O. (2019). PVT1 Signaling Is a Mediator of Cancer Progression. *Frontiers in Oncology*, *9*. https://doi.org/10.3389/fonc.2019.00502

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., … Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–1789. https://doi.org/10.1101/gr.132159.111

Diederichs, S. (2014). The four dimensions of noncoding RNA conservation. *Trends in Genetics*, *30*(4), 121–123. https://doi.org/10.1016/j.tig.2014.01.004

Dinger, M. E., Pang, K. C., Mercer, T. R., & Mattick, J. S. (2008). Differentiating Protein-Coding and Noncoding RNA: Challenges and Ambiguities. *PLoS Computational Biology*, *4*(11), e1000176. https://doi.org/10.1371/journal.pcbi.1000176

Dittmar, K. A., Goodenbour, J. M., & Pan, T. (2006). Tissue-Specific Differences in Human Transfer RNA Expression. *PLoS Genetics*, *2*(12), e221. https://doi.org/10.1371/journal.pgen.0020221

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R. F., … Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, *489*(7414), 101–108. https://doi.org/10.1038/nature11233

Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N. J., van Nostrand, E. L., Pratt, G. A., Yeo, G. W., Graveley, B. R., & Burge, C. B. (2018). Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Molecular Cell*, *70*(5), 854-867.e9. https://doi.org/10.1016/j.molcel.2018.05.001

Du, W. W., Fang, L., Yang, W., Wu, N., Awan, F. M., Yang, Z., & Yang, B. B. (2017). Induction of tumor apoptosis through a circular RNA enhancing Foxo3 activity. *Cell Death & Differentiation*, *24*(2), 357–370. https://doi.org/10.1038/cdd.2016.133

Du, W. W., Yang, W., Liu, E., Yang, Z., Dhaliwal, P., & Yang, B. B. (2016). Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Research*, *44*(6), 2846–2858. https://doi.org/10.1093/nar/gkw027

Dudekula, D. B., Panda, A. C., Grammatikakis, I., De, S., Abdelmohsen, K., & Gorospe, M. (2016). CircInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biology*, *13*(1), 34–42. https://doi.org/10.1080/15476286.2015.1128065

Duret, L., Chureau, C., Samain, S., Weissenbach, J., & Avner, P. (2006). The *Xist* RNA Gene Evolved in Eutherians by Pseudogenization of a Protein-Coding Gene. *Science*, *312*(5780), 1653–1655. https://doi.org/10.1126/science.1126316

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., & Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, *411*(6836), 494–498. https://doi.org/10.1038/35078107

Elguindy, M. M., & Mendell, J. T. (2021). NORAD-induced Pumilio phase separation is required for genome stability. *Nature*, *595*(7866), 303–308. https://doi.org/10.1038/s41586-021-03633-w

Elkayam, E., Kuhn, C.-D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J., & Joshua-Tor, L. (2012). The Structure of Human Argonaute-2 in Complex with miR-20a. *Cell*, *150*(1), 100–110. https://doi.org/10.1016/j.cell.2012.05.017

Engström, P. G., Suzuki, H., Ninomiya, N., Akalin, A., Sessa, L., Lavorgna, G., Brozzi, A., Luzi, L., Tan, S. L., Yang, L., Kunarso, G., Ng, E. L.-C., Batalov, S., Wahlestedt, C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Wells, C., … Lipovich, L. (2006). Complex Loci in Human and Mouse Genomes. *PLoS Genetics*, *2*(4), e47. https://doi.org/10.1371/journal.pgen.0020047

Esposito, R., Bosch, N., Lanzós, A., Polidori, T., Pulido-Quetglas, C., & Johnson, R. (2019). Hacking the Cancer Genome: Profiling Therapeutically Actionable Long Non-coding RNAs Using CRISPR-Cas9 Screening. *Cancer Cell*, *35*(4), 545–557. https://doi.org/10.1016/j.ccell.2019.01.019

Esteller, M. (2011). Non-coding RNAs in human disease. *Nature Reviews Genetics*, *12*(12), 861–874. https://doi.org/10.1038/nrg3074

Fabian, M. R., & Sonenberg, N. (2012). The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nature Structural & Molecular Biology*, *19*(6), 586–593. https://doi.org/10.1038/nsmb.2296

Franke, A., & Baker, B. S. (1999). The rox1 and rox2 RNAs Are Essential Components of the Compensasome, which Mediates Dosage Compensation in Drosophila. *Molecular Cell*, *4*(1), 117–122. https://doi.org/10.1016/S1097-2765(00)80193-8

Franklin, D. A., He, Y., Leslie, P. L., Tikunov, A. P., Fenger, N., Macdonald, J. M., & Zhang, Y. (2016). p53 coordinates DNA repair with nucleotide synthesis by suppressing PFKFB3 expression and promoting the pentose phosphate pathway. *Scientific Reports*, *6*(1), 38067. https://doi.org/10.1038/srep38067

Fu, J., Shi, H., Wang, B., Zhan, T., Shao, Y., Ye, L., Wu, S., Yu, C., & Zheng, L. (2020). LncRNA PVT1 links Myc to glycolytic metabolism upon CD4+ T cell activation and Sjögren's syndrome-like autoimmune response. *Journal of Autoimmunity*, *107*, 102358. https://doi.org/10.1016/j.jaut.2019.102358

Fuda, N. J., Ardehali, M. B., & Lis, J. T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*, *461*(7261), 186–192. https://doi.org/10.1038/nature08449

Furlan, G., Gutierrez Hernandez, N., Huret, C., Galupa, R., van Bemmel, J. G., Romito, A., Heard, E., Morey, C., & Rougeulle, C. (2018). The Ftx Noncoding Locus Controls X Chromosome Inactivation Independently of Its RNA Products. *Molecular Cell*, *70*(3), 462-472.e8. https://doi.org/10.1016/j.molcel.2018.03.024

Gagnon, K. T., & Corey, D. R. (2012). Argonaute and the Nuclear RNAs: New Pathways for RNA-Mediated Control of Gene Expression. *Nucleic Acid Therapeutics*, *22*(1), 3–16. https://doi.org/10.1089/nat.2011.0330

Ganot, P., Bortolin, M.-L., & Kiss, T. (1997). Site-Specific Pseudouridine Formation in Preribosomal RNA Is Guided by Small Nucleolar RNAs. *Cell*, *89*(5), 799–809. https://doi.org/10.1016/S0092-8674(00)80263-9

Gao, F., Cai, Y., Kapranov, P., & Xu, D. (2020). Reverse-genetics studies of lncRNAs—what we have learnt and paths forward. *Genome Biology*, *21*(1), 93. https://doi.org/10.1186/s13059-020-01994-5

Gebetsberger, J., Wyss, L., Mleczko, A. M., Reuther, J., & Polacek, N. (2017). A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation

during stress. *RNA Biology*, *14*(10), 1364–1373. https://doi.org/10.1080/15476286.2016.1257470
Ghetti, M., Vannini, I., Storlazzi, C. T., Martinelli, G., & Simonetti, G. (2020). Linear and circular PVT1 in hematological malignancies and immune response: two faces of the same coin. *Molecular Cancer*, *19*(1), 69. https://doi.org/10.1186/s12943-020-01187-5
Gil, N., & Ulitsky, I. (2020). Regulation of gene expression by cis-acting long non-coding RNAs. *Nature Reviews Genetics*, *21*(2), 102–117. https://doi.org/10.1038/s41576-019-0184-5
Gillinder, K. R., Tuckey, H., Bell, C. C., Magor, G. W., Huang, S., Ilsley, M. D., & Perkins, A. C. (2017). Direct targets of pSTAT5 signalling in erythropoiesis. *PLOS ONE*, *12*(7), e0180922. https://doi.org/10.1371/journal.pone.0180922
Girard, A., Sachidanandam, R., Hannon, G. J., & Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, *442*(7099), 199–202. https://doi.org/10.1038/nature04917
Golomb, L., Volarevic, S., & Oren, M. (2014). p53 and ribosome biogenesis stress: The essentials. *FEBS Letters*, *588*(16), 2571–2579. https://doi.org/10.1016/j.febslet.2014.04.014
Gotea, V., Petrykowska, H. M., & Elnitski, L. (2013). Bidirectional Promoters as Important Drivers for the Emergence of Species-Specific Transcripts. *PLoS ONE*, *8*(2), e57323. https://doi.org/10.1371/journal.pone.0057323
Goudarzi, M., Berg, K., Pieper, L. M., & Schier, A. F. (2019). Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility. *ELife*, *8*. https://doi.org/10.7554/eLife.40815
Graf, J., & Kretz, M. (2020). From structure to function: Route to understanding lncRNA mechanism. *BioEssays*, *42*(12), 2000027. https://doi.org/10.1002/bies.202000027
Graham, M., & Adams, J. M. (1986). Chromosome 8 breakpoint far 3′ of the c-myc oncogene in a Burkitt's lymphoma 2;8 variant translocation is equivalent to the murine pvt-1 locus. *The EMBO Journal*, *5*(11), 2845–2851. https://doi.org/10.1002/j.1460-2075.1986.tb04578.x
Graindorge, A., Pinheiro, I., Nawrocka, A., Mallory, A. C., Tsvetkov, P., Gil, N., Carolis, C., Buchholz, F., Ulitsky, I., Heard, E., Taipale, M., & Shkumatava, A. (2019). In-cell identification and measurement of RNA-protein interactions. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-13235-w
Grandi, P., Rybin, V., Baßler, J., Petfalski, E., Strauß, D., Marzioch, M., Schäfer, T., Kuster, B., Tschochner, H., Tollervey, D., Gavin, A.-C., & Hurt, E. (2002). 90S Pre-Ribosomes Include the 35S Pre-rRNA, the U3 snoRNP, and 40S Subunit Processing Factors but Predominantly Lack 60S Synthesis Factors. *Molecular Cell*, *10*(1), 105–115. https://doi.org/10.1016/S1097-2765(02)00579-8
Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics*, *27*(7), 1017–1018. https://doi.org/10.1093/bioinformatics/btr064
Grantham, R., Gautier, C., Gouy, M., Mercier, R., & Pavé, A. (1980). Codon catalog usage and the genome hypothesis. *Nucleic Acids Research*, *8*(1), 197–197. https://doi.org/10.1093/nar/8.1.197-c
Gräwe, C., Stelloo, S., van Hout, F. A. H., & Vermeulen, M. (2021). RNA-Centric Methods: Toward the Interactome of Specific RNA Transcripts. *Trends in Biotechnology*, *39*(9), 890–900. https://doi.org/10.1016/j.tibtech.2020.11.011
Grivna, S. T., Beyret, E., Wang, Z., & Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes & Development*, *20*(13), 1709–1714. https://doi.org/10.1101/gad.1434406

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., & Herrmann, B. G. (2013). The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. *Developmental Cell*, *24*(2), 206–214. https://doi.org/10.1016/j.devcel.2012.12.012

Guan, Y., Kuo, W.-L., Stilwell, J. L., Takano, H., Lapuk, A. v., Fridlyand, J., Mao, J.-H., Yu, M., Miller, M. A., Santos, J. L., Kalloger, S. E., Carlson, J. W., Ginzinger, D. G., Celniker, S. E., Mills, G. B., Huntsman, D. G., & Gray, J. W. (2007). Amplification of *PVT1* Contributes to the Pathophysiology of Ovarian and Breast Cancer. *Clinical Cancer Research*, *13*(19), 5745–5755. https://doi.org/10.1158/1078-0432.CCR-06-2882

Guella, I., Pistocchi, A., Asselta, R., Rimoldi, V., Ghilardi, A., Sironi, F., Trotta, L., Primignani, P., Zini, M., Zecchinelli, A., Coviello, D., Pezzoli, G., del Giacco, L., Duga, S., & Goldwurm, S. (2011). Mutational screening and zebrafish functional analysis of GIGYF2 as a Parkinson-disease gene. *Neurobiology of Aging*, *32*(11), 1994–2005. https://doi.org/10.1016/j.neurobiolaging.2009.12.016

Guo, C., Li, J., Guo, M., Bai, R., Lei, G., Sun, H., Tong, S., He, K., & He, L. (2022). Aberrant expressions of MIAT and PVT1 in serum exosomes of schizophrenia patients. *Schizophrenia Research*, *240*, 71–72. https://doi.org/10.1016/j.schres.2021.12.013

Guo, C.-J., Ma, X.-K., Xing, Y.-H., Zheng, C.-C., Xu, Y.-F., Shan, L., Zhang, J., Wang, S., Wang, Y., Carmichael, G. G., Yang, L., & Chen, L.-L. (2020). Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell*, *181*(3), 621-636.e22. https://doi.org/10.1016/j.cell.2020.03.006

Guo, D., Wang, Y., Ren, K., & Han, X. (2018). Knockdown of LncRNA PVT1 inhibits tumorigenesis in non-small-cell lung cancer by regulating miR-497 expression. *Experimental Cell Research*, *362*(1), 172–179. https://doi.org/10.1016/j.yexcr.2017.11.014

Guo, J., Hao, C., Wang, C., & Li, L. (2018). Long noncoding RNA PVT1 modulates hepatocellular carcinoma cell proliferation and apoptosis by recruiting EZH2. *Cancer Cell International*, *18*(1), 98. https://doi.org/10.1186/s12935-018-0582-3

Guo, Y., Niu, C., Breslin, P., Tang, M., Zhang, S., Wei, W., Kini, A. R., Paner, G. P., Alkan, S., Morris, S. W., Diaz, M., Stiff, P. J., & Zhang, J. (2009). c-Myc–mediated control of cell fate in megakaryocyte-erythrocyte progenitors. *Blood*, *114*(10), 2097–2106. https://doi.org/10.1182/blood-2009-01-197947

Gutschner, T., & Diederichs, S. (2012). The hallmarks of cancer. *RNA Biology*, *9*(6), 703–719. https://doi.org/10.4161/rna.20481

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., Huarte, M., Zuk, O., Carey, B. W., Cassady, J. P., Cabili, M. N., Jaenisch, R., Mikkelsen, T. S., Jacks, T., Hacohen, N., Bernstein, B. E., Kellis, M., Regev, A., Rinn, J. L., & Lander, E. S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, *458*(7235), 223–227. https://doi.org/10.1038/nature07672

Guttman, M., & Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature*, *482*(7385), 339–346. https://doi.org/10.1038/nature10887

Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., & Lander, E. S. (2013). Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. *Cell*, *154*(1), 240–251. https://doi.org/10.1016/j.cell.2013.06.009

Hadjiolov, A. A., Venkov, P. V., & Tsanev, R. G. (1966). Ribonucleic acids fractionation by density-gradient centrifugation and by agar gel electrophoresis: A comparison. *Analytical Biochemistry*, *17*(2), 263–267. https://doi.org/10.1016/0003-2697(66)90204-1

Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., Ule, J., & Zavolan, M. (2021). CLIP and complementary methods. *Nature Reviews Methods Primers*, *1*(1), 20. https://doi.org/10.1038/s43586-021-00018-1

Haiser, H. J., Karginov, F. v., Hannon, G. J., & Elliot, M. A. (2008). Developmentally regulated cleavage of tRNAs in the bacterium Streptomyces coelicolor. *Nucleic Acids Research*, *36*(3), 732–741. https://doi.org/10.1093/nar/gkm1096

Hamilton, A. J., & Baulcombe, D. C. (1999). A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. *Science*, *286*(5441), 950–952. https://doi.org/10.1126/science.286.5441.950

Han, J., LaVigne, C. A., Jones, B. T., Zhang, H., Gillett, F., & Mendell, J. T. (2020). A ubiquitin ligase mediates target-directed microRNA decay independently of tailing and trimming. *Science*, *370*(6523). https://doi.org/10.1126/science.abc9546

Han, X., Zhang, J., Liu, Y., Fan, X., Ai, S., Luo, Y., Li, X., Jin, H., Luo, S., Zheng, H., Yue, Y., Chang, Z., Yang, Z., Tang, F., He, A., & Shen, X. (2019). The lncRNA *Hand2os1* / *Uph* locus orchestrates heart development through regulation of precise expression of *Hand2*. *Development*, *146*(13). https://doi.org/10.1242/dev.176198

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., & Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature*, *495*(7441), 384–388. https://doi.org/10.1038/nature11993

Hanson, R. L., Craig, D. W., Millis, M. P., Yeatts, K. A., Kobes, S., Pearson, J. v., Lee, A. M., Knowler, W. C., Nelson, R. G., & Wolford, J. K. (2007). Identification of *PVT1* as a Candidate Gene for End-Stage Renal Disease in Type 2 Diabetes Using a Pooling-Based Genome-Wide Single Nucleotide Polymorphism Association Study. *Diabetes*, *56*(4), 975–983. https://doi.org/10.2337/db06-1072

Harris, S. L., & Levine, A. J. (2005). The p53 pathway: positive and negative feedback loops. *Oncogene*, *24*(17), 2899–2908. https://doi.org/10.1038/sj.onc.1208615

He, F., Brown, A. H., & Jacobson, A. (1997). Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. *Molecular and Cellular Biology*, *17*(3), 1580–1594. https://doi.org/10.1128/MCB.17.3.1580

He, F., Song, Z., Chen, H., Chen, Z., Yang, P., Li, W., Yang, Z., Zhang, T., Wang, F., Wei, J., Wei, F., Wang, Q., & Cao, J. (2019). Long noncoding RNA PVT1-214 promotes proliferation and invasion of colorectal cancer by stabilizing Lin28 and interacting with miR-128. *Oncogene*, *38*(2), 164–179. https://doi.org/10.1038/s41388-018-0432-8

He, Y., Jing, Y., Wei, F., Tang, Y., Yang, L., Luo, J., Yang, P., Ni, Q., Pang, J., Liao, Q., Xiong, F., Guo, C., Xiang, B., Li, X., Zhou, M., Li, Y., Xiong, W., Zeng, Z., & Li, G. (2018). Long non-coding RNA PVT1 predicts poor prognosis and induces radioresistance by regulating DNA repair and cell apoptosis in nasopharyngeal carcinoma. *Cell Death & Disease*, *9*(2), 235. https://doi.org/10.1038/s41419-018-0265-y

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Henras, A. K., Plisson-Chastang, C., O'Donohue, M.-F., Chakraborty, A., & Gleizes, P.-E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdisciplinary Reviews: RNA*, *6*(2), 225–242. https://doi.org/10.1002/wrna.1269

Herriges, M. J., Swarr, D. T., Morley, M. P., Rathi, K. S., Peng, T., Stewart, K. M., & Morrisey, E. E. (2014). Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development. *Genes & Development*, *28*(12), 1363–1379. https://doi.org/10.1101/gad.238782.114

Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Ulitsky, I. (2015). Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of

Transcriptomes in 17 Species. *Cell Reports*, *11*(7), 1110–1122. https://doi.org/10.1016/j.celrep.2015.04.023
Ho, L., & Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature*, *463*(7280), 474–484. https://doi.org/10.1038/nature08911
Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I., & Zamecnik, P. C. (1958). A SOLUBLE RIBONUCLEIC ACID INTERMEDIATE IN PROTEIN SYNTHESIS. *Journal of Biological Chemistry*, *231*(1), 241–257. https://doi.org/10.1016/S0021-9258(19)77302-5
Holmes, Z. E., Hamilton, D. J., Hwang, T., Parsonnet, N. v., Rinn, J. L., Wuttke, D. S., & Batey, R. T. (2020). The Sox2 transcription factor binds RNA. *Nature Communications*, *11*(1), 1805. https://doi.org/10.1038/s41467-020-15571-8
Hon, C.-C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J. L., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T. M., Severin, J., Lizio, M., Kawaji, H., Kasukawa, T., Itoh, M., Burroughs, A. M., Noma, S., Djebali, S., Alam, T., Medvedeva, Y. A., … Forrest, A. R. R. (2017). An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature*, *543*(7644), 199–204. https://doi.org/10.1038/nature21374
Housman, G., & Ulitsky, I. (2016). Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1859*(1), 31–40. https://doi.org/10.1016/j.bbagrm.2015.07.017
Hsieh, L.-C., Lin, S.-I., Kuo, H.-F., & Chiou, T.-J. (2010). Abundance of tRNA-derived small RNAs in phosphate-starved Arabidopsis roots. *Plant Signaling & Behavior*, *5*(5), 537–539. https://doi.org/10.4161/psb.11029
Hsu, M. T., & Coca-Prados, M. (1979). Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature*, *280*(5720), 339–340. https://doi.org/10.1038/280339a0
Hu, L., Xu, Z., Hu, B., & Lu, Z. J. (2017). COME: a robust coding potential calculation tool for lncRNA identification and characterization based on multiple features. *Nucleic Acids Research*, *45*(1), e2–e2. https://doi.org/10.1093/nar/gkw798
Huang, T., Liu, H. W., Chen, J. Q., Wang, S. H., Hao, L. Q., Liu, M., & Wang, B. (2017). The long noncoding RNA PVT1 functions as a competing endogenous RNA by sponging miR-186 in gastric cancer. *Biomedicine & Pharmacotherapy*, *88*, 302–308. https://doi.org/10.1016/j.biopha.2017.01.049
Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nature Medicine*, *21*(11), 1253–1261. https://doi.org/10.1038/nm.3981
Huntzinger, E., & Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews Genetics*, *12*(2), 99–110. https://doi.org/10.1038/nrg2936
Huppi, K., Pitt, J. J., Wahlberg, B. M., & Caplen, N. J. (2012). The 8q24 Gene Desert: An Oasis of Non-Coding Transcriptional Activity. *Frontiers in Genetics*, *3*. https://doi.org/10.3389/fgene.2012.00069
Huppi, K., Volfovsky, N., Runfola, T., Jones, T. L., Mackiewicz, M., Martin, S. E., Mushinski, J. F., Stephens, R., & Caplen, N. J. (2008). The Identification of MicroRNAs in a Genomically Unstable Region of Human Chromosome 8q24. *Molecular Cancer Research*, *6*(2), 212–221. https://doi.org/10.1158/1541-7786.MCR-07-0105
Ipsaro, J. J., Haase, A. D., Knott, S. R., Joshua-Tor, L., & Hannon, G. J. (2012). The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature*, *491*(7423), 279–283. https://doi.org/10.1038/nature11502
Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.-L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J. H., & Ackerman, S. L. (2014). Ribosome stalling induced by mutation of

a CNS-specific tRNA causes neurodegeneration. *Science*, *345*(6195), 455–459. https://doi.org/10.1126/science.1249749

Isoda, T., Moore, A. J., He, Z., Chandra, V., Aida, M., Denholtz, M., Piet van Hamburg, J., Fisch, K. M., Chang, A. N., Fahl, S. P., Wiest, D. L., & Murre, C. (2017). Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. *Cell*, *171*(1), 103-119.e18. https://doi.org/10.1016/j.cell.2017.09.001

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H. T., Orejuela, M. R., Piechotta, M., Levanon, E. Y., Landthaler, M., Dieterich, C., & Rajewsky, N. (2015). Analysis of Intron Sequences Reveals Hallmarks of Circular RNA Biogenesis in Animals. *Cell Reports*, *10*(2), 170–177. https://doi.org/10.1016/j.celrep.2014.12.019

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T. R., Prensner, J. R., Evans, J. R., Zhao, S., Poliakov, A., Cao, X., Dhanasekaran, S. M., Wu, Y.-M., Robinson, D. R., Beer, D. G., Feng, F. Y., Iyer, H. K., & Chinnaiyan, A. M. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nature Genetics*, *47*(3), 199–208. https://doi.org/10.1038/ng.3192

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, *3*(3), 318–356. https://doi.org/10.1016/S0022-2836(61)80072-7

Jayapal, S. R., Lee, K. L., Ji, P., Kaldis, P., Lim, B., & Lodish, H. F. (2010). Down-regulation of Myc Is Essential for Terminal Erythroid Maturation. *Journal of Biological Chemistry*, *285*(51), 40252–40265. https://doi.org/10.1074/jbc.M110.181073

Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., Marzluff, W. F., & Sharpless, N. E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, *19*(2), 141–157. https://doi.org/10.1261/rna.035667.112

Jiang, P., Du, W., Wang, X., Mancuso, A., Gao, X., Wu, M., & Yang, X. (2011). p53 regulates biosynthesis through direct inactivation of glucose-6-phosphate dehydrogenase. *Nature Cell Biology*, *13*(3), 310–316. https://doi.org/10.1038/ncb2172

Jo, M. H., Shin, S., Jung, S.-R., Kim, E., Song, J.-J., & Hohng, S. (2015). Human Argonaute 2 Has Diverse Reaction Pathways on Target RNAs. *Molecular Cell*, *59*(1), 117–124. https://doi.org/10.1016/j.molcel.2015.04.027

Jöchl, C., Rederstorff, M., Hertel, J., Stadler, P. F., Hofacker, I. L., Schrettl, M., Haas, H., & Hüttenhofer, A. (2008). Small ncRNA transcriptome analysis from Aspergillus fumigatus suggests a novel mechanism for regulation of protein synthesis. *Nucleic Acids Research*, *36*(8), 2677–2689. https://doi.org/10.1093/nar/gkn123

Jonas, S., & Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nature Reviews Genetics*, *16*(7), 421–433. https://doi.org/10.1038/nrg3965

Jurica, M. S., & Moore, M. J. (2003). Pre-mRNA Splicing. *Molecular Cell*, *12*(1), 5–14. https://doi.org/10.1016/S1097-2765(03)00270-3

Kafina, M. D., & Paw, B. H. (2018). *Using the Zebrafish as an Approach to Examine the Mechanisms of Vertebrate Erythropoiesis* (pp. 11–36). https://doi.org/10.1007/978-1-4939-7428-3_2

Kapusta, A., Kronenberg, Z., Lynch, V. J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., & Feschotte, C. (2013). Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *PLoS Genetics*, *9*(4), e1003470. https://doi.org/10.1371/journal.pgen.1003470

Katayama, S., Kanamori, M., & Hayashizaki, Y. (2004). Integrated analysis of the genome and the transcriptome by FANTOM. *Briefings in Bioinformatics*, *5*(3), 249–258. https://doi.org/10.1093/bib/5.3.249

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C. C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K. C., Hallinan, J., Mattick, J., … Wahlestedt, C. (2005). Antisense Transcription in the Mammalian Transcriptome. *Science*, *309*(5740), 1564–1566. https://doi.org/10.1126/science.1112009

Kikuchi, Y., Tokita, S., Hirama, T., Kochin, V., Nakatsugawa, M., Shinkawa, T., Hirohashi, Y., Tsukahara, T., Hata, F., Takemasa, I., Sato, N., Kanaseki, T., & Torigoe, T. (2021). CD8+ T–cell Immune Surveillance against a Tumor Antigen Encoded by the Oncogenic Long Noncoding RNA *PVT1*. *Cancer Immunology Research*, *9*(11), 1342–1353. https://doi.org/10.1158/2326-6066.CIR-20-0964

Kim, H., & D'Andrea, A. D. (2012). Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes & Development*, *26*(13), 1393–1408. https://doi.org/10.1101/gad.195248.112

Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P. F., Kreiman, G., & Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, *465*(7295), 182–187. https://doi.org/10.1038/nature09033

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., & Schilling, T. F. (1995). *Stages of Embryonic Development of the Zebrafish*.

Kirchner, S., & Ignatova, Z. (2015). Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nature Reviews Genetics*, *16*(2), 98–112. https://doi.org/10.1038/nrg3861

Kirk, J. M., Kim, S. O., Inoue, K., Smola, M. J., Lee, D. M., Schertzer, M. D., Wooten, J. S., Baker, A. R., Sprague, D., Collins, D. W., Horning, C. R., Wang, S., Chen, Q., Weeks, K. M., Mucha, P. J., & Calabrese, J. M. (2018). Functional classification of long non-coding RNAs by k-mer content. *Nature Genetics*, *50*(10), 1474–1482. https://doi.org/10.1038/s41588-018-0207-8

Kishore, S., Gruber, A. R., Jedlinski, D. J., Syed, A. P., Jorjani, H., & Zavolan, M. (2013). Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biology*, *14*(5), R45. https://doi.org/10.1186/gb-2013-14-5-r45

Kiss, T. (2004). Biogenesis of small nuclear RNPs. *Journal of Cell Science*, *117*(25), 5949–5951. https://doi.org/10.1242/jcs.01487

Kiss-László, Z., Henry, Y., Bachellerie, J.-P., Caizergues-Ferrer, M., & Kiss, T. (1996). Site-Specific Ribose Methylation of Preribosomal RNA: A Novel Function for Small Nucleolar RNAs. *Cell*, *85*(7), 1077–1088. https://doi.org/10.1016/S0092-8674(00)81308-2

Kleaveland, B., Shi, C. Y., Stefano, J., & Bartel, D. P. (2018). A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. *Cell*, *174*(2), 350-362.e17. https://doi.org/10.1016/j.cell.2018.05.022

Kloc, M., Wilk, K., Vargas, D., Shirato, Y., Bilinski, S., & Etkin, L. D. (2005). Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of *Xenopus* oocytes. *Development*, *132*(15), 3445–3457. https://doi.org/10.1242/dev.01919

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., & Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*, *35*(suppl_2), W345–W349. https://doi.org/10.1093/nar/gkm391

Kong, R., Zhang, E., Yin, D., You, L., Xu, T., Chen, W., Xia, R., Wan, L., Sun, M., Wang, Z., De, W., & Zhang, Z. (2015). Long noncoding RNA PVT1 indicates a poor prognosis of gastric cancer and promotes cell proliferation through epigenetically regulating p15 and p16. *Molecular Cancer*, *14*(1), 82. https://doi.org/10.1186/s12943-015-0355-8

Kopp, F., Elguindy, M. M., Yalvac, M. E., Zhang, H., Chen, B., Gillett, F. A., Lee, S., Sivakumar, S., Yu, H., Xie, Y., Mishra, P., Sahenk, Z., & Mendell, J. T. (2019). PUMILIO hyperactivity drives premature aging of Norad-deficient mice. *ELife*, *8*. https://doi.org/10.7554/eLife.42650

Kopp, F., & Mendell, J. T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*, *172*(3), 393–407. https://doi.org/10.1016/j.cell.2018.01.011

Kotian, S., Banerjee, T., Lockhart, A., Huang, K., Catalyurek, U. v, & Parvin, J. D. (2014). NUSAP1 influences the DNA damage response by controlling BRCA1 protein levels. *Cancer Biology & Therapy*, *15*(5), 533–543. https://doi.org/10.4161/cbt.28019

Kramer, A. C., Weber, J., Zhang, Y., Tolar, J., Gibbens, Y. Y., Shevik, M., & Lund, T. C. (2017). TP53 Modulates Oxidative Stress in Gata1 + Erythroid Cells. *Stem Cell Reports*, *8*(2), 360–372. https://doi.org/10.1016/j.stemcr.2016.12.025

Kristensen, L. S., Andersen, M. S., Stagsted, L. V. W., Ebbesen, K. K., Hansen, T. B., & Kjems, J. (2019). The biogenesis, biology and characterization of circular RNAs. *Nature Reviews Genetics*, *20*(11), 675–691. https://doi.org/10.1038/s41576-019-0158-7

Kulkeaw, K., & Sugiyama, D. (2012). Zebrafish erythropoiesis and the utility of fish as models of anemia. *Stem Cell Research & Therapy*, *3*(6), 55. https://doi.org/10.1186/scrt146

Kumari, P., & Sampath, K. (2015). cncRNAs: Bi-functional RNAs with protein coding and non-coding functions. *Seminars in Cell & Developmental Biology*, *47–48*, 40–51. https://doi.org/10.1016/j.semcdb.2015.10.024

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., … Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Langheinrich, U., Hennen, E., Stott, G., & Vacun, G. (2002). Zebrafish as a Model Organism for the Identification and Characterization of Drugs and Genes Affecting p53 Signaling. *Current Biology*, *12*(23), 2023–2028. https://doi.org/10.1016/S0960-9822(02)01319-2

Latos, P. A., & Hemberger, M. (2016). From the stem of the placental tree: trophoblast stem cells and their progeny. *Development*, *143*(20), 3650–3660. https://doi.org/10.1242/dev.133462

Lau, N. C., Seto, A. G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D. P., & Kingston, R. E. (2006). Characterization of the piRNA Complex from Rat Testes. *Science*, *313*(5785), 363–367. https://doi.org/10.1126/science.1130164

Lavalou, P., Eckert, H., Damy, L., Constanty, F., Majello, S., Bitetti, A., Graindorge, A., & Shkumatava, A. (2019). *Strategies for genetic inactivation of long noncoding RNAs in zebrafish*. https://doi.org/10.1261/rna

le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., Hur, J. K., Aravin, A. A., & Tóth, K. F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes & Development*, *27*(4), 390–399. https://doi.org/10.1101/gad.209841.112

Lee, J., Wu, Y., Harada, B. T., Li, Y., Zhao, J., He, C., Ma, Y., & Wu, X. (2021). N ^{6} -methyladenosine modification of lncRNA *Pvt1* governs epidermal stemness. *The EMBO Journal*, *40*(8). https://doi.org/10.15252/embj.2020106276

Lee, J.-S., & Mendell, J. T. (2020). Antisense-Mediated Transcript Knockdown Triggers Premature Transcription Termination. *Molecular Cell*, *77*(5), 1044-1054.e3. https://doi.org/10.1016/j.molcel.2019.12.011

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, *75*(5), 843–854. https://doi.org/10.1016/0092-8674(93)90529-Y

Lee, T. I., & Young, R. A. (2013). Transcriptional Regulation and Its Misregulation in Disease. *Cell*, *152*(6), 1237–1251. https://doi.org/10.1016/j.cell.2013.02.014

Lekka, E., & Hall, J. (2018). Noncoding <scp>RNA</scp> s in disease. *FEBS Letters*, *592*(17), 2884–2900. https://doi.org/10.1002/1873-3468.13182

Levitz, R., Chapman, D., Amitsur, M., Green, R., Snyder, L., & Kaufmann, G. (1990). The optional E. coli prr locus encodes a latent form of phage T4-induced anticodon nuclease. *The EMBO Journal*, *9*(5), 1383–1389. https://doi.org/10.1002/j.1460-2075.1990.tb08253.x

Li, J., & Liu, C. (2019). Coding or Noncoding, the Converging Concepts of RNAs. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00496

Li, M., Shao, F., Qian, Q., Yu, W., Zhang, Z., Chen, B., Su, D., Guo, Y., Phan, A.-V., Song, L., Stephens, S. B., Sebag, J., Imai, Y., Yang, L., & Cao, H. (2021). A putative long noncoding RNA-encoded micropeptide maintains cellular homeostasis in pancreatic β cells. *Molecular Therapy - Nucleic Acids*, *26*, 307–320. https://doi.org/10.1016/j.omtn.2021.06.027

Li, P., Lin, X.-J., Yang, Y., Yang, A.-K., Di, J.-M., Jiang, Q.-W., Huang, J.-R., Yuan, M.-L., Xing, Z.-H., Wei, M.-N., Li, Y., Yuan, X.-H., Shi, Z., Liu, H., & Ye, J. (2019). Reciprocal regulation of miR-1205 and E2F1 modulates progression of laryngeal squamous cell carcinoma. *Cell Death & Disease*, *10*(12), 916. https://doi.org/10.1038/s41419-019-2154-4

Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W., Dai, L., Zhu, P., Chang, Z., Wu, Q., Zhao, Y., Jia, Y., Xu, P., Liu, H., & Shan, G. (2015). Exon-intron circular RNAs regulate transcription in the nucleus. *Nature Structural & Molecular Biology*, *22*(3), 256–264. https://doi.org/10.1038/nsmb.2959

Liu, D.-W., Zhang, J.-H., Liu, F.-X., Wang, X.-T., Pan, S.-K., Jiang, D.-K., Zhao, Z.-H., & Liu, Z.-S. (2019). Silencing of long noncoding RNA PVT1 inhibits podocyte damage and apoptosis in diabetic nephropathy by upregulating FOXA1. *Experimental & Molecular Medicine*, *51*(8), 1–15. https://doi.org/10.1038/s12276-019-0259-6

Lin, N., Chang, K.-Y., Li, Z., Gates, K., Rana, Z. A., Dang, J., Zhang, D., Han, T., Yang, C.-S., Cunningham, T. J., Head, S. R., Duester, G., Dong, P. D. S., & Rana, T. M. (2014). An Evolutionarily Conserved Long Noncoding RNA TUNA Controls Pluripotency and Neural Lineage Commitment. *Molecular Cell*, *53*(6), 1005–1019. https://doi.org/10.1016/j.molcel.2014.01.021

Liu, J., Gough, J., & Rost, B. (2006). Distinguishing Protein-Coding from Non-Coding RNAs through Support Vector Machines. *PLoS Genetics*, *2*(4), e29. https://doi.org/10.1371/journal.pgen.0020029

Liu, S., Chen, W., Hu, H., Zhang, T., Wu, T., Li, X., Li, Y., Kong, Q., Lu, H., & Lu, Z. (2021). Long noncoding RNA PVT1 promotes breast cancer proliferation and metastasis by binding miR-128-3p and UPF1. *Breast Cancer Research*, *23*(1), 115. https://doi.org/10.1186/s13058-021-01491-y

Liu, T. X., Howlett, N. G., Deng, M., Langenau, D. M., Hsu, K., Rhodes, J., Kanki, J. P., D'Andrea, A. D., & Look, A. T. (2003). Knockdown of Zebrafish Fancd2 Causes Developmental Abnormalities via p53-Dependent Apoptosis. *Developmental Cell*, *5*(6), 903–914. https://doi.org/10.1016/S1534-5807(03)00339-3

Liu, X., Bi, L., Wang, Q., Wen, M., Li, C., Ren, Y., Jiao, Q., Mao, J.-H., Wang, C., Wei, G., & Wang, Y. (2018). miR-1204 targets VDR to promotes epithelial-mesenchymal transition and metastasis in breast cancer. *Oncogene*, *37*(25), 3426–3439. https://doi.org/10.1038/s41388-018-0215-2

Loewenstein, Y., Raimondo, D., Redfern, O. C., Watson, J., Frishman, D., Linial, M., Orengo, C., Thornton, J., & Tramontano, A. (2009). Protein function annotation by homology-based inference. *Genome Biology*, *10*(2), 207. https://doi.org/10.1186/gb-2009-10-2-207

Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., Rheinheimer, S., Meder, B., Stähler, C., Meese, E., & Keller, A. (2016). Distribution of miRNA expression across human tissues. *Nucleic Acids Research*, *44*(8), 3865–3877. https://doi.org/10.1093/nar/gkw116

Luo, L. Z., Gopalakrishna-Pillai, S., Nay, S. L., Park, S.-W., Bates, S. E., Zeng, X., Iverson, L. E., & O'Connor, T. R. (2012). DNA Repair in Human Pluripotent Stem Cells Is Distinct from That in Non-Pluripotent Human Cells. *PLoS ONE*, *7*(3), e30541. https://doi.org/10.1371/journal.pone.0030541

Lyons, S. M., Gudanis, D., Coyne, S. M., Gdaniec, Z., & Ivanov, P. (2017). Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs. *Nature Communications*, *8*(1), 1127. https://doi.org/10.1038/s41467-017-01278-w

Maass, P. G., Glažar, P., Memczak, S., Dittmar, G., Hollfinger, I., Schreyer, L., Sauer, A. v., Toka, O., Aiuti, A., Luft, F. C., & Rajewsky, N. (2017). A map of human circular RNAs in clinically relevant tissues. *Journal of Molecular Medicine*, *95*(11), 1179–1189. https://doi.org/10.1007/s00109-017-1582-9

Marchese, F. P., Raimondi, I., & Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biology*, *18*(1), 206. https://doi.org/10.1186/s13059-017-1348-2

Martínez-Barriocanal, Á., Arango, D., & Dopeso, H. (2020). PVT1 Long Non-coding RNA in Gastrointestinal Cancer. *Frontiers in Oncology*, *10*. https://doi.org/10.3389/fonc.2020.00038

Matranga, C., Tomari, Y., Shin, C., Bartel, D. P., & Zamore, P. D. (2005). Passenger-Strand Cleavage Facilitates Assembly of siRNA into Ago2-Containing RNAi Enzyme Complexes. *Cell*, *123*(4), 607–620. https://doi.org/10.1016/j.cell.2005.08.044

Maxwell, E., & Fournier, M. (1995). THE SMALL NUCLEOLAR RNAs. *Annual Review of Biochemistry*, *64*(1), 897–934. https://doi.org/10.1146/annurev.bi.64.070195.004341

McHugh, C. A., Russell, P., & Guttman, M. (2014). Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biology*, *15*(1), 203. https://doi.org/10.1186/gb4152

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S. D., Gregersen, L. H., Munschauer, M., Loewer, A., Ziebold, U., Landthaler, M., Kocks, C., le Noble, F., & Rajewsky, N. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, *495*(7441), 333–338. https://doi.org/10.1038/nature11928

Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F., & Mattick, J. S. (2008). Specific expression of long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences*, *105*(2), 716–721. https://doi.org/10.1073/pnas.0706729105

Morais, P., Adachi, H., & Yu, Y.-T. (2021). Spliceosomal snRNA Epitranscriptomics. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.652129

Mori, T., Ato, S., Knudsen, J. R., Henriquez-Olguin, C., Li, Z., Wakabayashi, K., Suginohara, T., Higashida, K., Tamura, Y., Nakazato, K., Jensen, T. E., & Ogasawara, R. (2021). c-Myc overexpression increases ribosome biogenesis and protein synthesis

independent of mTORC1 activation in mouse skeletal muscle. *American Journal of Physiology-Endocrinology and Metabolism*, *321*(4), E551–E559. https://doi.org/10.1152/ajpendo.00164.2021

Much, C., Smallegan, M. J., Hwang, T., Hanson, S. D., Dumbović, G., & Rinn, J. L. (2022). Evolutionary divergence of *Firre* localization and expression. *RNA*, *28*(6), 842–853. https://doi.org/10.1261/rna.079070.121

Nakagawa, S., Ip, J. Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T., & Prasanth, K. v. (2012). Malat1 is not an essential component of nuclear speckles in mice. *RNA*, *18*(8), 1487–1499. https://doi.org/10.1261/rna.033217.112

Nakanishi, K., Weinberg, D. E., Bartel, D. P., & Patel, D. J. (2012). Structure of yeast Argonaute with guide RNA. *Nature*, *486*(7403), 368–374. https://doi.org/10.1038/nature11211

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J. C., Grützner, F., & Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, *505*(7485), 635–640. https://doi.org/10.1038/nature12943

Ni, J., Tien, A. L., & Fournier, M. J. (1997). Small Nucleolar RNAs Direct Site-Specific Synthesis of Pseudouridine in Ribosomal RNA. *Cell*, *89*(4), 565–573. https://doi.org/10.1016/S0092-8674(00)80238-X

Nicoloso, M., Qu, L.-H., Michot, B., & Bachellerie, J.-P. (1996). Intron-encoded, Antisense Small Nucleolar RNAs: The Characterization of Nine Novel Species Points to Their Direct Role as Guides for the 2′-O-ribose Methylation of rRNAs. *Journal of Molecular Biology*, *260*(2), 178–195. https://doi.org/10.1006/jmbi.1996.0391

Nishimasu, H., Ishizu, H., Saito, K., Fukuhara, S., Kamatani, M. K., Bonnefond, L., Matsumoto, N., Nishizawa, T., Nakanaga, K., Aoki, J., Ishitani, R., Siomi, H., Siomi, M. C., & Nureki, O. (2012). Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature*, *491*(7423), 284–287. https://doi.org/10.1038/nature11509

Niu, L., Lou, F., Sun, Y., Sun, L., Cai, X., Liu, Z., Zhou, H., Wang, H., Wang, Z., Bai, J., Yin, Q., Zhang, J., Chen, L., Peng, D., Xu, Z., Gao, Y., Tang, S., Fan, L., & Wang, H. (2020). A micropeptide encoded by lncRNA MIR155HG suppresses autoimmune inflammation via modulating antigen presentation. *Science Advances*, *6*(21). https://doi.org/10.1126/sciadv.aaz2059

Noller, H. F., Lancaster, L., Mohan, S., & Zhou, J. (2017). Ribosome structural dynamics in translocation: yet another functional role for ribosomal RNA. *Quarterly Reviews of Biophysics*, *50*, e12. https://doi.org/10.1017/S0033583517000117

Northcott, P. A., Shih, D. J. H., Peacock, J., Garzia, L., Sorana Morrissy, A., Zichner, T., Stütz, A. M., Korshunov, A., Reimand, J., Schumacher, S. E., Beroukhim, R., Ellison, D. W., Marshall, C. R., Lionel, A. C., Mack, S., Dubuc, A., Yao, Y., Ramaswamy, V., Luu, B., … Taylor, M. D. (2012). Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature*, *488*(7409), 49–56. https://doi.org/10.1038/nature11327

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. v., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., … Phillippy, A. M. (2022). The complete sequence of a human genome. *Science*, *376*(6588), 44–53. https://doi.org/10.1126/science.abj6987

O'brien, B. R. A. (1961). Identification of Haemoglobin by its Catalase Reaction with Peroxide and *o* -Dianisidine. *Stain Technology*, *36*(2), 57–61. https://doi.org/10.3109/10520296109113244

Ogunwobi, O. O., & Kumar, A. (2019). Chemoresistance Mediated by ceRNA Networks Associated With the PVT1 Locus. *Frontiers in Oncology*, *9*. https://doi.org/10.3389/fonc.2019.00834

Ohno S. (1972). So much "junk" DNA in our genome. *Brookhaven Symp Biol*, *23*, 366–370.

Okamura, K., & Lai, E. C. (2008). Endogenous small interfering RNAs in animals. *Nature Reviews Molecular Cell Biology*, *9*(9), 673–678. https://doi.org/10.1038/nrm2479

Olivero, C. E., Martínez-Terroba, E., Zimmer, J., Liao, C., Tesfaye, E., Hooshdaran, N., Schofield, J. A., Bendor, J., Fang, D., Simon, M. D., Zamudio, J. R., & Dimitrova, N. (2020). p53 Activates the Long Noncoding RNA Pvt1b to Inhibit Myc and Suppress Tumorigenesis. *Molecular Cell*, *77*(4), 761-774.e8. https://doi.org/10.1016/j.molcel.2019.12.014

Onagoruwa, O. T., Pal, G., Ochu, C., & Ogunwobi, O. O. (2020). Oncogenic Role of PVT1 and Therapeutic Implications. *Frontiers in Oncology*, *10*. https://doi.org/10.3389/fonc.2020.00017

Paci, P., Colombo, T., & Farina, L. (2014). Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. *BMC Systems Biology*, *8*(1), 83. https://doi.org/10.1186/1752-0509-8-83

Paffett-Lugassy, N., Hsia, N., Fraenkel, P. G., Paw, B., Leshinsky, I., Barut, B., Bahary, N., Caro, J., Handin, R., & Zon, L. I. (2007). Functional conservation of erythropoietin signaling in zebrafish. *Blood*, *110*(7), 2718–2726. https://doi.org/10.1182/blood-2006-04-016535

Pal, G., Di, L., Orunmuyi, A., Olapade-Olaopa, E. O., Qiu, W., & Ogunwobi, O. O. (2020). Population Differentiation at the PVT1 Gene Locus: Implications for Prostate Cancer. *G3 Genes|Genomes|Genetics*, *10*(7), 2257–2264. https://doi.org/10.1534/g3.120.401291

Pal, G., & Ogunwobi, O. O. (2019). Copy number-based quantification assay for non-invasive detection of PVT1-derived transcripts. *PLOS ONE*, *14*(12), e0226620. https://doi.org/10.1371/journal.pone.0226620

Palade, G. E. (1955). A SMALL PARTICULATE COMPONENT OF THE CYTOPLASM. *The Journal of Biophysical and Biochemical Cytology*, *1*(1), 59–68. https://doi.org/10.1083/jcb.1.1.59

Palcau, A. C., Canu, V., Donzelli, S., Strano, S., Pulito, C., & Blandino, G. (2022). CircPVT1: a pivotal circular node intersecting Long Non-Coding-PVT1 and c-MYC oncogenic signals. *Molecular Cancer*, *21*(1), 33. https://doi.org/10.1186/s12943-022-01514-y

Pane, A., Wehr, K., & Schüpbach, T. (2007). zucchini and squash Encode Two Putative Nucleases Required for rasiRNA Production in the Drosophila Germline. *Developmental Cell*, *12*(6), 851–862. https://doi.org/10.1016/j.devcel.2007.03.022

Parisien, M., Wang, X., & Pan, T. (2013). Diversity of human tRNA genes from the 1000-genomes project. *RNA Biology*, *10*(12), 1853–1867. https://doi.org/10.4161/rna.27361

Pauli, A., Valen, E., Lin, M. F., Garber, M., Vastenhouw, N. L., Levin, J. Z., Fan, L., Sandelin, A., Rinn, J. L., Regev, A., & Schier, A. F. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Research*, *22*(3), 577–591. https://doi.org/10.1101/gr.133009.111

Perry, R. B.-T., & Ulitsky, I. (2016). The functions of long noncoding RNAs in development and stem cells. *Development*, *143*(21), 3882–3894. https://doi.org/10.1242/dev.140962

Pijuan-Sala, B., Griffiths, J. A., Guibentif, C., Hiscock, T. W., Jawaid, W., Calero-Nieto, F. J., Mulas, C., Ibarra-Soria, X., Tyser, R. C. v., Ho, D. L. L., Reik, W., Srinivas, S., Simons, B. D., Nichols, J., Marioni, J. C., & Göttgens, B. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*, *566*(7745), 490–495. https://doi.org/10.1038/s41586-019-0933-9

Piquet, S., le Parc, F., Bai, S.-K., Chevallier, O., Adam, S., & Polo, S. E. (2018). The Histone Chaperone FACT Coordinates H2A.X-Dependent Signaling and Repair of DNA Damage. *Molecular Cell*, *72*(5), 888-901.e7. https://doi.org/10.1016/j.molcel.2018.09.010

Piwecka, M., Glažar, P., Hernandez-Miranda, L. R., Memczak, S., Wolf, S. A., Rybak-Wolf, A., Filipchyk, A., Klironomos, F., Cerda Jara, C. A., Fenske, P., Trimbuch, T., Zywitza, V., Plass, M., Schreyer, L., Ayoub, S., Kocks, C., Kühn, R., Rosenmund, C., Birchmeier, C., & Rajewsky, N. (2017). Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science*, *357*(6357). https://doi.org/10.1126/science.aam8526

Plaster, N., Sonntag, C., Busse, C. E., & Hammerschmidt, M. (2006). p53 deficiency rescues apoptosis and differentiation of multiple cell types in zebrafish flathead mutants deficient for zygotic DNA polymerase δ1. *Cell Death & Differentiation*, *13*(2), 223–235. https://doi.org/10.1038/sj.cdd.4401747

Ponjavic, J., Ponting, C. P., & Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Research*, *17*(5), 556–565. https://doi.org/10.1101/gr.6036807

Poortinga, G., Hannan, K. M., Snelling, H., Walkley, C. R., Jenkins, A., Sharkey, K., Wall, M., Brandenburger, Y., Palatsides, M., Pearson, R. B., McArthur, G. A., & Hannan, R. D. (2004). MAD1 and c-MYC regulate UBF and rDNA transcription during granulocyte differentiation. *The EMBO Journal*, *23*(16), 3325–3335. https://doi.org/10.1038/sj.emboj.7600335

Poortinga, G., Wall, M., Sanij, E., Siwicki, K., Ellul, J., Brown, D., Holloway, T. P., Hannan, R. D., & McArthur, G. A. (2011). c-MYC coordinately regulates ribosomal gene chromatin remodeling and Pol I availability during granulocyte differentiation. *Nucleic Acids Research*, *39*(8), 3267–3281. https://doi.org/10.1093/nar/gkq1205

Qian, Y., Mao, Z., Shi, Y., Liu, Z., Cao, Q., & Zhang, Q. (2018). Comprehensive Analysis of miRNA-mRNA-lncRNA Networks in Non-Smoking and Smoking Patients with Chronic Obstructive Pulmonary Disease. *Cellular Physiology and Biochemistry*, *50*(3), 1140–1153. https://doi.org/10.1159/000494541

Qin, X., Chen, Y., Chen, S., Liu, X., Zeng, W., Tian, F., Lin, Y., & Fan, C. (2019). Plasmacytoma variant translocation 1 (*PVT1*) regulates trophoblast viability, proliferation, and migration and is downregulated in spontaneous abortion. *American Journal of Reproductive Immunology*, *81*(1), e13048. https://doi.org/10.1111/aji.13048

Ramanagoudr-Bhojappa, R., Carrington, B., Ramaswami, M., Bishop, K., Robbins, G. M., Jones, M., Harper, U., Frederickson, S. C., Kimble, D. C., Sood, R., & Chandrasekharappa, S. C. (2018). Multiplexed CRISPR/Cas9-mediated knockout of 19 Fanconi anemia pathway genes in zebrafish revealed their roles in growth, sexual development and fertility. *PLOS Genetics*, *14*(12), e1007821.

Ramanathan, M., Porter, D. F., & Khavari, P. A. (2019). Methods to study RNA–protein interactions. *Nature Methods*, *16*(3), 225–234. https://doi.org/10.1038/s41592-019-0330-1

Reinardy, H. C., Dharamshi, J., Jha, A. N., & Henry, T. B. (2013). Changes in expression profiles of genes associated with DNA repair following induction of DNA damage in larval zebrafish Danio rerio. *Mutagenesis*, *28*(5), 601–608. https://doi.org/10.1093/mutage/get038

Rich, A., & RajBhandary, U. L. (1976). Transfer RNA: Molecular Structure, Sequence, and Properties. *Annual Review of Biochemistry*, *45*(1), 805–860. https://doi.org/10.1146/annurev.bi.45.070176.004105

Rodríguez-Marí, A., Cañestro, C., BreMiller, R. A., Nguyen-Johnson, A., Asakawa, K., Kawakami, K., & Postlethwait, J. H. (2010). Sex Reversal in Zebrafish fancl Mutants Is Caused by Tp53-Mediated Germ Cell Apoptosis. *PLoS Genetics*, *6*(7), e1001034. https://doi.org/10.1371/journal.pgen.1001034

Rodríguez-Marí, A., & Postlethwait, J. H. (2011). *The Role of Fanconi Anemia/BRCA Genes in Zebrafish Sex Determination* (pp. 461–490). https://doi.org/10.1016/B978-0-12-381320-6.00020-5

Roeder, R. G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Letters*, *579*(4), 909–915. https://doi.org/10.1016/j.febslet.2004.12.007

Ron, M., & Ulitsky, I. (2022). Context-specific effects of sequence elements on subcellular localization of linear and circular RNAs. *Nature Communications*, *13*(1), 2481. https://doi.org/10.1038/s41467-022-30183-0

Ros, S., Flöter, J., Kaymak, I., da Costa, C., Houddane, A., Dubuis, S., Griffiths, B., Mitter, R., Walz, S., Blake, S., Behrens, A., Brindle, K. M., Zamboni, N., Rider, M. H., & Schulze, A. (2017). 6-Phosphofructo-2-kinase/fructose-2,6-biphosphatase 4 is essential for p53-null cancer cells. *Oncogene*, *36*(23), 3287–3299. https://doi.org/10.1038/onc.2016.477

Rospo, G., Lorenzato, A., Amirouchene-Angelozzi, N., Magrì, A., Cancelliere, C., Corti, G., Negrino, C., Amodio, V., Montone, M., Bartolini, A., Barault, L., Novara, L., Isella, C., Medico, E., Bertotti, A., Trusolino, L., Germano, G., di Nicolantonio, F., & Bardelli, A. (2019). Evolving neoantigen profiles in colorectal cancers with DNA repair defects. *Genome Medicine*, *11*(1), 42. https://doi.org/10.1186/s13073-019-0654-6

Ross, C. J., Rom, A., Spinrad, A., Gelbard-Solodkin, D., Degani, N., & Ulitsky, I. (2021). Uncovering deeply conserved motif combinations in rapidly evolving noncoding sequences. *Genome Biology*, *22*(1). https://doi.org/10.1186/s13059-020-02247-1

Ross, C. J., & Ulitsky, I. (2022). Discovering functional motifs in long noncoding <scp>RNAs</scp>. *WIREs RNA*. https://doi.org/10.1002/wrna.1708

Ruan, X., Li, P., Chen, Y., Shi, Y., Pirooznia, M., Seifuddin, F., Suemizu, H., Ohnishi, Y., Yoneda, N., Nishiwaki, M., Shepherdson, J., Suresh, A., Singh, K., Ma, Y., Jiang, C., & Cao, H. (2020). In vivo functional analysis of non-conserved human lncRNAs associated with cardiometabolic traits. *Nature Communications*, *11*(1), 45. https://doi.org/10.1038/s41467-019-13688-z

Ruiz-Orera, J., Messeguer, X., Subirana, J. A., & Alba, M. M. (2014). Long non-coding RNAs as a source of new peptides. *ELife*, *3*. https://doi.org/10.7554/eLife.03523

Sakai, K., Ohta, T., Minoshima, S., Kudoh, J., Wang, Y., de Jong, P. J., & Shimizu, N. (1995). Human ribosomal RNA gene cluster: identification of the proximal end containing a novel tandem repeat sequence. *Genomics*, *26*(3), 521–526. https://doi.org/10.1016/0888-7543(95)80170-Q

Salomon, W. E., Jolly, S. M., Moore, M. J., Zamore, P. D., & Serebrov, V. (2015). Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell*, *162*(1), 84–95. https://doi.org/10.1016/j.cell.2015.06.029

Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L., & Brown, P. O. (2013). Cell-Type Specific Features of Circular RNA Expression. *PLoS Genetics*, *9*(9), e1003777. https://doi.org/10.1371/journal.pgen.1003777

Salzman, J., Gawad, C., Wang, P. L., Lacayo, N., & Brown, P. O. (2012). Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types. *PLoS ONE*, *7*(2), e30733. https://doi.org/10.1371/journal.pone.0030733

Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J., & Kleinschmidt, A. K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proceedings of the National Academy of Sciences*, *73*(11), 3852–3856. https://doi.org/10.1073/pnas.73.11.3852

Sarangdhar, M. A., Chaubey, D., Srikakulam, N., & Pillai, B. (2018). Parentally inherited long non-coding RNA Cyrano is involved in zebrafish neurodevelopment. *Nucleic Acids Research*, *46*(18), 9726–9735. https://doi.org/10.1093/nar/gky628

Sartorelli, V., & Lauberth, S. M. (2020). Enhancer RNAs are an important regulatory layer of the epigenome. *Nature Structural & Molecular Biology*, *27*(6), 521–528. https://doi.org/10.1038/s41594-020-0446-0

Sato, K., Akiyama, M., & Sakakibara, Y. (2021). RNA secondary structure prediction using deep learning with thermodynamic integration. *Nature Communications*, *12*(1), 941. https://doi.org/10.1038/s41467-021-21194-4

Sauvageau, M., Goff, L. A., Lodato, S., Bonev, B., Groff, A. F., Gerhardinger, C., Sanchez-Gomez, D. B., Hacisuleyman, E., Li, E., Spence, M., Liapis, S. C., Mallard, W., Morse, M., Swerdel, M. R., D'Ecclessis, M. F., Moore, J. C., Lai, V., Gong, G., Yancopoulos, G. D., … Rinn, J. L. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *ELife*, *2013*(2). https://doi.org/10.7554/eLife.01749

Schimmel, P. (2018). The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nature Reviews Molecular Cell Biology*, *19*(1), 45–58. https://doi.org/10.1038/nrm.2017.77

Schirle, N. T., & MacRae, I. J. (2012). The Crystal Structure of Human Argonaute2. *Science*, *336*(6084), 1037–1040. https://doi.org/10.1126/science.1221551

Schwartz, J. C., Younger, S. T., Nguyen, N.-B., Hardy, D. B., Monia, B. P., Corey, D. R., & Janowski, B. A. (2008). Antisense transcripts are targets for activating small RNAs. *Nature Structural & Molecular Biology*, *15*(8), 842–848. https://doi.org/10.1038/nsmb.1444

Schwartzenberg-Bar-Yoseph, F., Armoni, M., & Karnieli, E. (2004). The Tumor Suppressor p53 Down-Regulates Glucose Transporters ***GLUT1*** and ***GLUT4*** Gene Expression. *Cancer Research*, *64*(7), 2627–2633. https://doi.org/10.1158/0008-5472.CAN-03-0846

Scott, M., Lubin, B., Zuo, L., & Kuypers, F. (1991). Erythrocyte defense against hydrogen peroxide: Preeminent importance of catalase. *Translational Research*, *118*(1).

Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., Young, R. A., & Sharp, P. A. (2008). Divergent Transcription from Active Promoters. *Science*, *322*(5909), 1849–1851. https://doi.org/10.1126/science.1162253

Senís, E., Esgleas, M., Najas, S., Jiménez-Sábado, V., Bertani, C., Giménez-Alejandre, M., Escriche, A., Ruiz-Orera, J., Hergueta-Redondo, M., Jiménez, M., Giralt, A., Nuciforo, P., Albà, M. M., Peinado, H., Toro, D. del, Hove-Madsen, L., Götz, M., & Abad, M. (2021). TUNAR lncRNA Encodes a Microprotein that Regulates Neural Differentiation and Neurite Formation by Modulating Calcium Dynamics. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/fcell.2021.747667

Sharma, U., Conine, C. C., Shea, J. M., Boskovic, A., Derr, A. G., Bing, X. Y., Belleannee, C., Kucukural, A., Serra, R. W., Sun, F., Song, L., Carone, B. R., Ricci, E. P., Li, X. Z., Fauquier, L., Moore, M. J., Sullivan, R., Mello, C. C., Garber, M., & Rando, O. J. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science*, *351*(6271), 391–396. https://doi.org/10.1126/science.aad6780

Shi, C. Y., Kingston, E. R., Kleaveland, B., Lin, D. H., Stubna, M. W., & Bartel, D. P. (2020). The ZSWIM8 ubiquitin ligase mediates target-directed microRNA degradation. *Science*, *370*(6523). https://doi.org/10.1126/science.abc9359

Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E., & Weeks, K. M. (2014). RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nature Methods*, *11*(9), 959–965. https://doi.org/10.1038/nmeth.3029

Simsek, D., Tiu, G. C., Flynn, R. A., Byeon, G. W., Leppek, K., Xu, A. F., Chang, H. Y., & Barna, M. (2017). The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell*, *169*(6), 1051-1065.e18. https://doi.org/10.1016/j.cell.2017.05.022

Sleutels, F., Zwart, R., & Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, *415*(6873), 810–813. https://doi.org/10.1038/415810a

Sloan, K. E., Bohnsack, M. T., & Watkins, N. J. (2013). The 5S RNP Couples p53 Homeostasis to Ribosome Biogenesis and Nucleolar Stress. *Cell Reports*, *5*(1), 237–247. https://doi.org/10.1016/j.celrep.2013.08.049

Smith, M. A., Gesell, T., Stadler, P. F., & Mattick, J. S. (2013). Widespread purifying selection on RNA structure in mammals. *Nucleic Acids Research*, *41*(17), 8220–8236. https://doi.org/10.1093/nar/gkt596

Somyajit, K., Subramanya, S., & Nagaraju, G. (2010). RAD51C: a novel cancer susceptibility gene is linked to Fanconi anemia and breast cancer. *Carcinogenesis*, *31*(12), 2031–2038. https://doi.org/10.1093/carcin/bgq210

Somyajit, K., Subramanya, S., & Nagaraju, G. (2012). Distinct Roles of FANCO/RAD51C Protein in DNA Damage Signaling and Repair. *Journal of Biological Chemistry*, *287*(5), 3366–3380. https://doi.org/10.1074/jbc.M111.311241

Sousa, C., Johansson, C., Charon, C., Manyani, H., Sautter, C., Kondorosi, A., & Crespi, M. (2001). Translational and Structural Requirements of the Early Nodulin Gene *enod40* , a Short-Open Reading Frame-Containing RNA, for Elicitation of a Cell-Specific Growth Response in the Alfalfa Root Cortex. *Molecular and Cellular Biology*, *21*(1), 354–366. https://doi.org/10.1128/MCB.21.1.354-366.2001

Statello, L., Guo, C.-J., Chen, L.-L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology*, *22*(2), 96–118. https://doi.org/10.1038/s41580-020-00315-9

Sun, F., Wu, K., Yao, Z., Mu, X., Zheng, Z., Sun, M., Wang, Y., Liu, Z., & Zhu, Y. (2020). Long Noncoding RNA PVT1 Promotes Prostate Cancer Metastasis by Increasing NOP2 Expression via Targeting Tumor Suppressor MicroRNAs . *OncoTargets and Therapy*, *Volume 13*, 6755–6765. https://doi.org/10.2147/OTT.S242441

Sun, K., Chen, X., Jiang, P., Song, X., Wang, H., & Sun, H. (2013). iSeeRNA: identification of long intergenic non-coding RNA transcripts from transcriptome sequencing data. *BMC Genomics*, *14*(S2), S7. https://doi.org/10.1186/1471-2164-14-S2-S7

Sun, Y., Ren, D., Zhou, Y., Shen, J., Wu, H., & Jin, X. (2021). Histone acetyltransferase 1 promotes gemcitabine resistance by regulating the PVT1/EZH2 complex in pancreatic cancer. *Cell Death & Disease*, *12*(10), 878. https://doi.org/10.1038/s41419-021-04118-4

Taatjes, D. J. (2010). The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends in Biochemical Sciences*, *35*(6), 315–322. https://doi.org/10.1016/j.tibs.2010.02.004

Taiana, E., Ronchetti, D., Todoerti, K., Nobili, L., Tassone, P., Amodio, N., & Neri, A. (2020). LncRNA NEAT1 in Paraspeckles: A Structural Scaffold for Cellular DNA Damage Response Systems? *Non-Coding RNA*, *6*(3), 26. https://doi.org/10.3390/ncrna6030026

Taipale, M., Krykbaeva, I., Koeva, M., Kayatekin, C., Westover, K. D., Karras, G. I., & Lindquist, S. (2012). Quantitative Analysis of Hsp90-Client Interactions Reveals Principles of Substrate Recognition. *Cell*, *150*(5), 987–1001. https://doi.org/10.1016/j.cell.2012.06.047

Taliaferro, J. M., Lambert, N. J., Sudmant, P. H., Dominguez, D., Merkin, J. J., Alexis, M. S., Bazile, C. A., & Burge, C. B. (2016). RNA Sequence Context Effects Measured In Vitro Predict In Vivo Protein Binding and Regulation. *Molecular Cell*, *64*(2), 294–306. https://doi.org/10.1016/j.molcel.2016.08.035

The ENCODE Project Consortium. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*(7146), 799–816. https://doi.org/10.1038/nature05874

Thompson, D. M., Lu, C., Green, P. J., & Parker, R. (2008). tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA*, *14*(10), 2095–2103. https://doi.org/10.1261/rna.1232808

Tian, L., Hu, X., He, Y., Wu, Z., Li, D., & Zhang, H. (2018). Construction of lncRNA-miRNA-mRNA networks reveals functional lncRNAs in abdominal aortic aneurysm. *Experimental and Therapeutic Medicine*. https://doi.org/10.3892/etm.2018.6690

Tolomeo, D., Agostini, A., Visci, G., Traversa, D., & Storlazzi, C. T. (2021). PVT1: A long non-coding RNA recurrently involved in neoplasia-associated fusion transcripts. *Gene*, *779*, 145497. https://doi.org/10.1016/j.gene.2021.145497

Traversa, D., Simonetti, G., Tolomeo, D., Visci, G., Macchia, G., Ghetti, M., Martinelli, G., Kristensen, L. S., & Storlazzi, C. T. (2022). Unravelling similarities and differences in the role of circular and linear PVT1 in cancer and human disease. *British Journal of Cancer*, *126*(6), 835–850. https://doi.org/10.1038/s41416-021-01584-7

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., Freier, S. M., Bennett, C. F., Sharma, A., Bubulya, P. A., Blencowe, B. J., Prasanth, S. G., & Prasanth, K. v. (2010). The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. *Molecular Cell*, *39*(6), 925–938. https://doi.org/10.1016/j.molcel.2010.08.011

Tseng, Y.-Y., & Bagchi, A. (2015). The PVT1-MYC duet in cancer. *Molecular & Cellular Oncology*, *2*(2), e974467. https://doi.org/10.4161/23723556.2014.974467

Tseng, Y.-Y., Moriarity, B. S., Gong, W., Akiyama, R., Tiwari, A., Kawakami, H., Ronning, P., Reuland, B., Guenther, K., Beadnell, T. C., Essig, J., Otto, G. M., O'Sullivan, M. G., Largaespada, D. A., Schwertfeger, K. L., Marahrens, Y., Kawakami, Y., & Bagchi, A. (2014). PVT1 dependence in cancer with MYC copy-number increase. *Nature*, *512*(7512), 82–86. https://doi.org/10.1038/nature13311

Tuck, A. C., Natarajan, K. N., Rice, G. M., Borawski, J., Mohn, F., Rankova, A., Flemr, M., Wenger, A., Nutiu, R., Teichmann, S., & Bühler, M. (2018). Distinctive features of lincRNA gene expression suggest widespread RNA-independent functions. *Life Science Alliance*, *1*(4), e201800124. https://doi.org/10.26508/lsa.201800124

Turunen, J. J., Niemelä, E. H., Verma, B., & Frilander, M. J. (2013). The significant other: splicing by the minor spliceosome. *Wiley Interdisciplinary Reviews: RNA*, *4*(1), 61–76. https://doi.org/10.1002/wrna.1141

Tycowski, K. T., Smith, C. M., Shu, M.-D., & Steitz, J. A. (1996). A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in *Xenopus*. *Proceedings of the National Academy of Sciences*, *93*(25), 14480–14485. https://doi.org/10.1073/pnas.93.25.14480

Uechi, T., & Kenmochi, N. (2019). Zebrafish Models of Diamond-Blackfan Anemia: A Tool for Understanding the Disease Pathogenesis and Drug Discovery. *Pharmaceuticals*, *12*(4), 151. https://doi.org/10.3390/ph12040151

Ulitsky, I., & Bartel, D. P. (2013). lincRNAs: Genomics, Evolution, and Mechanisms. *Cell*, *154*(1), 26–46. https://doi.org/10.1016/j.cell.2013.06.020

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H., & Bartel, D. P. (2011). Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. *Cell*, *147*(7), 1537–1550. https://doi.org/10.1016/j.cell.2011.11.055

Unfried, J. P., & Ulitsky, I. (2022). Substoichiometric action of long noncoding RNAs. *Nature Cell Biology*, *24*(5), 608–615. https://doi.org/10.1038/s41556-022-00911-1

Uroda, T., Anastasakou, E., Rossi, A., Teulon, J.-M., Pellequer, J.-L., Annibale, P., Pessey, O., Inga, A., Chillón, I., & Marcia, M. (2019). Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway. *Molecular Cell*, *75*(5), 982-995.e9. https://doi.org/10.1016/j.molcel.2019.07.025

Vagin, V. v., Sigova, A., Li, C., Seitz, H., Gvozdev, V., & Zamore, P. D. (2006). A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. *Science*, *313*(5785), 320–324. https://doi.org/10.1126/science.1129333

Vaz, F., Hanenberg, H., Schuster, B., Barker, K., Wiek, C., Erven, V., Neveling, K., Endt, D., Kesterton, I., Autore, F., Fraternali, F., Freund, M., Hartmann, L., Grimwade, D., Roberts, R. G., Schaal, H., Mohammed, S., Rahman, N., Schindler, D., & Mathew, C. G. (2010). Mutation of the RAD51C gene in a Fanconi anemia–like disorder. *Nature Genetics*, *42*(5), 406–409. https://doi.org/10.1038/ng.570

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., … Zhu, X. (2001). The Sequence of the Human Genome. *Science*, *291*(5507), 1304–1351. https://doi.org/10.1126/science.1058040

Verduci, L., Ferraiuolo, M., Sacconi, A., Ganci, F., Vitale, J., Colombo, T., Paci, P., Strano, S., Macino, G., Rajewsky, N., & Blandino, G. (2017). The oncogenic role of circPVT1 in head and neck squamous cell carcinoma is mediated through the mutant p53/YAP/TEAD transcription-competent complex. *Genome Biology*, *18*(1), 237. https://doi.org/10.1186/s13059-017-1368-y

Wahlestedt, C. (2006). Natural antisense and noncoding RNA transcripts as potential drug targets. *Drug Discovery Today*, *11*(11–12), 503–508. https://doi.org/10.1016/j.drudis.2006.04.013

Wahlestedt, C. (2013). Targeting long non-coding RNA to therapeutically upregulate gene expression. *Nature Reviews Drug Discovery*, *12*(6), 433–446. https://doi.org/10.1038/nrd4018

Wan, B., Wu, H., Lv, D., Zhou, X., Zhong, L., Lei, B., Zhang, S., & Mao, X. (2018). Downregulation of lncRNA PVT1 expression inhibits proliferation and migration by regulating p38 expression in prostate cancer. *Oncology Letters*. https://doi.org/10.3892/ol.2018.9305

Wan, L., Sun, M., Liu, G.-J., Wei, C.-C., Zhang, E.-B., Kong, R., Xu, T.-P., Huang, M.-D., & Wang, Z.-X. (2016). Long Noncoding RNA *PVT1* Promotes Non–Small Cell Lung Cancer Cell Proliferation through Epigenetically Regulating LATS2 Expression. *Molecular Cancer Therapeutics*, *15*(5), 1082–1094. https://doi.org/10.1158/1535-7163.MCT-15-0707

Wang, H., Wei, M., Kang, Y., Xing, J., & Zhao, Y. (2020). Circular RNA circ_PVT1 induces epithelial-mesenchymal transition to promote metastasis of cervical cancer. *Aging*, *12*(20), 20139–20151. https://doi.org/10.18632/aging.103679

Wang, J., Zhang, J., Zheng, H., Li, J., Liu, D., Li, H., Samudrala, R., Yu, J., & Wong, G. K.-S. (2004). Neutral evolution of 'non-coding' complementary DNAs. *Nature*, *431*(7010), 1–2. https://doi.org/10.1038/nature03016

Wang, L., Li, J., Zhou, H., Zhang, W., Gao, J., & Zheng, P. (2021). A novel lncRNA Discn fine-tunes replication protein A (RPA) availability to promote genomic stability. *Nature Communications*, *12*(1), 5572. https://doi.org/10.1038/s41467-021-25827-6

Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J.-P., & Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Research*, *41*(6), e74–e74. https://doi.org/10.1093/nar/gkt006

Wang, L., Xiong, H., Wu, F., Zhang, Y., Wang, J., Zhao, L., Guo, X., Chang, L.-J., Zhang, Y., You, M. J., Koochekpour, S., Saleem, M., Huang, H., Lu, J., & Deng, Y. (2014). Hexokinase 2-Mediated Warburg Effect Is Required for PTEN- and p53-Deficiency-Driven Prostate Cancer Growth. *Cell Reports*, *8*(5), 1461–1474. https://doi.org/10.1016/j.celrep.2014.07.053

Wang, P. L., Bao, Y., Yee, M.-C., Barrett, S. P., Hogan, G. J., Olsen, M. N., Dinneny, J. R., Brown, P. O., & Salzman, J. (2014). Circular RNA Is Expressed across the Eukaryotic Tree of Life. *PLoS ONE*, *9*(3), e90859. https://doi.org/10.1371/journal.pone.0090859

Wang, Q., Lee, I., Ren, J., Ajay, S. S., Lee, Y. S., & Bao, X. (2013). Identification and Functional Characterization of tRNA-derived RNA Fragments (tRFs) in Respiratory Syncytial Virus Infection. *Molecular Therapy*, *21*(2), 368–379. https://doi.org/10.1038/mt.2012.237

Wang, Q., Lu, X., Li, C., Zhang, W., Lv, Y., Wang, L., Wu, L., Meng, L., Fan, Y., Ding, H., Long, W., & Lv, M. (2019). Down-regulated long non-coding RNA PVT1 contributes to gestational diabetes mellitus and preeclampsia via regulation of human trophoblast cells. *Biomedicine & Pharmacotherapy*, *120*, 109501. https://doi.org/10.1016/j.biopha.2019.109501

Wang, W., Zhou, R., Wu, Y., Liu, Y., Su, W., Xiong, W., & Zeng, Z. (2019). PVT1 Promotes Cancer Progression via MicroRNAs. *Frontiers in Oncology*, *9*. https://doi.org/10.3389/fonc.2019.00609

Wang, Y., Chen, W., Lian, J., Zhang, H., Yu, B., Zhang, M., Wei, F., Wu, J., Jiang, J., Jia, Y., Mo, F., zhang, S., Liang, X., Mou, X., & Tang, J. (2020). The lncRNA PVT1 regulates nasopharyngeal carcinoma cell proliferation via activating the KAT2A acetyltransferase and stabilizing HIF-1α. *Cell Death & Differentiation*, *27*(2), 695–710. https://doi.org/10.1038/s41418-019-0381-y

Watson, J., & Crick, F. (1953). Molecular Structure of Deoxypentose Nucleic Acids. *Nature*.

Webb, E., Adams, J. M., & Cory, S. (1984). Variant (6 ; 15) translocation in a murine plasmacytoma occurs near an immunoglobulin κ gene but far from the myc oncogene. *Nature*, *312*(5996), 777–779. https://doi.org/10.1038/312777a0

Weinberg, R. A., & Penman, S. (1968). Small molecular weight monodisperse nuclear RNA. *Journal of Molecular Biology*, *38*(3), 289–304. https://doi.org/10.1016/0022-2836(68)90387-2

Westholm, J. O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., Celniker, S. E., Graveley, B. R., & Lai, E. C. (2014). Genome-wide Analysis of Drosophila Circular RNAs Reveals Their Structural and Sequence Properties and Age-Dependent Neural Accumulation. *Cell Reports*, *9*(5), 1966–1980. https://doi.org/10.1016/j.celrep.2014.10.062

Williams, A. B., & Schumacher, B. (2016). p53 in the DNA-Damage-Repair Process. *Cold Spring Harbor Perspectives in Medicine*, *6*(5), a026070. https://doi.org/10.1101/cshperspect.a026070

Wilusz, J. E., JnBaptiste, C. K., Lu, L. Y., Kuhn, C.-D., Joshua-Tor, L., & Sharp, P. A. (2012). A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. *Genes & Development*, *26*(21), 2392–2407. https://doi.org/10.1101/gad.204438.112

Wu, M., Zhang, S., Chen, X., Xu, H., & Li, X. (2019). Expression and function of lncRNA MALAT-1 in the embryonic development of zebrafish. *Gene*, *680*, 65–71. https://doi.org/10.1016/j.gene.2018.09.037

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, *2*(3). https://doi.org/10.1016/j.xinn.2021.100141

Wu, X., & Sharp, P. A. (2013). Divergent Transcription: A Driving Force for New Gene Origination? *Cell*, *155*(5), 990–996. https://doi.org/10.1016/j.cell.2013.10.048

Wu, X.-Z., Cui, H.-P., Lv, H.-J., & Feng, L. (2019). Knockdown of lncRNA PVT1 inhibits retinoblastoma progression by sponging miR-488-3p. *Biomedicine & Pharmacotherapy*, *112*, 108627. https://doi.org/10.1016/j.biopha.2019.108627

Wutz, A., Rasmussen, T. P., & Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nature Genetics*, *30*(2), 167–174. https://doi.org/10.1038/ng820

Xia, S., Feng, J., Lei, L., Hu, J., Xia, L., Wang, J., Xiang, Y., Liu, L., Zhong, S., Han, L., & He, C. (2016). Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Briefings in Bioinformatics*, bbw081. https://doi.org/10.1093/bib/bbw081

Xie, Y., Yao, L., Yu, X., Ruan, Y., Li, Z., & Guo, J. (2020). Action mechanisms and research methods of tRNA-derived small RNAs. *Signal Transduction and Targeted Therapy*, *5*(1), 109. https://doi.org/10.1038/s41392-020-00217-4

Xu, H., Guo, S., Li, W., & Yu, P. (2015). The circular RNA Cdr1as, via miR-7 and its targets, regulates insulin transcription and secretion in islet cells. *Scientific Reports*, *5*(1), 12453. https://doi.org/10.1038/srep12453

Xu, M., Wang, Y., Weng, W., Wei, P., Qi, P., Zhang, Q., Tan, C., Ni, S., Dong, L., Yang, Y., Lin, W., Xu, Q., Huang, D., Huang, Z., Ma, Y., Zhang, W., Sheng, W., & Du, X. (2017). A Positive Feedback Loop of lncRNA- *PVT1* and FOXM1 Facilitates Gastric Cancer Growth and Invasion. *Clinical Cancer Research*, *23*(8), 2071–2080. https://doi.org/10.1158/1078-0432.CCR-16-0742

Xu, X., Wang, L., Zang, Q., Li, S., Li, L., Wang, Z., He, J., Qiang, B., Han, W., Zhang, R., Peng, X., & Abliz, Z. (2021). Rewiring of purine metabolism in response to acidosis stress in glioma stem cells. *Cell Death & Disease*, *12*(3), 277. https://doi.org/10.1038/s41419-021-03543-9

Xu, Y., Lian, Y., Zhang, Y., Huang, S., Zuo, Q., Yang, N., Chen, Y., Wu, D., & Sun, L. (2017). The long non-coding RNA *PVT1* represses *ANGPTL4* transcription through binding with EZH2 in trophoblast cell. *Journal of Cellular and Molecular Medicine*. https://doi.org/10.1111/jcmm.13405

Yamanishi, A., Matsuba, A., Kondo, R., Akamatsu, R., Tanaka, S., Tokunaga, M., Horie, K., Kokubu, C., Ishida, Y., & Takeda, J. (2018). Collection of homozygous mutant mouse embryonic stem cells arising from autodiploidization during haploid gene trap mutagenesis. *Nucleic Acids Research*, *46*(10), e63–e63. https://doi.org/10.1093/nar/gky183

Yang, A., Wang, H., & Yang, X. (2017). Long non-coding RNA PVT1 indicates a poor prognosis of glioma and promotes cell proliferation and invasion via target EZH2. *Bioscience Reports*, *37*(6). https://doi.org/10.1042/BSR20170871

Yang, L., Peng, X., Jin, H., & Liu, J. (2019). Long non-coding RNA PVT1 promotes autophagy as ceRNA to target ATG3 by sponging microRNA-365 in hepatocellular carcinoma. *Gene*, *697*, 94–102. https://doi.org/10.1016/j.gene.2019.02.036

Yang, T., Zhou, H., Liu, P., Yan, L., Yao, W., Chen, K., Zeng, J., Li, H., Hu, J., Xu, H., & Ye, Z. (2017). lncRNA PVT1 and its splicing variant function as competing endogenous RNA to regulate clear cell renal cell carcinoma progression. *Oncotarget*, *8*(49), 85353–85367. https://doi.org/10.18632/oncotarget.19743

Yin, Y., Lu, J. Y., Zhang, X., Shao, W., Xu, Y., Li, P., Hong, Y., Cui, L., Shan, G., Tian, B., Zhang, Q. C., & Shen, X. (2020). U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature*, *580*(7801), 147–150. https://doi.org/10.1038/s41586-020-2105-3

Yu, L., Gong, X., Sun, L., Zhou, Q., Lu, B., & Zhu, L. (2016). The Circular RNA Cdr1as Act as an Oncogene in Hepatocellular Carcinoma through Targeting miR-7 Expression. *PLOS ONE*, *11*(7), e0158347. https://doi.org/10.1371/journal.pone.0158347

Zampetaki, A., Albrecht, A., & Steinhofel, K. (2018). Long Non-coding RNA Structure and Function: Is There a Link? *Frontiers in Physiology*, *9*. https://doi.org/10.3389/fphys.2018.01201

Zeng, Y., Du, W. W., Wu, Y., Yang, Z., Awan, F. M., Li, X., Yang, W., Zhang, C., Yang, Q., Yee, A., Chen, Y., Yang, F., Sun, H., Huang, R., Yee, A. J., Li, R.-K., Wu, Z., Backx, P. H., & Yang, B. B. (2017). A Circular RNA Binds To and Activates AKT Phosphorylation and Nuclear Localization Reducing Apoptosis and Enhancing Cardiac Repair. *Theranostics*, *7*(16), 3842–3855. https://doi.org/10.7150/thno.19764

Zhang, C., Liu, J., Wu, R., Liang, Y., Lin, M., Liu, J., Chan, C. S., Hu, W., & Feng, Z. (2014). Tumor suppressor p53 negatively regulates glycolysis stimulated by hypoxia through its target RRAD. *Oncotarget*, *5*(14), 5535–5546. https://doi.org/10.18632/oncotarget.2137

Zhang, M., Zhao, K., Xu, X., Yang, Y., Yan, S., Wei, P., Liu, H., Xu, J., Xiao, F., Zhou, H., Yang, X., Huang, N., Liu, J., He, K., Xie, K., Zhang, G., Huang, S., & Zhang, N. (2018). A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nature Communications*, *9*(1), 4475. https://doi.org/10.1038/s41467-018-06862-2

Zhang, S., Zhang, G., & Liu, J. (2016). Long noncoding RNA PVT1 promotes cervical cancer progression through epigenetically silencing miR-200b. *APMIS*, *124*(8), 649–658. https://doi.org/10.1111/apm.12555

Zhang, X., Rice, K., Wang, Y., Chen, W., Zhong, Y., Nakayama, Y., Zhou, Y., & Klibanski, A. (2010). Maternally Expressed Gene 3 (MEG3) Noncoding Ribonucleic Acid: Isoform Structure, Expression, and Functions. *Endocrinology*, *151*(3), 939–947. https://doi.org/10.1210/en.2009-0657

Zhang, Y., Chen, M., & Chen, C. (2021). Using the Zebrafish as a Genetic Model to Study Erythropoiesis. *International Journal of Molecular Sciences*, *22*(19), 10475. https://doi.org/10.3390/ijms221910475

Zhang, Y., Dang, Y.-W., Wang, X., Yang, X., Zhang, R., Lv, Z.-L., & Chen, G. (2017). Comprehensive analysis of long non-coding RNA PVT1 gene interaction regulatory network in hepatocellular carcinoma using gene microarray and bioinformatics. In *Am J Transl Res* (Vol. 9, Issue 9). http://kobas.cbi.pku.edu.cn/

Zhang, Y., Vastenhouw, N. L., Feng, J., Fu, K., Wang, C., Ge, Y., Pauli, A., van Hummelen, P., Schier, A. F., & Liu, X. S. (2014). Canonical nucleosome organization at promoters forms during genome activation. *Genome Research*, *24*(2), 260–266. https://doi.org/10.1101/gr.157750.113

Zhang, Y., Zhang, X.-O., Chen, T., Xiang, J.-F., Yin, Q.-F., Xing, Y.-H., Zhu, S., Yang, L., & Chen, L.-L. (2013). Circular Intronic Long Noncoding RNAs. *Molecular Cell*, *51*(6), 792–806. https://doi.org/10.1016/j.molcel.2013.08.017

Zhao, J., Du, P., Cui, P., Qin, Y., Hu, C., Wu, J., Zhou, Z., Zhang, W., Qin, L., & Huang, G. (2018). LncRNA PVT1 promotes angiogenesis via activating the STAT3/VEGFA axis in gastric cancer. *Oncogene*, *37*(30), 4094–4109. https://doi.org/10.1038/s41388-018-0250-z

Zheng, J., Yu, F., Dong, P., Wu, L., Zhang, Y., Hu, Y., & Zheng, L. (2016). Long non-coding RNA PVT1 activates hepatic stellate cells through competitively binding microRNA-152. *Oncotarget*, *7*(39), 62886–62897. https://doi.org/10.18632/oncotarget.11709

Zheng, Q., Bao, C., Guo, W., Li, S., Chen, J., Chen, B., Luo, Y., Lyu, D., Li, Y., Shi, G., Liang, L., Gu, J., He, X., & Huang, S. (2016). Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nature Communications*, *7*(1), 11215. https://doi.org/10.1038/ncomms11215

Zhou, Q., Chen, J., Feng, J., & Wang, J. (2016). Long noncoding RNA PVT1 modulates thyroid cancer cell proliferation by recruiting EZH2 and regulating thyroid-stimulating hormone receptor (TSHR). *Tumor Biology*, *37*(3), 3105–3113. https://doi.org/10.1007/s13277-015-4149-9

Ziv, O., Farberov, S., Lau, J. Y., Miska, E., Kudla, G., & Ulitsky, I. (2021). Structural features within the NORAD long noncoding RNA underlie efficient repression of Pumilio activity. *BioRxiv*. https://doi.org/10.1101/2021.11.19.469243

Ziv, O., Gabryelska, M. M., Lun, A. T. L., Gebert, L. F. R., Sheu-Gruttadauria, J., Meredith, L. W., Liu, Z.-Y., Kwok, C. K., Qin, C.-F., MacRae, I. J., Goodfellow, I., Marioni, J. C., Kudla, G., & Miska, E. A. (2018). COMRADES determines in vivo RNA structures and interactions. *Nature Methods*, *15*(10), 785–788. https://doi.org/10.1038/s41592-018-0121-0

ANNEX

© 2021. Published by The Company of Biologists Ltd | Development (2021) 148, dev182741. doi:10.1242/dev.182741

REVIEW

lncRNAs in development and differentiation: from sequence motifs to functional characterization

Florian Constanty and Alena Shkumatava*

ABSTRACT

The number of long noncoding RNAs (lncRNAs) with characterized developmental and cellular functions continues to increase, but our understanding of the molecular mechanisms underlying lncRNA functions, and how they are dictated by RNA sequences, remains limited. Relatively short, conserved sequence motifs embedded in lncRNA transcripts are often important determinants of lncRNA localization, stability and interactions. Identifying such RNA motifs remains challenging due to the substantial length of lncRNA transcripts and the rapid evolutionary turnover of lncRNA sequences. Nevertheless, the recent discovery of specific RNA elements, together with their experimental interrogation, has enabled the first step in classifying heterogeneous lncRNAs into sub-groups with similar molecular mechanisms and functions. In this Review, we focus on lncRNAs with roles in development, cell differentiation and normal physiology in vertebrates, and we discuss the sequence elements defining their functions. We also summarize progress on the discovery of regulatory RNA sequence elements, as well as their molecular functions and interaction partners.

KEY WORDS: RNA regulatory motifs, RNA-protein complexes, Cell differentiation, Development, In vivo functions, LncRNAs

Introduction

Thousands of genomic loci are transcribed into so-called long noncoding RNAs (lncRNAs) (Rinn and Chang, 2012). A fraction of these has been demonstrated to regulate a range of important biological processes, such as development and cell differentiation (Mallory and Shkumatava, 2015; Perry and Ulitsky, 2016; Ulitsky and Bartel, 2013). LncRNAs have also been linked to human diseases, including nearly every major cancer type as well as various neurological disorders (Bian and Sun, 2011; Esposito et al., 2019; Gutschner and Diederichs, 2012; Huarte, 2015; Li et al., 2018). Although the annotation of lncRNAs and the identification of their biological functions has progressed over the past decade, lncRNAs are still poorly understood at the molecular level, partially due to the diversity of their modes of action and a lack of suitable molecular tools to dissect them.

Three sub-groups of lncRNAs can be distinguished based on their molecular mechanisms of action: (1) lncRNA loci that contain enhancers regulating gene expression; (2) lncRNA loci for which the act of transcription but not the transcript per se is important for regulating neighbouring genes; and (3) lncRNA loci that carry out their cellular functions via lncRNA transcripts that interact with DNA, other RNAs and proteins (Marchese et al., 2017; Wang and Chang, 2011). Notably, some lncRNAs may belong to more than one of these subgroups. The lncRNAs in the final group share biogenesis pathways and post-transcriptional features with mRNAs, i.e. they are transcribed by RNA polymerase II and are capped, spliced and polyadenylated (Cabili et al., 2011; Guttman and Rinn, 2012). For these lncRNAs, the linear transcript sequence often defines function. Comparative analyses of lncRNAs in different vertebrate species has shown that lncRNA sequences undergo rapid evolutionarily turnover, resulting in their overall poor conservation (Hezroni et al., 2015; Kutter et al., 2012; Necsulea et al., 2014; Ulitsky et al., 2011; Washietl et al., 2014). Nonetheless, signatures of primary sequence conservation have been detected in ∼100 lncRNAs conserved from mammals to fish and in ∼1000 lncRNAs conserved among mammals (Hezroni et al., 2015). For the majority of conserved lncRNAs, only short stretches of sequence are retained (Hezroni et al., 2015; Ulitsky et al., 2011). These short sequences may represent functional elements of lncRNA transcripts that facilitate interactions with other RNAs, proteins or genomic loci.

In addition to their linear RNA sequences, the secondary and tertiary structures of lncRNA transcripts have been proposed to be functional determinants (Diederichs, 2014; Guttman and Rinn, 2012; Wutz et al., 2002). Several functionally conserved RNA structure examples support this idea (Holmes et al., 2020; Uroda et al., 2019; Zhang et al., 2010). Moreover, recent development of new technologies to determine RNA structures *in vivo* have demonstrated that RNA structures can be engaged in important long-range interactions (Lu et al., 2016; Ziv et al., 2018). However, it is unclear how prevalent the impact of RNA structure on lncRNA functionality is (Rivas et al., 2017). Furthermore, the unbiased analysis of sequence and context preferences for multiple human RNA-binding proteins (RBPs), which are important drivers of lncRNA function, has demonstrated that RBPs bind short, often low complexity, RNA motifs (Dominguez et al., 2018). RBP-binding specificity is further supported by additional features, such as local RNA secondary structure and flanking nucleotides (Dominguez et al., 2018), suggesting that the functional elements controlling lncRNAs reside in their sequence and adjacent secondary structure motifs, rather than in their global tertiary structure.

In this Review, we discuss a set of vertebrate lncRNAs with characterized developmental, cell biological and physiological functions defined by specific lncRNA sequence elements (Table 1). We discuss how such sequence elements and motifs regulate the biogenesis, localization and interactions of lncRNAs. In addition, we survey recent efforts to experimentally identify functional lncRNA motifs and highlight computational algorithms that have been developed to predict these functions based on lncRNA sequences.

The regulation of lncRNA biogenesis and processing by RNA motifs

To generate functional noncoding transcripts, tight regulation of lncRNA biogenesis is required. Although the majority of mature

Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, Paris 75005, France.

*Author for correspondence (alena.shkumatava@curie.fr)

A.S., 0000-0001-9286-4700

Development (2021) 148, dev182741. doi:10.1242/dev.182741

Table 1. Examples of lncRNA motifs with regulatory functions in development, cell differentiation and animal physiology

lncRNA	RNA motif(s)	Function	References
NEAT1_2	Two U-rich and one A-rich motifs	3′ END alternative RNAse P processing	Brown et al. (2012); Sunwoo et al. (2009); Wilusz et al. (2012)
MALAT1	Two U-rich and one A-rich motifs	3′ END alternative RNAse P processing	Brown et al. (2012); Wilusz et al. (2008, 2012)
Firre	Local repeating RNA domains	Nuclear retention signal bound by hnRNPU	Hacisuleyman et al. (2014, 2016)
MALAT1	RNSP1 binding sites	Paraspeckle localization	Miyagawa et al. (2012)
MEG3	Nuclear retention element	Nuclear retention signal involving U1 snRNP	Azam et al. (2019)
BORG	AGCCC motifs	Nuclear retention signal	Zhang et al. (2014)
CTN-RNA	A-to-I editing site	I-edited RNA as a nuclear retention signal bound by paraspeckle $p54^{nrb}$ protein	Prasanth et al. (2005)
libra/NREP	Near-perfect miR-29 binding site	Regulation of miR-29b levels by TDMD	Bitetti et al. (2018)
Cyrano	Near-perfect miR-7 binding site	Regulation of miR-7 levels by TDMD	Kleaveland et al. (2018); Ulitsky et al. (2011)
NORAD	UGURUAUA	Sequestration of PUM proteins	Lee et al. (2016); Tichon et al. (2016)
Megamind	A conserved region of ∼200bp	RNP complex formation	Lin et al. (2014); Ulitsky et al. (2011)
Braveheart	G-rich motif (AGIL)	Interaction with the transcription factor CNBP	Xue et al. (2016)
GAS5	GR mimic-binding site	Decoy of the GRs	Hudson et al. (2014)
PINT	Conserved PRC2 binding site	Downregulation of pro-invasion gene expression	Marin-Bejar et al. (2013); Marín-Béjar et al. (2017); Sauvageau et al. (2013)

lncRNA transcripts are generated by the same biogenesis pathways as those regulating mRNAs, two particular lncRNA transcripts – *MALAT1* (*metastasis-associated lung adenocarcinoma transcript 1*) and *NEAT1* (*nuclear enriched abundant transcript 1*) – are processed in a manner that involves conserved motifs (Wilusz et al., 2012) (Fig. 1A).

Both *MALAT1* and *NEAT1_2* (the long isoform of *NEAT1*) are structural components of nuclear bodies, with *MALAT1* localizing to speckles and *NEAT1_2* to paraspeckles (Clemson et al., 2009; Hutchinson et al., 2007; Nakagawa et al., 2011; Sasaki et al., 2009; Sunwoo et al., 2009). Whereas *NEAT1_2* is essential for paraspeckle formation (Clemson et al., 2009; Mao et al., 2011; Naganuma and Hirose, 2013; Naganuma et al., 2012; Sasaki et al., 2009), depletion of *MALAT1* does not noticeably affect speckles (Nakagawa et al., 2012; Zhang et al., 2012). At the organismal level, *Neat1* is required for normal mammary gland development, lactation and corpus luteum formation during the establishment of pregnancy (Nakagawa et al., 2014; Standaert et al., 2014). By contrast, *Malat1* does not appear to be required for normal development, as mouse and zebrafish null alleles of *Malat1* do not exhibit any apparent morphological defects (Eissmann et al., 2012; Lavalou et al., 2019; Nakagawa et al., 2012; Zhang et al., 2012). However, in-depth analysis of one of the mouse null alleles has shown retinal vascularization defects (Michalik et al., 2014). In addition, there is growing evidence that *Malat1* plays an essential role in disease, affecting cancer progression and metastasis (Arun and Spector, 2019).

During their biogenesis, *MALAT1* and *NEAT1* transcripts are processed by an unusual RNase P cleavage step directed by structured tRNA-like motifs (Wilusz et al., 2012). Upon RNase P cleavage, *MALAT1*, which is conserved from fish to human (Ulitsky et al., 2011), is stabilized by a highly conserved triple helical structure located immediately upstream of the RNase P cleavage site; this event prevents exonucleolytic degradation from the 3′ end of the transcript (Brown et al., 2012; Wilusz et al., 2012) (Fig. 1A). The mammalian-specific *NEAT1* locus encodes two lncRNA isoforms generated from the same promoter. Although the short isoform, termed *NEAT1_1* (3.7 kb in humans), is produced by canonical cleavage and polyadenylation using an upstream polyadenylation signal, the long isoform, termed *NEAT1_2* (23 kb in humans), is processed by RNase P cleavage and stabilized by a triple helical structure, similar to that regulating *MALAT1* biogenesis (Fig. 1A). In both cases, the 3′-end stabilizing triple helical structure is generated by conserved motifs of ∼10 nucleotides (nt) in length – two U-rich motifs and an A-rich tract (Brown et al., 2014, 2012; Wilusz et al., 2008, 2012) (Fig. 1A). This biogenesis mechanism remains a unique feature of *MALAT1* and *NEAT1*, as no additional lncRNA transcripts processed by RNase P cleavage and harbouring similar triple-helix forming motifs have been identified so far.

Motifs and sequences that control the subcellular localization of lncRNA transcripts

The subcellular localization of lncRNAs is also tightly regulated and is an important determinant of lncRNA function (Chen, 2016; Guo et al., 2020). In the nucleus, lncRNAs regulate chromatin organization, transcription, RNA maturation and nuclear protein activity (Sun et al., 2018; Wang and Chang, 2011). In the cytoplasm, by contrast, lncRNAs interact with microRNAs (miRNAs) and regulate translation efficiency and cytoplasmic protein activity (Noh et al., 2018; Paraskevopoulou and Hatzigeorgiou, 2016). Notably, lncRNAs tend to be less enriched in the cytoplasm compared with mRNAs (Derrien et al., 2012; Mukherjee et al., 2017). Specifically, for lncRNAs whose functions are linked to activities in the nucleus, repression of their export to the cytoplasm is required. Many nuclear lncRNAs have been studied in detail and specific RNA elements that facilitate their binding to proteins defining nuclear localization have been determined. Below, we highlight examples that illustrate the variety of different RNA elements and molecular mechanisms directing the nuclear enrichment of lncRNAs.

The nuclear localization of *MALAT1* can be explained in part by the single-exon structure of its transcript, which does not engage with exon junction complexes promoting nuclear export (Hutchinson et al., 2007). In addition, two distinct sequence elements have been found to promote the specific enrichment of *MALAT1* in nuclear speckles (Miyagawa et al., 2012). These nuclear retention sequences of ∼600 nt are bound *in vitro* by the nuclear speckle protein RNSP1. Depletion of RNSP1 and two other key nuclear speckle proteins, SRM160 and IBP160, leads to diffusion of *MALAT1* into the nucleoplasm, suggesting that these proteins may drive the nuclear retention of *MALAT1*. However, sequence and structure conservation analyses did not detect any common motifs within the two *MALAT1* nuclear localization elements (Miyagawa et al., 2012).

In contrast to *MALAT1*, the lncRNA *Firre* (*functional intergenic repeating RNA element*) is a spliced and polyadenylated nuclear transcript, the knockout of which leads to cell-specific defects in

Development (2021) 148, dev182741. doi:10.1242/dev.182741

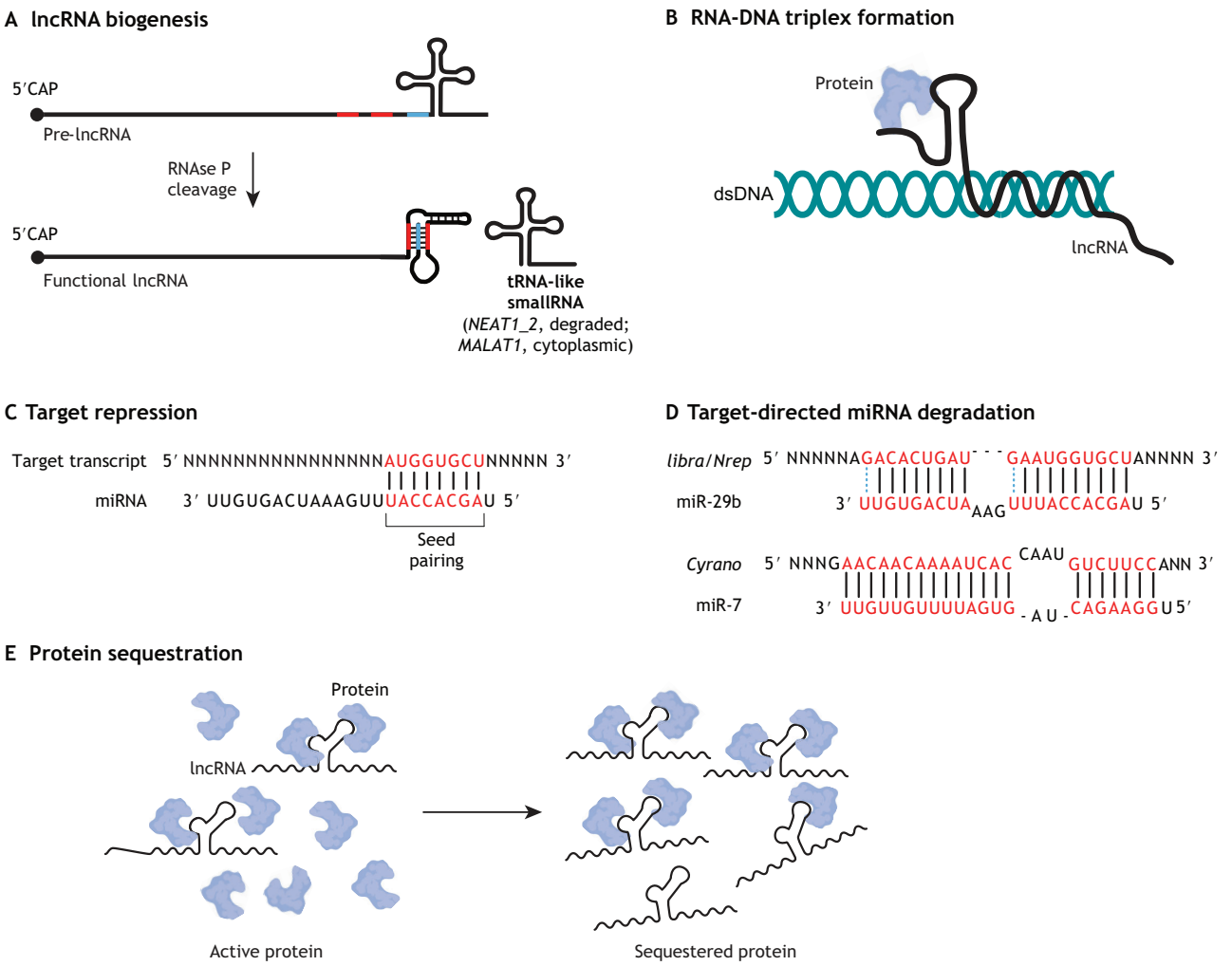


Fig. 1. Functions of conserved lncRNA motifs. (A) Two U-rich motifs (red) and one A-rich motif (blue) are essential for the RNAse P-mediated cleavage and processing of the lncRNAs *NEAT1_2* and *MALAT1* (black). This event is also directed by structured tRNA-like motifs. (B) lncRNAs (black) can also act as guides during the formation of RNA-DNA triplex structures that interact with proteins (blue). (C) Canonical animal miRNA-binding sites (with partial miRNA-target RNA pairing complementarity) can regulate post-transcriptional target repression. (D) Nearly perfect miRNA-binding sites regulate miRNA decay through target-directed miRNA degradation (TDMD). The vertical dotted lines represent wobble pairing. (E) lncRNAs (black) can also sequester proteins (blue) to inhibit their functionality.

hematopoietic populations in mice (Lewandowski et al., 2019). Although *Firre* is syntenically conserved in almost all eutherians (Hezroni et al., 2015), its sequence comprising multiple repetitive elements of ∼150 nt, termed repeating RNA domains (RRDs), has diverged between primate and rodent lineages (Hacisuleyman et al., 2016). Despite this divergence in sequence evolution resulting in limited sequence similarity between primate and rodent RRD elements, orthologous RRDs share the same function and act as strong nuclear retention signals (Hacisuleyman et al., 2014, 2016). RRDs drive nuclear retention of the *Firre* transcript and are sufficient to localize an otherwise cytoplasmic mRNA to the nucleus when added to the 3′ end of this mRNA (Hacisuleyman et al., 2016). Mechanistically, it has been proposed that RRD elements interact with the nuclear matrix protein HNRNPU, which insures correct localization of RNAs to the nucleus (Hacisuleyman et al., 2014, 2016).

The nuclear retention of the lncRNA *MEG3* is also regulated by a linear sequence motif of 356 nt (Azam et al., 2019). Deletion of this nuclear retention element (NRE), which is bound by several components of U1 snRNP, effectively relocates *MEG3* reporter transcripts from the nucleus to the cytoplasm (Azam et al., 2019). By contrast, the function of *MEG3*, which acts as a tumour suppressor that stimulates the p53 pathway and controls the cell cycle, is mediated by a tertiary structure that is conserved in mammals (Uroda et al., 2019).

The spliced and polyadenylated lncRNA *BORG* regulates BMP-induced differentiation of C2C12 cells into osteoblastic cells and is exclusively localized to the nucleus (Takeda et al., 1998). The nuclear localization of *BORG* is mediated by a short AGCCC RNA motif. Although the mechanism of nuclear retention controlled by this RNA motif remains unknown, the motif was found to be present in multiple other nuclear lncRNAs as well as in protein-coding RNAs, indicating that nuclear localization motifs are shared between noncoding and coding transcripts (Zhang et al., 2014).

The 3′UTR of *CTN-RNA*, a polyadenylated transcript that localizes to paraspeckles under regular cellular conditions,

Development (2021) 148, dev182741. doi:10.1242/dev.182741

harbours another type of a NRE (Prasanth et al., 2005). This NRE consists of a ∼100 nt forward repeat and three inverted repeats originating from SINE retro-elements. Each of the inverted repeats can form a stem loop with the forward repeat. The stem loop serves as a target site for ADAR enzymes to catalyse adenosine-to-inosine (A-to-I) RNA editing (Prasanth et al., 2005). It has been proposed that the paraspeckle protein p54^{nrb} interacts with *CTN-RNA* at I-edited residues, triggering retention of the transcript in the nucleus. Under cellular stress conditions, *CTN-RNA* is released from the nucleus by cleavage of its edited 3′UTR, generating an mRNA that contains a 5′UTR and a coding region. This cleaved transcript is then transported to the cytoplasm, where it is translated into the mCAT2 protein that modulates cellular uptake of L-arginine (Prasanth et al., 2005), demonstrating that tight regulation of *CTN-RNA* subcellular localization is essential during homeostatic and stress conditions.

From the few examples presented here, it is evident that the repertoire of RNA nuclear retention motifs and their associated interactors far exceeds the number of known protein nuclear localization signal motifs, presenting a major challenge in their identification. It remains to be seen whether individual transcripts adapted various nuclear retention mechanisms or whether common sets of lncRNA sequence elements execute their retention in the nucleus. High-throughput approaches (discussed below) have the potential to find common sets of motifs with similar functions and group them into mechanistically defined sub-classes.

Unbiased identification of lncRNA subcellular localization motifs

Identification of the aforementioned subcellular localization elements has been restricted to individual lncRNA and mRNA transcripts, and has been achieved using reporter constructs and classical genetics. However, with recent advances in comparative genomics, data analysis and systematic experimental approaches, several computational and high-throughput experimental strategies have been applied to unbiasedly decipher sequence elements dictating RNA localization.

To identify RNA motifs that drive nuclear localization, several independent studies have used a tiling-based strategy. In this approach, pools of thousands of tiled oligonucleotides covering human lncRNAs and selected 3′UTRs are evaluated for their ability to retain an otherwise cytoplasmic reporter transcript in the nucleus. The Ulitsky laboratory constructed a library of ∼5500 109-mers that tiled exons of 37 human lncRNAs and 13 3′UTRs of mRNAs enriched in the nucleus. The library was cloned into the 5′ and 3′ UTRs of a cytoplasmic GFP reporter and transfected into human cells; RNA from cytoplasmic and nuclear fractions was then sequenced and analysed for enrichment of sequence motifs. A 42 nt C-rich sequence element derived from an Alu-repeat named SIRLOIN was found to drive nuclear localization of a set of lncRNAs and mRNAs by interacting with the protein HNRNPK (Lubelsky and Ulitsky, 2018). In parallel, the Rinn laboratory used a similar tiling array-reporter method to identify *cis* elements driving nuclear retention (Shukla et al., 2018). The authors generated a pool of ∼12,000 110-mers tightly tiled across 38 lncRNAs with different subcellular localization patterns. Similarly, the oligonucleotides were fused to a cytoplasmic reporter transcript and the reporter-oligonucleotide pool was transfected into human cells, which were subjected to subcellular fractionation and RNA-seq (Shukla et al., 2018). This study identified 109 RNA elements, including a 15 nt C-rich motif, that enriched the cytoplasmic reporter in the nucleus. Interestingly, the identified C-rich motif shows sequence similarities to the SIRLOIN motif (Lubelsky and Ulitsky, 2018; Shukla et al., 2018). However, single-molecule *in situ* hybridization analyses showed that this short C-rich region alone is not sufficient to drive nuclear localization and requires a larger RNA element to retain the reporter transcript in the nucleus (Shukla et al., 2018). Furthermore, the Shen team used a tiling-based approach to search for lncRNA sequence elements that can mediate chromatin enrichment and identified a short 7 nt motif that is essential for reporter RNA localization to chromatin. It has been proposed that this motif is recognized by U1 snRNP to tether lncRNAs to chromatin (Yin et al., 2020).

In addition to the systematic experimental approaches described above, several computational approaches have been developed to evaluate the subcellular localization of lncRNAs based on their sequences. One approach generated a map of transposable element (TE) insertions in lncRNAs and found that a set of evolutionarily conserved TEs functionalized to nuclear retention elements (Carlevaro-Fita et al., 2019). Specifically, this study identified four TEs that significantly correlate with lncRNA nuclear enrichment, as well as a set of GC-rich TEs that correlates with lncRNA cytoplasmic localization (Carlevaro-Fita et al., 2019). In addition, efficient splicing has been found to be one of the main predictive factors for lncRNA cytoplasmic localization, whereas inefficient splicing strongly correlates with nuclear enrichment (Zuckerman and Ulitsky, 2019). Moreover, by training computational algorithms to compare sequences of lncRNAs with known localization and detecting a link between specific sequence motifs and subcellular localization, one can predict the subcellular localization of lncRNA transcripts (Cao et al., 2018; Su et al., 2018). Another approach used nuclear and cytoplasmic fractionated RNA-seq data to develop a deep learning algorithm that predicts lncRNA subcellular localization based on lncRNA sequences (Gudenas and Wang, 2018). Taken together, the ability to computationally predict and experimentally identify functional sequence elements that drive lncRNA subcellular localization is an important step towards understanding lncRNA functions.

Sequences that control the formation of lncRNA-DNA triplexes

Nuclear pyrimidine (T and C)-rich RNAs can interact with purine (A and G)-rich DNA via Hoogsteen-type hydrogen bonding to form RNA:DNA triple helices that may have regulatory functions (Fig. 1B) (Felsenfeld and Rich, 1957). As triplex formation requires pairing between double-stranded DNA and single-stranded RNA, the formation of the triplexes depends on the lncRNA primary sequence (Li et al., 2016). *Fendrr* is one example of a nuclear lncRNA that forms RNA:DNA triple helices and plays a regulatory role during development. Loss of function of *Fendrr*, which is exclusively expressed in the lateral plate mesoderm during mouse embryogenesis, leads to embryonic lethality due to heart and body wall developmental defects (Grote et al., 2013; Sauvageau et al., 2013). *Fendrr* interacts with both PRC2 (polycomb repressive complex 2) and TrxG/MLL complexes *in vivo*, and its loss results in changes in epigenetic modifications and gene expression of mesoderm differentiation factors (Grote et al., 2013). A 40 nt sequence element of *Fendrr* RNA was identified *in silico* and validated experimentally to form triplexes with promoter regions of two targets of *Fendrr*, *Foxf1* and *Pitx2*, resulting in increased PRC2 occupancy at these promoters and a subsequent inhibition of expression (Grote et al., 2013).

Although reports of DNA:lncRNA triplexes and their functional relevance have been restricted to specific examples, computational

Development (2021) 148, dev182741. doi:10.1242/dev.182741

methods have been developed to predict the formation of lncRNA-DNA triplex structures based on lncRNA sequence *in silico* (Buske et al., 2012; Kuo et al., 2019). One algorithm, called TDF (triplex domain finder), enables retrieval and statistical ranking of sequence elements already known to form triplex structures, including characterized domains of the lncRNAs *Fendrr*, *HOTAIR* and *MEG3* (Kuo et al., 2019). Importantly, the TDF algorithm has also identified and ranked novel sequence elements with triplex-forming potential and detected previously unknown DNA-binding domains in *MEG3* and in lncRNAs relevant to human cardiomyocyte differentiation (Kuo et al., 2019).

RNA motifs that mediate miRNA degradation

While RNA sequence elements of nuclear lncRNA transcripts have received a lot of attention, RNA motifs in cytoplasmic lncRNAs that have defined molecular mechanisms linked to cellular or developmental functions are as yet under-represented. One of the most studied interaction partners of cytoplasmic lncRNAs are miRNAs, which are small RNAs of ∼22 nt that regulate virtually every aspect of development and normal physiology by pairing with their targets (Alberti and Cochella, 2017; Bartel, 2018). Because one can easily predict miRNA interactions based on the presence of miRNA-binding sites in their target transcripts, miRNA-lncRNA interactions have been extensively investigated (Ulitsky, 2018). The majority of miRNA binding sites within vertebrate lncRNA transcripts are canonical miRNA sites with partial miRNA-lncRNA pairing complementarity (Fig. 1C), which can regulate lncRNA expression similar to miRNA-mediated regulation of mRNA transcripts. It has also been suggested that canonical miRNA target sites within lncRNAs can regulate miRNA activity via competition for binding sites (Ulitsky, 2018).

In addition to the canonical miRNA-binding sites with partial complementarity, several independent studies have found lncRNAs that harbour miRNA-binding sites with extensive, near-perfect complementarity (Fig. 1D), which is rather unusual for animal miRNAs and has not been reported in endogenous transcripts before (Bitetti et al., 2018; Kleaveland et al., 2018; Ulitsky et al., 2011). By using these high-complementarity miRNA sites in artificially engineered targets and natural viral RNAs, it has been demonstrated that the extensive target-miRNA pairing triggers efficient degradation of miRNAs by shortening (‘trimming’) or untemplated lengthening ‘tailing’) the 3′ end of miRNAs, a phenomenon known as target-directed miRNA degradation, or TDMD (Ameres et al., 2010; de la Mata et al., 2015; Marcinowski et al., 2012).

More recently, an endogenous transcript that harbours a RNA sequence element conserved among all vertebrates that binds and specifically degrades miR-29b was discovered (Bitetti et al., 2018). The conserved RNA sequence element is part of the cytoplasmic lncRNA *libra* in zebrafish and the 3′ UTR of the protein-coding gene *Nrep* in mammals. Genetic scrambling of the miR-29 site resulted in ectopic accumulation of miR-29b in the brain, leading to abnormal animal behaviour, thus demonstrating that the miRNA sites embedded in genome-encoded lncRNAs have crucial *in vivo* relevance (Bitetti et al., 2018). A similar activity of the endogenous high complementarity miR-7 site (Ulitsky et al., 2011) was reported for the lncRNA *Cyrano* (Kleaveland et al., 2018). The miR-7 site embedded within the *Cyrano* transcript is conserved among all vertebrates and triggers efficient miR-7 degradation *in vivo* in mouse tissues and in human cells (Kleaveland et al., 2018). The biological relevance of miR-7 degradation by TDMD remains to be addressed, as genetic perturbations of *Cyrano* do not lead to obvious morphological defects in mice or zebrafish (Goudarzi et al., 2019; Kleaveland et al., 2018; Lavalou et al., 2019), in contrast to previous morpholino-based knockdown zebrafish studies (Sarangdhar et al., 2018; Ulitsky et al., 2011). Nonetheless, it is possible that additional lncRNAs contain high complementarity miRNA sites that control miRNA turnover and potentially have key biological functions. This is further supported by another finding of a high-complementarity miRNA site for miR-30 located in the 3′UTR of the *Serpine1* (Ghini et al., 2018). Similar to *libra*/*NREP* and *Cyrano*, the endogenously encoded miR-30 site reduces levels of miR-30b-p5 and 30c-p5 by TDMD, regulating mitotic rate and apoptosis in mouse fibroblasts (Ghini et al., 2018).

LncRNA motifs that sequester cellular proteins and regulate normal physiology

Although multiple lncRNAs have been shown to have regulatory functions, detailed structure-function analyses have been performed only for a few of them. One of the striking examples of a lncRNA for which specific RNA sequences have been assigned a function is *NORAD* (*noncoding RNA activated by DNA damage*). *NORAD* is an abundant cytoplasmic transcript present at 80-1400 copies per cell that shows sequence conservation across mammals and is upregulated upon DNA damage (Lee et al., 2016; Tichon et al., 2016). Remarkably, *NORAD* contains at least 17 consensus binding sites for PUMILIO proteins (PUM1 and PUM2), which are highly conserved RBPs that bind to consensus motifs typically located in the 3′ untranslated regions (UTRs) of mRNAs, resulting in reduced translation and turnover of mRNA targets (Miller and Olivas, 2011). *NORAD*, with its multiple PUMILIO binding sites and high copy number, acts as a negative regulator of PUMILIO proteins by limiting their amount in the cell (Fig. 1E). NORAD-mediated limitation of PUMILIO modulates the abundance of PUMILIO targets, which are enriched for mRNAs with mitotic, DNA repair and DNA replication functions (Lee et al., 2016; Tichon et al., 2016). As a result, *NORAD* loss of function leads to excess repression of PUMILIO targets, resulting in chromosomal instability, aberrant mitosis and aneuploidy in human cells (Lee et al., 2016; Tichon et al., 2016). In mice, *Norad* deletion leads to genomic instability and mitochondrial dysfunction, with animals showing signs of premature aging; *Norad* null allele mice were found to have grey and thin fur, to show abnormal spine curvature characteristic of aging and to die earlier than wild-type mice (Kopp et al., 2019). While *NORAD* interacts with additional proteins (Munschauer et al., 2018; Tichon et al., 2018), a series of genetic experiments, including PUMILIO 2 overexpression in mice and *NORAD* expression in human cells deficient for *NORAD*, demonstrated that the aforementioned cellular and organismal phenotypes are linked to PUMILIO hyperactivity (Elguindy et al., 2019; Kopp et al., 2019; Lee et al., 2016; Tichon et al., 2016). This example demonstrates the importance of clustered RNA motifs/elements in establishing functional lncRNA-protein interactions that play key roles in cell biology and normal physiology.

LncRNA elements that mediate lncRNA-protein interactions

Interactions between lncRNAs and RBPs can lead to different molecular, cellular and developmental outcomes. Indeed, lncRNAs can interact with proteins to inhibit or enhance transcription (Feng, 2006; Wutz et al., 2002), stabilize proteins (Zhao et al., 2018) or sequester them (Kino et al., 2010; Lee et al., 2016; Tichon et al., 2016), regulate translation efficiency (Gong et al., 2015), and/or regulate scaffold protein complexes (Rinn et al., 2007). Because RBPs tend to bind low-complexity sequences (Dominguez et al., 2018), it is often difficult to identify protein consensus binding

Development (2021) 148, dev182741. doi:10.1242/dev.182741

motifs within RNA transcripts; however, it is possible to attribute protein binding to specific lncRNA regions. Indeed, it has been shown that, while lncRNA transcripts can interact with multiple RBPs, often only a short RNA region, embedded in the long transcript and interacting with a specific RBP or with a specific set of RBPs, is sufficient and required to drive the biological function of a number of lncRNAs.

One of the best characterized lncRNAs with a crucial developmental function is *Xist* (*X-inactive specific transcript*), which orchestrates transcriptional silencing of one of the female X chromosomes in placental mammals (Borsani et al., 1991; Brown, 1991; Moindrot and Brockdorff, 2016), a process known as X-chromosome inactivation (XCI) (Gendrel and Heard, 2014; Marahrens et al., 1997). *Xist* knockout leads to loss of X-inactivation and female-specific lethality in mice (Marahrens et al., 1997). One of the conserved regions of *Xist*, named the A-repeat, is composed of repetitive sequence elements that are believed to be derived from TEs (Elisaphenko et al., 2008) and is required for the initiation of XCI (Nesterova et al., 2001; Wutz et al., 2002). The A-repeat interacts with multiple RBPs (Chu et al., 2015b; Graindorge et al., 2019; Lu et al., 2016; McHugh et al., 2015; Minajigi et al., 2015; Moindrot et al., 2015; Monfort et al., 2015), among which SPEN (also known as SHARP in human and MINT in mouse) is essential for initiating XCI in mouse embryonic stem cells (ESCs) and preimplantation embryos (Carter et al., 2020; Dossin et al., 2020). By specifically binding the *Xist* A-repeat region, SPEN recruits additional silencing factors to initiate XCI, bridging *Xist* RNA, transcription machinery and chromatin remodellers (Chu et al., 2015b; Dossin et al., 2020; McHugh et al., 2015; Minajigi et al., 2015). Interestingly, in mouse ESCs, Spen binds and silences endogenous retroviral element (ERV) RNAs that resemble the A-repeat of *Xist* (Carter et al., 2020). Insertion of an ERV into an A-repeat-deficient *Xist* transcript rescues *Xist*-mediated gene silencing, suggesting that *Xist* might have adapted a functional TE RNA-protein interaction for dose compensation (Carter et al., 2020).

Braveheart is a mouse ∼590 nt lncRNA that carries out a regulatory function in cardiovascular lineage commitment (Klattenhoff et al., 2013). It contains a short G-rich motif termed AGIL that is necessary for the differentiation of mouse ESCs to cardiomyocytes (Xue et al., 2016). This motif interacts with the zinc-finger transcription factor CNBP/ZNF9 (Xue et al., 2016), a heart-specific transcription factor known to bind single-stranded G-rich DNA and RNA motifs (Calcaterra et al., 2010; Chen, 2003). It has been proposed that *Braveheart* and CNBP function together to regulate cardiac gene expression (Xue et al., 2016). Thus, *Braveheart* directs cell fate via a short RNA motif that interacts with a tissue-specific transcription factor (Xue et al., 2016).

The lncRNA *growth arrest-specific 5* (*GAS5*) is one of the first studied lncRNAs (Schneider et al., 1988). *GAS5* regulates the survival of female germline stem cells *in vitro* (Wang et al., 2018) and the self-renewal of human ESCs (Xu et al., 2016). In certain cases, *GAS5* acts as a glucocorticoid receptor (GR) decoy (Kino et al., 2010), participating in the negative-feedback loop of activated GRs, directing cell fate and regulating transcription. The region implicated in this interaction is located in the last exon of the 12-exon *GAS5* transcript and includes a motif that 'mimics' genomic GR-binding sites (Hudson et al., 2014). Interestingly, a single nucleotide substitution is sufficient to abolish the GR-*GAS5* interaction. Sequence alignment showed that this recognition site is conserved among the haplorhine lineage (a suborder of primates), suggesting that other RNAs containing this motif may also interact with GRs (Hudson et al., 2014).

Megamind (also known as *Tunar*) harbours a region of ∼200 nt that is deeply conserved among vertebrates (Lin et al., 2014; Ulitsky et al., 2011). Knockdown of *Tunar* in mouse ESCs inhibits their differentiation into neural lineages (Lin et al., 2014). It has been proposed that *Tunar* carries out its regulatory function in pluripotency through an interaction with a protein complex composed of PTBP1, HNRNPK and NCL. Specifically, RNA pulldown experiments have shown that the conserved sequence binds to PTBP1, HNRNPK and NCL proteins with an affinity comparable with that of the full-length *Tunar* transcript, whereas binding affinity of the proteins to a synthetic transcript without the conserved sequence is decreased. The conserved region of *Tunar* therefore appears to act as an RBP-binding platform driving cellular differentiation (Lin et al., 2014).

lncRNA-protein interactions that play regulatory functions in development and organ growth can also contribute to human diseases. For example, the ubiquitously expressed lncRNA *Pint* (p53-induced noncoding transcript), which is under the direct control of p53 (Marín-Béjar et al., 2013), has been implicated in cancer. In mouse cells, *Pint* controls cell proliferation and survival by interacting with PRC2 and regulating the expression of hundreds of genes (Marín-Béjar et al., 2013). In line with this, it has been demonstrated that $Pint^{-/-}$ mice are noticeably smaller and have reduced body weight compared with wild-type mice (Sauvageau et al., 2013). Similar to mouse *Pint*, human *PINT* is regulated by p53 and interacts with PRC2. Of note, *PINT* inhibits the migration and invasion of cancer cells, and this function is carried out by a short sequence motif conserved in mammals that is required for the interaction between *PINT* and PRC2, and the downregulation of pro-invasion genes (Marín-Béjar et al., 2017).

Identifying RNA-binding proteins associated with lncRNA elements

As highlighted above, a number of lncRNA molecular functions are mediated by interactions with proteins. As such, reliable identification of lncRNA-associated proteins is key for delineating the molecular mechanisms of lncRNA action. Unbiased identification of lncRNA associated proteins is typically carried out by applying so-called RNA-centric approaches, which identify RBPs that interact with a specific RNA of interest. The most commonly used strategy relies on cross-linking and probe-based affinity capture of a test RNA followed by mass spectrometry-based identification of the co-purified proteins (Fig. 2A) (Chu et al., 2015a). Despite the high potential of this methodology, the identification of lncRNA-protein interactions via this approach remains technically challenging due to the general inefficiency of RNA pull-downs. Recently, an alternative approach termed incPRINT, which enables in-cell identification of protein interactions with any RNA of interest, was developed (Graindorge et al., 2019). This technique is based on screening a library of tagged proteins (including the majority of all known human RBPs, transcription factors and chromatin modifiers) with a test RNA that is tethered to a luciferase detector. Instead of pulling down the test RNA, high-throughput immunoprecipitation of thousands of tagged RBPs is performed followed by luciferase detection of their interactions with the test RNA (Graindorge et al., 2019; Sabaté-Cadenas and Shkumatava, 2020). incPRINT is suitable for the identification of in-cell RNA-interacting proteomes of full-length transcripts of various endogenous abundancies, and for mapping proteins that bind to different regions of longer RNA transcripts (Graindorge et al., 2019).

An alternative method to identify proteins that bind to specific sequence motifs harnesses existing data from crosslinking and

Development (2021) 148, dev182741. doi:10.1242/dev.182741

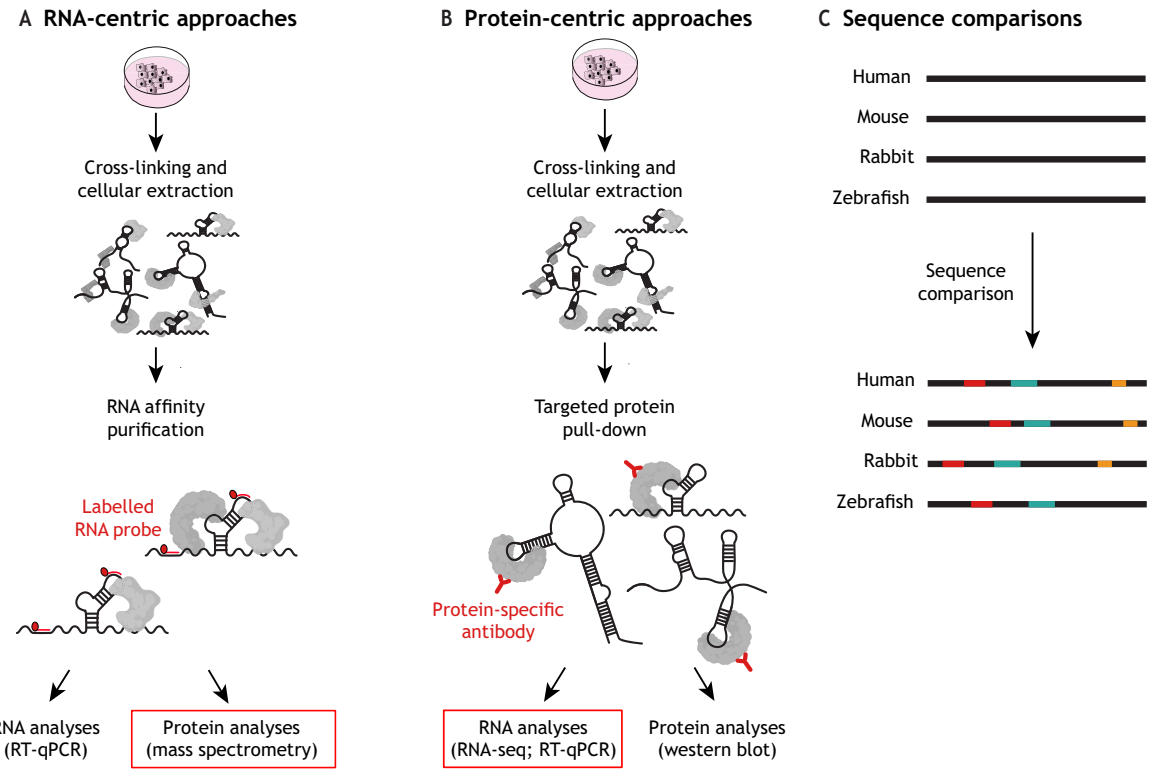


Fig. 2. Identifying lncRNA-binding proteins and conserved sequence elements. (A) RNA-centric approaches allow the identification of proteins that can bind to specific RNAs. In these approaches, probe-based affinity capture of a test RNA is performed after UV or chemical crosslinking and subsequent cell lysis. Following biochemical purification of RNA-protein complexes, RNA-binding proteins are identified by mass-spectrometry. (B) Protein-centric approaches enable the identification of RNAs bound by specific proteins. CLIP (crosslinking and immunoprecipitation) approaches use UV or chemical crosslinking to covalently bind protein-RNA complexes, allowing stringent purification of RNA-binding proteins with specific antibodies along with small fragments of RNA, which are amplified and sequenced. (C) Sequence comparisons of homologous lncRNAs enable discoveries of conserved sequence elements.

immunoprecipitation (CLIP) assays. CLIP methods are based on chemical or UV crosslinking followed by immunoprecipitation of a test protein to identify all RNA fragments that crosslink to the protein of interest, which in turn is followed by sequencing (Darnell, 2012; Jensen and Darnell, 2008) (Fig. 2B). Currently, the eCLIP data from the ENCODE project includes ∼150 RBPs (Van Nostrand et al., 2020a, 2020b) the binding sites of which can be mapped and analyzed in specific transcripts of interest. Although the application of CLIP data to an RNA of interest is limited by the number of proteins and cell lines for which the data are available (currently, predominantly K562 or HepG2 cells), it is powerful as an orthogonal approach for validating RNA-centric methods (Graindorge et al., 2019) and for mapping RBP-binding sites identified by eCLIP (Van Nostrand et al., 2017) to RNA sequence elements (Kirk et al., 2018; Lee et al., 2016; Tichon et al., 2018). The identification of RNA motifs and their associated proteins can also enable the prediction of RNA-protein interactions based on lncRNA sequences, although this will require the development of algorithms that are able to detect short sequence homologies. These predictions can then be tested by the aforementioned RNA-protein interaction technologies.

RNA motif discovery as a starting point for lncRNA functional classification

One of the major bottlenecks in identifying physiologically important lncRNAs is the difficulty in establishing a connection between lncRNA sequence and function. Moreover, because even lncRNAs with well-characterized cellular and organismal functions represent a highly heterogeneous group, it is becoming increasingly important to classify them into functional subgroups. The identification of sequence motifs linked to specific molecular events, such as localization and protein binding, and the subsequent discovery of these motifs in other lncRNAs and/or mRNAs, might help in assigning specific functions to individual lncRNAs, as experimental interrogations are often tedious. Indeed, the comparative analysis of lncRNA sequences across different species (Fig. 2C) (Chen et al., 2016; Hezroni et al., 2015; Necsulea et al., 2014) can be a powerful approach for studying the functions and molecular mechanisms of action of lncRNAs, especially when followed by experimental validations. Such alignment-based comparative approaches have led to the identification of conserved sequence motifs within lncRNA transcripts such as *Megamind*, *Cyrano*, *NEAT1*, *MALAT1*, *NORAD*, *PINT* and many more (Cornelis et al., 2016; Lee et al., 2016; Marín-Béjar et al., 2017; Tichon et al., 2016; Ulitsky et al., 2011; Wilusz et al., 2012). Moreover, uncovering RNA elements with known functions within as-yet uncharacterized lncRNA sequences can facilitate the discovery of novel lncRNA functions. For example, a *SRA1*-like (steroid receptor RNA activator 1) sequence element that mediates androgen receptor (AR) binding was found in the lncRNA *SLNCR1* (Schmidt et al., 2016) and, based on the known function of this motif as a co-activator of hormonal pathways (Lanz et al., 1999; Novikova et al., 2012), it was possible to predict the function of *SLNCR1*. As predicted, targeting this specific binding site within *SLNCR1* prevented its interaction with AR, leading to inhibition of melanogenesis (Schmidt et al., 2020).

Although powerful, alignment-based approaches require relatively long stretches of sequence conservation between homologs. However, most lncRNAs lack long stretches with high sequence conservation. Detection of rather short, conserved sequence motifs within long and rapidly evolving transcripts would be an important step towards assigning sequences to biological and molecular functions. Application of the SEEKR algorithm, developed to detect the presence of similar short sequence motifs of three to eight nucleotides (termed *k*-mers), recently allowed the functional classification of lncRNAs with related functions (Kirk et al., 2018). Reasoning that lncRNAs with shared functions harbour a shared set of similar *k*-mers while lacking linear homology, SEEKR groups lncRNAs based on their *k*-mer profiles. Subgroups of lncRNAs with similar *k*-mer contents share biological properties, such as protein-binding and subcellular localization, even between groups with no apparent evolutionary relationships. Based on both computational data and experimental validation, it has been proposed that no linear sequence homology is required for lncRNAs to support conserved functions. Instead, lncRNA functions are driven by similar sets of short *k*-mer motifs that represent RNA-binding protein motifs (Kirk et al., 2018). The classification of lncRNAs into functional groups based on short motifs is very promising but there is a need for improved algorithms for unbiased identification of short conserved motifs within lncRNA transcripts across more distant species.

Conclusions and perspectives

The regulatory functions of the lncRNA sequence motifs discussed here demonstrate the importance of primary RNA sequences and the necessity of careful analyses of lncRNA sequence elements, as many of them are shared with protein-coding RNAs. RNA motifs regulate a range of different functions, including transcript biogenesis, subcellular localization and transcript interactions with proteins, DNA and other RNAs, such as miRNAs and protein-coding mRNAs. The identification and further characterization of regulatory RNA motifs will facilitate the prediction of lncRNA interactions and mechanisms of action. Establishing a database of functional RNA elements identified experimentally as well as RNA motifs determined by computational approaches would enable the classification of lncRNAs into functional subgroups. Because several thousand of lncRNAs have been annotated and it is likely that only a minor fraction of them are functional, the identification of conserved RNA motifs would help to prioritize lncRNAs for experimental interrogation to identify transcripts with important cellular functions.

The unbiased identification of regulatory sequence elements within lncRNA transcripts has demonstrated that some coding mRNAs employ the same regulatory RNA motifs. As such, the characterization of functional RNA sequence elements might also help us to understand regulatory sequences that are embedded in mRNA transcripts. Indeed, studying regulatory RNA motifs within mRNA transcripts is often challenging, because the impact of genetic manipulation of sequences located in untranslated regions of mRNAs cannot always be easily uncoupled from their impact on protein production and function. Hence, the characterization of RNA sequence elements in noncoding RNAs could provide a more accessible route and would allow predictions of mRNA regulation by these elements. Furthermore, regulatory RNA sequence motifs represent an untapped source of potential therapeutic targets. As the realization that RNA molecules are druggable targets (Warner et al., 2018), there has been an effort in the field to find small molecule inhibitors that target specific RNA transcripts. However, it is challenging to identify selective drugs that bind specific RNAs without broad effects. Targeting regulatory RNA motifs may provide this specificity. Furthermore, many proteins have been found to be undruggable (Warner et al., 2018); thus, targeting RNA motifs that regulate the transcription, stability and/or translation of mRNAs impacting protein production could open the door to novel treatment strategies.

Acknowledgements

We thank all members of the Shkumatava laboratory and Dina Zielinsky for useful discussions.

Competing interests

The authors declare no competing or financial interests.

Funding

The authors' research is supported by LabEx DEEP, France (ANR-11-LABX-0044 and ANR-10-IDEX-0001-02) and by a SDSV n°577, université PSL doctoral fellowship to F.C.

References

Alberti, C. and Cochella, L. (2017). A framework for understanding the roles of miRNAs in animal development. *Development* **144**, 2548-2559. doi:10.1242/dev.146613

Ameres, S. L., Horwich, M. D., Hung, J.-H., Xu, J., Ghildiyal, M., Weng, Z. and Zamore, P. D. (2010). Target RNA-directed trimming and tailing of small silencing RNAs. *Science* **328**, 1534-1539. doi:10.1126/science.1187058

Arun, G. and Spector, D. L. (2019). MALAT1 long non-coding RNA and breast cancer. *RNA Biol.* **16**, 860-863. doi:10.1080/15476286.2019.1592072

Azam, S., Hou, S., Zhu, B., Wang, W., Hao, T., Bu, X., Khan, M. and Lei, H. (2019). Nuclear retention element recruits U1 snRNP components to restrain spliced lncRNAs in the nucleus. *RNA Biol.* **16**, 1001-1009. doi:10.1080/15476286.2019.1620061

Bartel, D. P. (2018). Metazoan MicroRNAs. *Cell* **173**, 20-51. doi:10.1016/j.cell.2018.03.006

Bian, S. and Sun, T. (2011). Functions of noncoding RNAs in neural development and neurological diseases. *Mol. Neurobiol.* **44**, 359-373. doi:10.1007/s12035-011-8211-3

Bitetti, A., Mallory, A. C., Golini, E., Carrieri, C., Carreño Gutiérrez, H., Perlas, E., Pérez-Rico, Y. A., Tocchini-Valentini, G. P., Enright, A. J., Norton, W. H. J. et al. (2018). MicroRNA degradation by a conserved target RNA regulates animal behavior. *Nat. Struct. Mol. Biol.* **25**, 244-251. doi:10.1038/s41594-018-0032-x

Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C. et al. (1991). Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**, 325-329. doi:10.1038/351325a0

Brown, S. D. M. (1991). XIST and the mapping of the X chromosome inactivation centre. *BioEssays* **13**, 607-612. doi:10.1002/bies.950131112

Brown, J. A., Valenstein, M. L., Yario, T. A., Tycowski, K. T. and Steitz, J. A. (2012). Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc. Natl. Acad. Sci. USA* **109**, 19202-19207. doi:10.1073/pnas.1217338109

Brown, J. A., Bulkley, D., Wang, J., Valenstein, M. L., Yario, T. A., Steitz, T. A. and Steitz, J. A. (2014). Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat. Struct. Mol. Biol.* **21**, 633-640. doi:10.1038/nsmb.2844

Buske, F. A., Bauer, D. C., Mattick, J. S. and Bailey, T. L. (2012). Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res.* **22**, 1372-1381. doi:10.1101/gr.130237.111

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. and Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915-1927. doi:10.1101/gad.17446611

Calcaterra, N. B., Armas, P., Weiner, A. M. J. and Borgognone, M. (2010). CNBP: A multifunctional nucleic acid chaperone involved in cell death and proliferation control. *IUBMB Life* **62**, 707-714. doi:10.1002/iub.379

Cao, Z., Pan, X., Yang, Y., Huang, Y. and Shen, H.-B. (2018). The lncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. *Bioinformatics* **34**, 2185-2194. doi:10.1093/bioinformatics/bty085

Carlevaro-Fita, J., Polidori, T., Das, M., Navarro, C., Zoller, T. I. and Johnson, R. (2019). Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. *Genome Res.* **29**, 208-222. doi:10.1101/gr.229922.117

Carter, A. C., Xu, J., Nakamoto, M. Y., Wei, Y., Zarnegar, B. J., Shi, Q., Broughton, J. P., Ransom, R. C., Salhotra, A., Nagaraja, S. D. et al. (2020).

Spen links RNA-mediated endogenous retrovirus silencing and X chromosome inactivation. *eLife* **9**, e54508. doi:10.7554/eLife.54508
Chen, W. (2003). The zinc-finger protein CNBP is required for forebrain formation in the mouse. *Development* **130**, 1367-1379. doi:10.1242/dev.00349
Chen, L.-L. (2016). Linking long noncoding RNA localization and function. *Trends Biochem. Sci.* **41**, 761-772. doi:10.1016/j.tibs.2016.07.003
Chen, J., Shishkin, A. A., Zhu, X., Kadri, S., Maza, I., Guttman, M., Hanna, J. H., Regev, A. and Garber, M. (2016). Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol.* **17**, 19. doi:10.1186/s13059-016-0880-9
Chu, C., Spitale, R. C. and Chang, H. Y. (2015a). Technologies to probe functions and mechanisms of long noncoding RNAs. *Nat. Struct. Mol. Biol.* **22**, 29-35. doi:10.1038/nsmb.2921
Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., Magnuson, T., Heard, E. and Chang, H. Y. (2015b). Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416. doi:10.1016/j.cell.2015.03.025
Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A. and Lawrence, J. B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of Paraspeckles. *Mol. Cell* **33**, 717-726. doi:10.1016/j.molcel.2009.01.026
Cornelis, G., Souquere, S., Vernochet, C., Heidmann, T. and Pierron, G. (2016). Functional conservation of the lncRNA NEAT1 in the ancestrally diverged marsupial lineage: Evidence for NEAT1 expression and associated paraspeckle assembly during late gestation in the opossum *Monodelphis domestica*. *RNA Biol.* **13**, 826-836. doi:10.1080/15476286.2016.1197482
Darnell, R. (2012). CLIP (cross-linking and immunoprecipitation) identification of RNAs bound by a specific protein. *Cold Spring Harb. Protoc.* **2012**, 1146-1160. doi:10.1101/pdb.prot072132
de la Mata, M., Gaidatzis, D., Vitanescu, M., Stadler, M. B., Wentzel, C., Scheiffele, P., Filipowicz, W. and Grosshans, H. (2015). Potent degradation of neuronal miRNAs induced by highly complementary targets. *EMBO Rep.* **16**, 500-511. doi:10.15252/embr.201540078
Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G. et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775-1789. doi:10.1101/gr.132159.111
Diederichs, S. (2014). The four dimensions of noncoding RNA conservation. *Trends Genet.* **30**, 121-123. doi:10.1016/j.tig.2014.01.004
Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N. J., Van Nostrand, E. L., Pratt, G. A. et al. (2018). Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell* **70**, 854-867.e859. doi:10.1016/j.molcel.2018.05.001
Dossin, F., Pinheiro, I., Żylicz, J. J., Roensch, J., Collombet, S., Le Saux, A., Chelmicki, T., Attia, M., Kapoor, V., Zhan, Y. et al. (2020). SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature* **578**, 455-460. doi:10.1038/s41586-020-1974-9
Eissmann, M., Gutschner, T., Hämmerle, M., Günther, S., Caudron-Herger, M., Gross, M., Schirmacher, P., Rippe, K., Braun, T., Zörnig, M. et al. (2012). Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biol.* **9**, 1076-1087. doi:10.4161/rna.21089
Elguindy, M. M., Kopp, F., Goodarzi, M., Rehfeld, F., Thomas, A., Chang, T.-C. and Mendell, J. T. (2019). PUMILIO, but not RBMX, binding is required for regulation of genomic stability by noncoding RNA NORAD. *eLife* **8**, e48625. doi:10.7554/eLife.48625
Elisaphenko, E. A., Kolesnikov, N. N., Shevchenko, A. I., Rogozin, I. B., Nesterova, T. B., Brockdorff, N. and Zakian, S. M. (2008). A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. *PLoS ONE* **3**, e2521. doi:10.1371/journal.pone.0002521
Esposito, R., Bosch, N., Lanzós, A., Polidori, T., Pulido-Quetglas, C. and Johnson, R. (2019). Hacking the Cancer genome: profiling therapeutically actionable long non-coding RNAs using CRISPR-Cas9 screening. *Cancer Cell* **35**, 545-557. doi:10.1016/j.ccell.2019.01.019
Felsenfeld, G. and Rich, A. (1957). Studies on the formation of two- and three-stranded polyribonucleotides. *Biochim. Biophys. Acta* **26**, 457-468. doi:10.1016/0006-3002(57)90091-4
Feng, J. (2006). The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* **20**, 1470-1484. doi:10.1101/gad.1416106
Gendrel, A.-V. and Heard, E. (2014). Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. *Annu. Rev. Cell Dev. Biol.* **30**, 561-580. doi:10.1146/annurev-cellbio-101512-122415
Ghini, F., Rubolino, C., Climent, M., Simeone, I., Marzi, M. J. and Nicassio, F. (2018). Endogenous transcripts control miRNA levels and activity in mammalian cells by target-directed miRNA degradation. *Nat. Commun.* **9**, 3119. doi:10.1038/s41467-018-05182-9
Gong, C., Li, Z., Ramanujan, K., Clay, I., Zhang, Y., Lemire-Brachat, S. and Glass, D. J. (2015). A long non-coding RNA, LncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. *Dev. Cell* **34**, 181-191. doi:10.1016/j.devcel.2015.05.009
Goudarzi, M., Berg, K., Pieper, L. M. and Schier, A. F. (2019). Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility. *eLife* **8**, e40815. doi:10.7554/eLife.40815
Graindorge, A., Pinheiro, I., Nawrocka, A., Mallory, A. C., Tsvetkov, P., Gil, N., Carolis, C., Buchholz, F., Ulitsky, I., Heard, E. et al. (2019). In-cell identification and measurement of RNA-protein interactions. *Nat. Commun.* **10**, 5317. doi:10.1038/s41467-019-13235-w
Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M. et al. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev. Cell* **24**, 206-214. doi:10.1016/j.devcel.2012.12.012
Gudenas, B. L. and Wang, L. (2018). Prediction of LncRNA subcellular localization with deep learning from sequence features. *Sci. Rep.* **8**, 16385. doi:10.1038/s41598-018-34708-w
Guo, C.-J., Ma, X.-K., Xing, Y.-H., Zheng, C.-C., Xu, Y.-F., Shan, L., Zhang, J., Wang, S., Wang, Y., Carmichael, G. G. et al. (2020). Distinct processing of lncRNAs contributes to non-conserved functions in stem cells. *Cell* **181**, 621-636.e622. doi:10.1016/j.cell.2020.03.006
Gutschner, T. and Diederichs, S. (2012). The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol.* **9**, 703-719. doi:10.4161/rna.20481
Guttman, M. and Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339-346. doi:10.1038/nature10887
Hacisuleyman, E., Goff, L. A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., McClanahan, P., Hendrickson, D. G., Sauvageau, M., Kelley, D. R. et al. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* **21**, 198-206. doi:10.1038/nsmb.2764
Hacisuleyman, E., Shukla, C. J., Weiner, C. L. and Rinn, J. L. (2016). Function and evolution of local repeats in the Firre locus. *Nat. Commun.* **7**, 11021. doi:10.1038/ncomms11021
Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P. and Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* **11**, 1110-1122. doi:10.1016/j.celrep.2015.04.023
Holmes, Z. E., Hamilton, D. J., Hwang, T., Parsonnet, N. V., Rinn, J. L., Wuttke, D. S. and Batey, R. T. (2020). The Sox2 transcription factor binds RNA. *Nat. Commun.* **11**, 1805. doi:10.1038/s41467-020-15571-8
Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nat. Med.* **21**, 1253-1261. doi:10.1038/nm.3981
Hudson, W. H., Pickard, M. R., de Vera, I. M. S., Kuiper, E. G., Mourtada-Maarabouni, M., Conn, G. L., Kojetin, D. J., Williams, G. T. and Ortlund, E. A. (2014). Conserved sequence-specific lincRNA–steroid receptor interactions drive transcriptional repression and direct cell fate. *Nat. Commun.* **5**, 5395. doi:10.1038/ncomms6395
Hutchinson, J. N., Ensminger, A. W., Clemson, C. M., Lynch, C. R., Lawrence, J. B. and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 39. doi:10.1186/1471-2164-8-39
Jensen, K. B. and Darnell, R. B. (2008). CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. *Methods Mol. Biol.* **488**, 85-98. doi:10.1007/978-1-60327-475-3_6
Kino, T., Hurt, D. E., Ichijo, T., Nader, N. and Chrousos, G. P. (2010). Noncoding RNA Gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* **3**, ra8. doi:10.1126/scisignal.2000568
Kirk, J. M., Kim, S. O., Inoue, K., Smola, M. J., Lee, D. M., Schertzer, M. D., Wooten, J. S., Baker, A. R., Sprague, D., Collins, D. W. et al. (2018). Functional classification of long non-coding RNAs by k-mer content. *Nat. Genet.* **50**, 1474-1482. doi:10.1038/s41588-018-0207-8
Klattenhoff, C. A., Scheuermann, J. C., Surface, L. E., Bradley, R. K., Fields, P. A., Steinhauser, M. L., Ding, H., Butty, V. L., Torrey, L., Haas, S. et al. (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* **152**, 570-583. doi:10.1016/j.cell.2013.01.003
Kleaveland, B., Shi, C. Y., Stefano, J. and Bartel, D. P. (2018). A network of noncoding regulatory RNAs acts in the mammalian brain. *Cell* **174**, 350-362.e317. doi:10.1016/j.cell.2018.05.022
Kopp, F., Elguindy, M. M., Yalvac, M. E., Zhang, H., Chen, B., Gillett, F. A., Lee, S., Sivakumar, S., Yu, H., Xie, Y. et al. (2019). PUMILIO hyperactivity drives premature aging of Norad-deficient mice. *eLife* **8**, e42650. doi:10.7554/eLife.42650
Kuo, C.-C., Hänzelmann, S., Sentürk Cetin, N., Frank, S., Zajzon, B., Derks, J.-P., Akhade, V. S., Ahuja, G., Kanduri, C., Grummt, I. et al. (2019). Detection of RNA–DNA binding sites in long noncoding RNAs. *Nucleic Acids Res.* **47**, e32-e32. doi:10.1093/nar/gkz037
Kutter, C., Watt, S., Stefflova, K., Wilson, M. D., Goncalves, A., Ponting, C. P., Odom, D. T. and Marques, A. C. (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* **8**, e1002841. doi:10.1371/journal.pgen.1002841
Lanz, R. B., McKenna, N. J., Onate, S. A., Albrecht, U., Wong, J., Tsai, S. Y., Tsai, M.-J. and O'Malley, B. W. (1999). A steroid receptor coactivator, SRA, functions

as an RNA and is present in an SRC-1 complex. *Cell* **97**, 17-27. doi:10.1016/S0092-8674(00)80711-4

Lavalou, P., Eckert, H., Damy, L., Constanty, F., Majello, S., Bitetti, A., Graindorge, A. and Shkumatava, A. (2019). Strategies for genetic inactivation of long noncoding RNAs in zebrafish. *RNA* **25**, 897-904. doi:10.1261/rna.069484.118

Lee, S., Kopp, F., Chang, T.-C., Sataluri, A., Chen, B., Sivakumar, S., Yu, H., Xie, Y. and Mendell, J. T. (2016). Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. *Cell* **164**, 69-80. doi:10.1016/j.cell.2015.12.017

Lewandowski, J. P., Lee, J. C., Hwang, T., Sunwoo, H., Goldstein, J. M., Groff, A. F., Chang, N. P., Mallard, W., Williams, A., Henao-Meija, J. et al. (2019). The Firre locus produces a trans-acting RNA molecule that functions in hematopoiesis. *Nat. Commun.* **10**, 5137. doi:10.1038/s41467-019-12970-4

Li, Y., Syed, J. and Sugiyama, H. (2016). RNA-DNA triplex formation by long noncoding RNAs. *Cell Chem. Biol.* **23**, 1325-1333. doi:10.1016/j.chembiol.2016.09.011

Li, L., Zhuang, Y., Zhao, X. and Li, X. (2018). Long non-coding RNA in neuronal development and neurological disorders. *Front. Genet.* **9**, 744. doi:10.3389/fgene.2018.00744

Lin, N., Chang, K.-Y., Li, Z., Gates, K., Rana, Z. A., Dang, J., Zhang, D., Han, T., Yang, C.-S., Cunningham, T. J. et al. (2014). An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol. Cell* **53**, 1005-1019. doi:10.1016/j.molcel.2014.01.021

Lu, Z., Zhang, Q. C., Lee, B., Flynn, R. A., Smith, M. A., Robinson, J. T., Davidovich, C., Gooding, A. R., Goodrich, K. J., Mattick, J. S. et al. (2016). RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165**, 1267-1279. doi:10.1016/j.cell.2016.04.028

Lubelsky, Y. and Ulitsky, I. (2018). Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature* **555**, 107-111. doi:10.1038/nature25757

Mallory, A. C. and Shkumatava, A. (2015). LncRNAs in vertebrates: advances and challenges. *Biochimie* **117**, 3-14. doi:10.1016/j.biochi.2015.03.014

Mao, Y. S., Sunwoo, H., Zhang, B. and Spector, D. L. (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13**, 95-101. doi:10.1038/ncb2140

Marahrens, Y., Panning, B., Dausman, J., Strauss, W. and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev.* **11**, 156-166. doi:10.1101/gad.11.2.156

Marchese, F. P., Raimondi, I. and Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* **18**, 206. doi:10.1186/s13059-017-1348-2

Marcinowski, L., Tanguy, M., Krmpotic, A., Rädle, B., Lisnić, V. J., Tuddenham, L., Chane-Woon-Ming, B., Ruzsics, Z., Erhard, F., Benkartek, C. et al. (2012). Degradation of cellular mir-27 by a novel, highly abundant viral transcript is important for efficient virus replication in vivo. *PLoS Pathog.* **8**, e1002510. doi:10.1371/journal.ppat.1002510

Marín-Béjar, O., Marchese, F. P., Athie, A., Sánchez, Y., González, J., Segura, V., Huang, L., Moreno, I., Navarro, A., Monzó, M. et al. (2013). Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. *Genome Biol.* **14**, R104. doi:10.1186/gb-2013-14-9-r104

Marín-Béjar, O., Mas, A. M., González, J., Martinez, D., Athie, A., Morales, X., Galduroz, M., Raimondi, I., Grossi, E., Guo, S. et al. (2017). The human lncRNA LINC-PINT inhibits tumor cell invasion through a highly conserved sequence element. *Genome Biol.* **18**, 202. doi:10.1186/s13059-017-1331-y

McHugh, C. A., Chen, C.-K., Chow, A., Surka, C. F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A. et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236. doi:10.1038/nature14443

Michalik, K. M., You, X., Manavski, Y., Doddaballapur, A., Zörnig, M., Braun, T., John, D., Ponomareva, Y., Chen, W., Uchida, S. et al. (2014). Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. *Circ. Res.* **114**, 1389-1397. doi:10.1161/CIRCRESAHA.114.303265

Miller, M. A. and Olivas, W. M. (2011). Roles of Puf proteins in mRNA degradation and translation. *Wiley Interdiscip. Rev. RNA* **2**, 471-492. doi:10.1002/wrna.69

Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W. et al. (2015). Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276. doi:10.1126/science.aab2276

Miyagawa, R., Tano, K., Mizuno, R., Nakamura, Y., Ijiri, K., Rakwal, R., Shibato, J., Masuo, Y., Mayeda, A., Hirose, T. et al. (2012). Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA* **18**, 738-751. doi:10.1261/rna.028639.111

Moindrot, B. and Brockdorff, N. (2016). RNA binding proteins implicated in Xist-mediated chromosome silencing. *Semin. Cell Dev. Biol.* **56**, 58-70. doi:10.1016/j.semcdb.2016.01.029

Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T. B. and Brockdorff, N. (2015). A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. *Cell Rep.* **12**, 562-572. doi:10.1016/j.celrep.2015.06.053

Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S. and Wutz, A. (2015). Identification of Spen as a crucial factor for xist function through forward genetic screening in haploid embryonic stem cells. *Cell Rep.* **12**, 554-561. doi:10.1016/j.celrep.2015.06.067

Mukherjee, N., Calviello, L., Hirsekorn, A., de Pretis, S., Pelizzola, M. and Ohler, U. (2017). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat. Struct. Mol. Biol.* **24**, 86-96. doi:10.1038/nsmb.3325

Munschauer, M., Nguyen, C. T., Sirokman, K., Hartigan, C. R., Hogstrom, L., Engreitz, J. M., Ulirsch, J. C., Fulco, C. P., Subramanian, V., Chen, J. et al. (2018). The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature* **561**, 132-136. doi:10.1038/s41586-018-0453-z

Naganuma, T. and Hirose, T. (2013). Paraspeckle formation during the biogenesis of long non-coding RNAs. *RNA Biol.* **10**, 456-461. doi:10.4161/rna.23547

Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y. F., Goshima, N. and Hirose, T. (2012). Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles: LncRNA processing for nuclear body architecture. *EMBO J.* **31**, 4020-4034. doi:10.1038/emboj.2012.251

Nakagawa, S., Naganuma, T., Shioi, G. and Hirose, T. (2011). Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *J. Cell Biol.* **193**, 31-39. doi:10.1083/jcb.201011110

Nakagawa, S., Ip, J. Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T. and Prasanth, K. V. (2012). Malat1 is not an essential component of nuclear speckles in mice. *RNA* **18**, 1487-1499. doi:10.1261/rna.033217.112

Nakagawa, S., Shimada, M., Yanaka, K., Mito, M., Arai, T., Takahashi, E., Fujita, Y., Fujimori, T., Standaert, L., Marine, J.-C. et al. (2014). The lncRNA Neat1 is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. *Development* **141**, 4618-4627. doi:10.1242/dev.110544

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J. C., Grützner, F. and Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635-640. doi:10.1038/nature12943

Nesterova, T. B., Slobodyanyuk, S. Y., Elisaphenko, E. A., Shevchenko, A. I., Johnston, C., Pavlova, M. E., Rogozin, I. B., Kolesnikov, N. N., Brockdorff, N. and Zakian, S. M. (2001). Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res.* **11**, 833-849. doi:10.1101/gr.174901

Noh, J. H., Kim, K. M., McClusky, W. G., Abdelmohsen, K. and Gorospe, M. (2018). Cytoplasmic functions of long noncoding RNAs. *Wiley Interdiscip. Rev. RNA* **9**, e1471. doi:10.1002/wrna.1471

Novikova, I. V., Hennelly, S. P. and Sanbonmatsu, K. Y. (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res.* **40**, 5034-5051. doi:10.1093/nar/gks071

Paraskevopoulou, M. D. and Hatzigeorgiou, A. G. (2016). Analyzing MiRNA–LncRNA interactions. In *Long Non-Coding RNAs* (ed. Y. Feng and L. Zhang), pp. 271-286. New York, NY: Springer New York.

Perry, R. B.-T. and Ulitsky, I. (2016). The functions of long noncoding RNAs in development and stem cells. *Development* **143**, 3882-3894. doi:10.1242/dev.140962

Prasanth, K. V., Prasanth, S. G., Xuan, Z., Hearn, S., Freier, S. M., Bennett, C. F., Zhang, M. Q. and Spector, D. L. (2005). Regulating gene expression through RNA nuclear retention. *Cell* **123**, 249-263. doi:10.1016/j.cell.2005.08.033

Rinn, J. L. and Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145-166. doi:10.1146/annurev-biochem-051410-092902

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E. et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323. doi:10.1016/j.cell.2007.05.022

Rivas, E., Clements, J. and Eddy, S. R. (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* **14**, 45-48. doi:10.1038/nmeth.4066

Sabaté-Cadenas, X. and Shkumatava, A. (2020). In-cell discovery of RNA-protein interactions. *Trends Biochem. Sci.* **45**, 272-273. doi:10.1016/j.tibs.2019.12.003

Sarangdhar, M. A., Chaubey, D., Srikakulam, N. and Pillai, B. (2018). Parentally inherited long non-coding RNA Cyrano is involved in zebrafish neurodevelopment. *Nucleic Acids Res.* **46**, 9726-9735. doi:10.1093/nar/gky628

Sasaki, Y. T. F., Ideue, T., Sano, M., Mituyama, T. and Hirose, T. (2009). MENε/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc. Natl. Acad. Sci. USA* **106**, 2525-2530. doi:10.1073/pnas.0807899106

Sauvageau, M., Goff, L. A., Lodato, S., Bonev, B., Groff, A. F., Gerhardinger, C., Sanchez-Gomez, D. B., Hacisuleyman, E., Li, E., Spence, M. et al. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* **2**, e01749. doi:10.7554/eLife.01749

Schmidt, K., Joyce, C. E., Buquicchio, F., Brown, A., Ritz, J., Distel, R. J., Yoon, C. H. and Novina, C. D. (2016). The lncRNA SLNCR1 mediates melanoma invasion through a conserved SRA1-like region. *Cell Rep.* **15**, 2025-2037. doi:10.1016/j.celrep.2016.04.018

Schmidt, K., Weidmann, C. A., Hilimire, T. A., Yee, E., Hatfield, B. M., Schneekloth, J. S., Weeks, K. M. and Novina, C. D. (2020). Targeting the oncogenic long non-coding RNA SLNCR1 by blocking its sequence-specific binding to the androgen receptor. *Cell Rep.* **30**, 541-554.e545. doi:10.1016/j.celrep.2019.12.011

Schneider, C., King, R. M. and Philipson, L. (1988). Genes specifically expressed at growth arrest of mammalian cells. *Cell* **54**, 787-793. doi:10.1016/S0092-8674(88)91065-3

Shukla, C. J., McCorkindale, A. L., Gerhardinger, C., Korthauer, K. D., Cabili, M. N., Shechner, D. M., Irizarry, R. A., Maass, P. G. and Rinn, J. L. (2018). High-throughput identification of RNA nuclear enrichment sequences. *EMBO J.* **37**, e98452. doi:10.15252/embj.201798452

Standaert, L., Adriaens, C., Radaelli, E., Van Keymeulen, A., Blanpain, C., Hirose, T., Nakagawa, S. and Marine, J.-C. (2014). The long noncoding RNA Neat1 is required for mammary gland development and lactation. *RNA* **20**, 1844-1849. doi:10.1261/rna.047332.114

Su, Z.-D., Huang, Y., Zhang, Z.-Y., Zhao, Y.-W., Wang, D., Chen, W., Chou, K.-C. and Lin, H. (2018). iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. *Bioinformatics* **34**, 4196-4204. doi:10.1093/bioinformatics/bty508

Sun, Q., Hao, Q. and Prasanth, K. V. (2018). nuclear long noncoding RNAs: key regulators of gene expression. *Trends Genet.* **34**, 142-157. doi:10.1016/j.tig.2017.11.005

Sunwoo, H., Dinger, M. E., Wilusz, J. E., Amaral, P. P., Mattick, J. S. and Spector, D. L. (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* **19**, 347-359. doi:10.1101/gr.087775.108

Takeda, K., Ichijo, H., Fujii, M., Mochida, Y., Saitoh, M., Nishitoh, H., Sampath, T. K. and Miyazono, K. (1998). Identification of a novel bone morphogenetic protein-responsive gene that may function as a noncoding RNA. *J. Biol. Chem.* **273**, 17079-17085. doi:10.1074/jbc.273.27.17079

Tichon, A., Gil, N., Lubelsky, Y., Havkin Solomon, T., Lemze, D., Itzkovitz, S., Stern-Ginossar, N. and Ulitsky, I. (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat. Commun.* **7**, 12209. doi:10.1038/ncomms12209

Tichon, A., Perry, R. B.-T., Stojic, L. and Ulitsky, I. (2018). SAM68 is required for regulation of Pumilio by the NORAD long noncoding RNA. *Genes Dev.* **32**, 70-78. doi:10.1101/gad.309138.117

Ulitsky, I. (2018). Interactions between short and long noncoding RNAs. *FEBS Lett.* **592**, 2874-2883. doi:10.1002/1873-3468.13085

Ulitsky, I. and Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26-46. doi:10.1016/j.cell.2013.06.020

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. and Bartel, D. P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537-1550. doi:10.1016/j.cell.2011.11.055

Uroda, T., Anastasakou, E., Rossi, A., Teulon, J.-M., Pellequer, J.-L., Annibale, P., Pessey, O., Inga, A., Chillón, I. and Marcia, M. (2019). Conserved pseudoknots in lncRNA MEG3 are essential for stimulation of the p53 pathway. *Mol. Cell* **75**, 982-995.e989. doi:10.1016/j.molcel.2019.07.025

Van Nostrand, E. L., Nguyen, T. B., Gelboin-Burkhart, C., Wang, R., Blue, S. M., Pratt, G. A., Louie, A. L. and Yeo, G. W. (2017). Robust, cost-effective profiling of RNA binding protein targets with single-end enhanced crosslinking and immunoprecipitation (seCLIP). *Methods Mol. Biol.* **1648**, 177-200. doi:10.1007/978-1-4939-7204-3_14

Van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., Blue, S. M., Chen, J. Y., Cody, N. A. L., Dominguez, D. et al. (2020a). A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711-719. doi:10.1038/s41586-020-2077-3

Van Nostrand, E. L., Pratt, G. A., Yee, B. A., Wheeler, E. C., Blue, S. M., Mueller, J., Park, S. S., Garcia, K. E., Gelboin-Burkhart, C., Nguyen, T. B. et al. (2020b). Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol.* **21**, 90. doi:10.1186/s13059-020-01982-9

Wang, K. C. and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43**, 904-914. doi:10.1016/j.molcel.2011.08.018

Wang, J., Gong, X., Tian, G. G., Hou, C., Zhu, X., Pei, X., Wang, Y. and Wu, J. (2018). Long noncoding RNA growth arrest-specific 5 promotes proliferation and survival of female germline stem cells in vitro. *Gene* **653**, 14-21. doi:10.1016/j.gene.2018.02.021

Warner, K. D., Hajdin, C. E. and Weeks, K. M. (2018). Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **17**, 547-558. doi:10.1038/nrd.2018.93

Washietl, S., Kellis, M. and Garber, M. (2014). Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res.* **24**, 616-628. doi:10.1101/gr.165035.113

Wilusz, J. E., Freier, S. M. and Spector, D. L. (2008). 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**, 919-932. doi:10.1016/j.cell.2008.10.012

Wilusz, J. E., JnBaptiste, C. K., Lu, L. Y., Kuhn, C.-D., Joshua-Tor, L. and Sharp, P. A. (2012). A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev.* **26**, 2392-2407. doi:10.1101/gad.204438.112

Wutz, A., Rasmussen, T. P. and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* **30**, 167-174. doi:10.1038/ng820

Xu, C., Zhang, Y., Wang, Q., Xu, Z., Jiang, J., Gao, Y., Gao, M., Kang, J., Wu, M., Xiong, J. et al. (2016). Long non-coding RNA GAS5 controls human embryonic stem cell self-renewal by maintaining NODAL signalling. *Nat. Commun.* **7**, 13287. doi:10.1038/ncomms13287

Xue, Z., Hennelly, S., Doyle, B., Gulati, A. A., Novikova, I. V., Sanbonmatsu, K. Y. and Boyer, L. A. (2016). A G-rich motif in the lncRNA braveheart interacts with a zinc-finger transcription factor to specify the cardiovascular lineage. *Mol. Cell* **64**, 37-50. doi:10.1016/j.molcel.2016.08.010

Yin, Y., Lu, J. Y., Zhang, X., Shao, W., Xu, Y., Li, P., Hong, Y., Cui, L., Shan, G., Tian, B. et al. (2020). U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature* **580**, 147-150. doi:10.1038/s41586-020-2105-3

Zhang, X., Rice, K., Wang, Y., Chen, W., Zhong, Y., Nakayama, Y., Zhou, Y. and Klibanski, A. (2010). Maternally Expressed Gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. *Endocrinology* **151**, 939-947. doi:10.1210/en.2009-0657

Zhang, B., Arun, G., Mao, Y. S., Lazar, Z., Hung, G., Bhattacharjee, G., Xiao, X., Booth, C. J., Wu, J., Zhang, C. et al. (2012). The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep.* **2**, 111-123. doi:10.1016/j.celrep.2012.06.003

Zhang, B., Gunawardane, L., Niazi, F., Jahanbani, F., Chen, X. and Valadkhan, S. (2014). A novel RNA motif mediates the strict nuclear localization of a long noncoding RNA. *Mol. Cell. Biol.* **34**, 2318-2329. doi:10.1128/MCB.01673-13

Zhao, J., Du, P., Cui, P., Qin, Y., Hu, C., Wu, J., Zhou, Z., Zhang, W., Qin, L. and Huang, G. (2018). LncRNA PVT1 promotes angiogenesis via activating the STAT3/VEGFA axis in gastric cancer. *Oncogene* **37**, 4094-4109. doi:10.1038/s41388-018-0250-z

Ziv, O., Gabryelska, M. M., Lun, A. T. L., Gebert, L. F. R., Sheu-Gruttadauria, J., Meredith, L. W., Liu, Z.-Y., Kwok, C. K., Qin, C.-F., MacRae, I. J. et al. (2018). COMRADES determines in vivo RNA structures and interactions. *Nat. Methods* **15**, 785-788. doi:10.1038/s41592-018-0121-0

Zuckerman, B. and Ulitsky, I. (2019). Predictive models of subcellular localization of long RNAs. *RNA* **25**, 557-572. doi:10.1261/rna.068288.118

RÉSUMÉ

Le dogme central de la biologie a été transformé par le séquençage du génome humain en découvrant un vaste nombre d'ARNs transcrits mais qui ne génèrent pas de protéines. Ces ARNs sont appelés non codant (ARNnc).

Nous avons défini la fonction de l'ARNnc *PVT1*, un transcrit dont la dérégulation est liée aux cancers. Le transcrit *PVT1* est présent chez toutes les espèces de vertébrés. Ainsi par comparaison de séquences nous avons trouvé des motifs ARNs conservés. L'ablation génétique de ces motifs chez le poisson amène à des défauts de développement et à une faible survie jusqu'à l'âge adulte. Alors que *PVT1* est exprimé dans les progéniteurs des érythrocytes chez la souris, chez le poisson, les mutants pour *PVT1* développent une anémie. Nous montrons que les motifs ARNs conservés interagissent avec des régulateurs de réparation de l'ADN, menant à l'idée que *PVT1* est impliqué dans la réparation de l'ADN pendant le développement des érythrocytes.

MOTS CLÉS

Cancer, Développement, ARN non-codant régulateur, motifs ARN, poisson-zèbre, PVT1

ABSTRACT

The central dogma of biology "DNA makes RNA, and RNA makes protein" has been transformed by the human genome sequencing uncovering a large number of transcribed RNAs that do not generate proteins, called non-coding RNAs (ncRNAs).

We defined the function of the ncRNA *PVT1*, a transcript that misregulation is linked to human cancers. *PVT1* transcript is present in all vertebrate species; by sequence comparison, we found conserved RNA motifs. Genetic removal of these motifs in zebrafish leads to developmental defects and poor survival to adulthood linked to the activation of the stress pathways. Consistent with *PVT1* expression in red blood cell progenitors during mice development, zebrafish *PVT1* mutants exhibit anemia suggesting potential conservation of the *PVT1* function. The conserved RNA motifs of *PVT1* have been found to interact with key regulators of the DNA repair pathway, leading to the idea that *PVT1* is integrated in the DNA repair network.

KEYWORDS

Cancer, Development, Long non-coding RNA, RNA motifs, zebrafish, PVT1