Conserved functions of rapidly evolving long noncoding RNAs : the long noncoding RNA menhir/CASC15 acts as a tumor suppressor in cutaneous melanoma

Perrine Lavalou

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'INSTITUT CURIE

Conserved functions of rapidly evolving long noncoding RNAs : the long noncoding RNA *menhir/CASC15* acts as a tumor suppressor in cutaneous melanoma

Ecole doctorale n°577

Structure et Dynamique des Systèmes Vivants

Spécialité SCIENCES DE LA VIE ET DE LA SANTE

COMPOSITION DU JURY :

M. CAPY Pierre
EGCE, Président du jury

Mme. HAREL-BELLAN Annick
CEA, Rapporteur

M. MUCHARDT Christian
Institut Pasteur, Rapporteur

Mme. MALLORY Allison
Institut Curie, Membre du jury

M. DEL BENE Filippo
Institut Curie, Invité

Mme. SHKUMATAVA Alena
Institut Curie, invitée

Soutenue par Perrine **LAVALOU**
le 21 Septembre 2018

Dirigée par **Allison MALLORY**

institutCurie PSL RESEARCH UNIVERSITY PARIS

Merci,

Bientôt 5 ans… une durée égale à celle passée à l'école primaire et peut être une évolution encore plus importante que cette période. Les rencontres, les soutiens et l'adversité m'ont permis de me construite et de me réaliser, en tant que scientifique mais surtout en tant qu'adulte.

5 ans de découvertes, de déceptions et de joies. 5 années très intenses.

Pour ces 5 années, je suis reconnaissante…

… Envers le Dr Shkumatava pour m'avoir rendue autonome, m'avoir laissé tester mes propres idées et résoudre moi-même les problèmes que j'ai rencontré au cours de ces 5 ans.

… Envers le Dr Allison Mallory, pour ton espoir sans failles, ton temps et ta capacité à m'écouter et à essayer de me calmer. Je suis également un petit peu désolée, je sais que mon stress est toxique.

… Envers mes rapporteurs, le Dr Harel-Bellan et le Dr Muchardt. Je vous remercie très sincèrement de prendre de votre temps pour examiner mon travail. Et envers le Dr Del Bene, invité lors de mon jury thèse, pour ces conseils précieux.

… Envers le Dr Capy, directeur de l'école doctorale et président de mon jury de thèse. Je ne suis pas certaine que ça ait été une bonne chose le jour où j'ai accepté cette bourse PSL vu la complexité des démarches administratives qui ont suivies. Mais je vous suis très reconnaissante de m'avoir écouté et surtout de m'avoir entendue

… Envers La Ligue Nationale Contre Le Cancer pour avoir soutenu mon projet d'étude de l'impact des lncARN sur le mélanome et avoir financé ma 4^{e} année (et pour leurs bracelets détecteurs d'UV aux Vieilles Charrues)

…Envers mon équipe. Merci pour ces liens que nous avons créés pendant 5 ans.
Je vais commencer par Angie. On a failli s'entretuer plusieurs fois mais tu resteras toujours pour moi le « father PhD » et j'espère que nous continuerons longtemps à trinquer à nos victoires et nos échecs à grands coups de pintes de cidre.
Yuvia, ma comparse de thèse, pour ton calme olympien en toute circonstance et ta gentillesse incroyable (sauf avec Angie).
Hélène, parce que ta joie de vivre est inestimable, et que tu as géré le labo d'une main de maitre… on ne serait pas allé bien loin sans toi
Arianna, ma voisine de bureau parce que tu m'as tellement de fois rendu le sourire ou fais rire aux éclats.
Sara, la douceur incarnée et la première à quitter le navire… j'aurais aimé que tu restes plus longtemps.
Antoine, mon compatriote. J'aurais peut être mieux fait de t'écouter il y a 4 ans. BREIZH ATAO !!

Allison je ne t'oublie pas ;)
Pierino non plus….

… Envers l'ensemble du 3e étage, Karine, Shahad, Giulia et les autres parce que vous êtes géniaux !!

… Envers Adam Hurlstone pour m'avoir aider à mettre en œuvre ce projet.

… Envers Emerald Perlas, Nicolas Servant, Virginie Dangles-Marie, Hélène Malka-Mahieu, Raghavendar Nagaraju, les gens de l'Histim et de la plateforme microscopie. Merci pour tout vos conseils, tout ce que vous m'avez appris et toute l'aide que vous m'avez apportées.

… Envers Armelle, Cédric, Tarek et Olivier. Merci d'avoir pris soin de mes bébés même s'ils étaient souvent moches.

… Envers Françoise, Marc et Mery, merci de m'avoir aider à reprendre pied.

… Envers les magistériens, Laure, Ludo, Christelle, Patrick, Chloé, Victoire et Tommaso, parce que nous allons y survivre et nous en ressortirons grandis.

… Envers les merveilleuses rencontres que j'ai faites pendant ces 5 années qui ont partagé mes peines et mes joies. Haser, Pauline, Andres, Julie, Joana, Romain, Thomas… promis je sors bientôt de mon coin !

… Envers mes amis de toujours. Anne laure, Julie et Mani, Kenan, Romi, Louis et Robin. Vous me manquez affreusement.

… Envers ma famille incroyable, on va fêter ça dignement !
… Envers mes sœurs Sarah et Chloé, même si nous sommes souvent comme chien et chat !
… Envers mes grands-parents, j'aurais aimé que vous soyez là
… Envers mes parents à qui je ne dis pas assez souvent que je les aime

… Pour Javier, merci d'y avoir survécu avec moi. Merci d'être toi.

Bientôt 5 ans,

5 années qui se terminent mais je n'oublierais jamais et dont je ne garderais que le meilleur.

TABLE OF CONTENTS

ABSTRACTS (English and French version) p.1
PREAMBLE p.7

Chapter 1: Conserved functions of rapidly evolving long noncoding RNAs **p.9**

INTRODUCTION

I. LncRNAs, as key players of gene regulation **p.9**

a. Noncoding RNAs challenge the central dogma of molecular biology p.9
b. What are lncRNAs? p.10
c. LncRNAs show multiple layers of conservation throughout evolution p.11
d. The functions of lncRNAs in gene regulation p.14

II. The *in vivo* function of lncRNAs **p.17**

a. LncRNAs have multiple developmental and physiological roles p.17
b. Misregulation of lncRNAs in cancer p.18
c. LncRNAs associated with melanoma progression p.19
i. Description of cutaneous melanoma
ii. LncRNAs as a new target for melanoma understanding and treatment
d. Zebrafish model to investigate melanoma progression p.24
i. Similarities between zebrafish and human skin
ii. Strategies to investigate melanoma in zebrafish
iii. Human melanoma-associated lncRNAs with conserved evolutionary features

III. Project objectives

RESULTS

I. Assessing the impact of lncRNAs on melanoma development: the lncRNA *menhir/CASC15* acts as a tumor suppressor in cutaneous melanoma **p.32**

a. A small-scale reverse genetic screen for melanoma progression p.32
b. Characterisation of lnc-sox4a/menhir mutant's and their melanoma-aggressiveness phenotype p.35
i. Characterisation of lnc-*sox4a/menhir* mutants
ii. Characterisation of lnc-*sox4a/menhir* mutants in $NRAS^{G12}$ melanoma
c. Determining the lnc-sox4a/menhir metastatic potential p.43
d. Rescue of the lnc-sox4a/menhir phenotype with the human ortholog CASC15 p.44
e. CASC15 expression profile in human cancer p.48
f. Melanoma induction triggers changes in skin pigmentation cells p.49

II. Determining the function of additional syntenic lncRNAs: linc-*myc* mutants exhibit multiple developmental defects **p.51**

DISCUSSION & PERSPECTIVES

I. Using zebrafish to investigate the *in vivo* function of lncRNAs in melanoma **p.56**

a. *Do rapidly evolving lncRNAs have conserved functions in cancer?* p.56
   i. *menhir,* a zebrafish melanoma tumor suppressor
   ii. *CASC15/menhir* conserved properties
   iii. The *CASC15* function in human cancer

b. *$NRAS^{G12}$ driven melanoma in zebrafish to analyse cancer-related functions of lncRNAs* p.59

c. Combining zebrafish therapeutic devices and RNA-based treatment to target cancer p.60
   i. Zebrafish as a devices to assess drug efficiency
   ii. RNA molecules as innovative therapeutics

d. *Outlook* p.62
   i. *menhir* melanoma invasiveness analysis
   ii. Identification of *CASC15/menhir* protein interactors to decipher lncRNA functionality
   iii. Incorporate *CASC15/menhir* role to cancer-related pathways
   iv. Determine *CASC15/menhir* differential function in cancer subtypes

II. Deciphering phenotype of lncRNA mutants **p.64**

a. *Lnc-myc functionality in zebrafish embryogenesis* p.65

b. *Outlooks* p.66

III. Synteny as a determinant of functional conservation? **p.67**

Chapter 2: A minimally invasive genome editing approach to inactivate lncRNAs in zebrafish **p.69**

INTRODUCTION **p.69**

I. Transient and stable strategies to inactivate lncRNAs *in vivo*

II. Insertion of RNA destabilizing elements as a minimally invasive approach to inactivate long noncoding RNA **p.71**

III. Achieving precise short sequence integration in the zebrafish genome **p.72**

IV. Overall objectives of the project **p.73**

RESULTS

I. CRISPR-Cas9 mediated genetic deletions as a strategy to inactivate lncRNAs in zebrafish **p.76**

a. Generation of zebrafish mutant lines for syntenic lncRNAs p.76

b. Phenotypic characterisation of long noncoding RNA zebrafish mutants: Characterisation of lnc-ppm1bb$^{-/-}$ p.77

II. Insertion of RNA Destabilisation Elements as a minimally invasive approach to inactivate lncRNAs **p.80**

a. Inhibition of the Non Homologous End Joining pathway to promote precise genomic insertion in zebrafish p.80

b. Technical challenges p.82

III. Assessing RNA destabilization efficiency **p.82**

DISCUSSION

I. Inactivating long noncoding RNA in vivo: finding the best strategy between transient and stable mutagenesis **p.87**

II. RNA Destabilising Element and premature polyA signal differential efficiency to inactivate lncRNA *in vivo* **p.88**

MATERIALS AND METHODS p.91

BIBLIOGRAPHY p.106

ABBREVIATIONS p.121

RESUME version longue française p.125

APPENDIX p.137

ABSTRACT

The identification of diverse target genes involved in cancer progression is crucial to decipher the mechanisms underlying cancer and to develop effective targeted treatment therapies. LncRNAs (long noncoding RNAs) are molecularly similar to messenger RNAs but lack protein-coding potential. Their deregulation and misexpression as well as the presence of mutations in lncRNA loci have been linked to diseases including cancers. Like protein-coding genes, vertebrate lncRNAs can be conserved at multiple levels (sequence, expression pattern, genomic position or synteny). In contrast to only 2% sequence conservation, more than 35% of zebrafish lncRNAs are conserved at the syntenic level to human, indicating an evolutionary pressure to preserve lncRNA position in the genome. To assess if positional conservation is a predictor of functional conservation, I implemented a reverse genetic screen assaying the role of lncRNAs in melanoma development using zebrafish, an optimal oncology model with multiple skin genetics, histological and physiological similarities with human.

Using CRISPR-Cas9 genome editing technology to generate syntenic lncRNA zebrafish loss of function mutants, I profiled the impact of lncRNA loss on melanoma induced by the human $NRAS^{G12}$ oncogene and in human melanoma cell xenografts. Among six candidate lncRNAs, we identified *menhir* (*ME*la*N*oma *HI*ndrance long noncoding *R*NA) as a melanoma tumor suppressor. *menhir* zebrafish mutants display impaired melanoma aggressiveness characterized by (1) accelerated tumorigenesis, (2) decreased mutant survival, (3) increased melanoma severity and (4) increased metastatic potential due to a higher permissiveness to melanoma cell invasion. To assess if the tumor suppressor function of zebrafish *menhir* is conserved throughout evolution, the human putative ortholog of *menhir* called *CASC15* (*Cancer Susceptibility 15*) was expressed in melanocytes of the zebrafish melanoma *menhir* mutant. Despite lack of sequence conservation, *CASC15* expression mitigated the mutant menhir melanoma phenotype as evidenced by reduced melanoma progression, decreased tumorigenesis and enhanced survival of zebrafish affected with melanoma.

Thus, our results identify a novel melanoma tumor suppressor lncRNA and show that conserved genomic location of lncRNAs can be used to posit functional conservation in vertebrates.

RESUME (Version courte)

L'identification de divers gènes cibles impliqués dans la progression cancéreuse est cruciale afin de décrypter les mécanismes sous-jacents au cancer et de développer des stratégies thérapeutiques efficaces. Les lncARN (longs ARN non codants) sont similaires aux ARN messagers d'un point de vue moléculaire, mais ne présentent pas de potentiel codant pour des protéines. Ils sont fréquemment dérégulés et mutés dans de nombreux types de cancers. Tout comme les gènes codants pour des protéines, les lncARN des vertébrés peuvent être conservés à plusieurs niveaux: séquence, profil d'expression ou position génomique (synténie). Seuls 2% des lncARN du poisson zèbre présentent une préservation de séquence avec l'homme, tandis que plus de 35% sont conservés au niveau synténique, indiquant la présence d'une pression évolutive préservant la position génomique des lncARN. Afin d'évaluer si ce phénomène synténique peut prédire la conservation fonctionnelle des lncARN, j'ai établi un criblage génétique inverse évaluant le rôle des lncARN dans le développement du mélanome. Ces études ont été effectuées chez le poisson zèbre, un modèle animal présentant de multiples similarités génétiques, histologiques et physiologiques avec la peau humaine.

En utilisant la technologie d'édition du génome CRISPR-Cas9 pour générer les lignées de poissons zèbres mutants pour une sélection de 6 lncARN candidats, j'ai mesuré l'impact de la perte de fonction de ces lncARN sur la progression du mélanome, induit chez le poisson zèbre via l'expression de l'oncogène humain $NRAS^{G12}$ et la xénogreffes de cellules de mélanome humain. Lors de cette étude, j'ai identifié *menhir* (*ME*la*N*oma *HI*ndrance long noncoding *R*NA) comme un gène suppresseur de tumeur dans le mélanome. En effet, les poisson zèbres mutants pour *menhir* présentent une altération de l'agressivité du mélanome caractérisée par (1) une augmentation de la tumorigenèse, (2) une baisse de la survie, (3) une augmentation de la sévérité du mélanome et (4) une augmentation du potentiel métastatique due à une plus grande permissivité à l'invasion des cellules du mélanome. Afin d'analyser si la fonction anti-oncogène de *menhir* est conservée dans l'évolution, nous avons exprimé l'homologue humain *CASC15* (*Cancer Susceptibility 15*) dans les mélanocytes des poissons zèbres mutants pour menhir affectés par le mélanome. Malgré l'absence de conservation de séquence, l'expression de *CASC15* atténue le phénotype d'agressivité du mélanome des poissons mutants pour menhir, entrainant une diminution de la progression du cancer, de la tumorigenèse et une amélioration de la survie des individus mutants affectés par le mélanome.

Par conséquent, mes résultats identifient un nouveau lncARN suppresseur de tumeur dans le mélanome et montrent que la conservation de position génomique peut être corrélée avec une conservation de fonction.

ABSTRACT (Mainstream)

Long noncoding RNAs (lncRNAs) are genes that do not encode for proteins and are often misregulated in human cancers, such as melanoma, a skin cancer responsible for the death of 50 000 people every year. However, the precise roles of lncRNAs during cancer development and progression remain largely elusive. Zebrafish is an experimental model that can mimic closely human melanoma progression through expression of a mutated transgene or grafts of human melanoma cells.

During my thesis, I have profiled the impact of several lncRNAs on melanoma progression at the organismal level and identified a lncRNA with a tumor suppressor function. Indeed, melanoma is more aggressive in zebrafish that lack this lncRNA. Furthermore, expression of the human putative ortholog lncRNA in the zebrafish mutant slows down cancer progression, demonstrating that the role of this lncRNA is conserved throughout evolution.

Taken together, my study identified the first lncRNA with the in vivo tumor suppressor function in melanoma development presenting a promising therapeutic target for melanoma treatment.

RESUME (Grand public)

Les longs ARN non codants (lncARN) sont des gènes ne codant pas pour des protéines qui sont fréquemment dérégulés dans les cancers humains, comme le mélanome, un cancer de la peau responsable du décès de 50 000 personnes par an. Toutefois, le rôle précis des lncARN dans le développement du cancer reste méconnu. Le poisson zèbre est un modèle d'étude qui permet de reproduire fidèlement la progression du mélanome grâce à des mutations ou des greffes de cellules de mélanome humain.

Lors de ma thèse, j'ai analysé l'impact de plusieurs lncARN sur le développement du mélanome à l'échelle de l'organisme entier et identifié un lncARN ayant une fonction de suppresseur de tumeur. En effet, le mélanome est plus agressif chez les poissons zèbres mutants n'exprimant pas ce gène. J'ai également observé que l'expression du lncARN humain chez le poisson mutant ralentit la progression du cancer, démontrant que le rôle de ce gène est conservé dans l'évolution.

Ce projet a donc permis de découvrir un nouveau gène suppresseur de tumeur ayant un fort potentiel thérapeutique pour le traitement du mélanome.

PREAMBLE

Long noncoding RNAs (or lncRNAs) have emerged as dynamic players of gene regulation. From *Drosophila* to human, several lncRNAs have been reported to modulate gene expression at both the transcriptional and post-transcriptional level. Although the important roles of a handful of lncRNAs have been shown *in vivo*, the function of the majority of lncRNAs remains elusive. We recently demonstrated that sequence conservation throughout evolution, although limited to less than 5% of vertebrate lncRNAs from zebrafish to mammals, can be used to identify functional lncRNAs among the thousands of annotated vertebrate lncRNA genes. However, compared to sequence conservation, a much larger cohort of vertebrate lncRNA genes is conserved at the level of genomic position, also known as synteny.

The broad aim of my thesis was to investigate if syntenic conservation could be used as a determinant of vertebrate lncRNA functional conservation. Using zebrafish as a vertebrate model, I report in Chapter 1 the characterization of the *in vivo* roles of a set of lncRNAs in development and disease using a collection of zebrafish lncRNA mutants I generated using CRISPR/Cas9 genome editing. Using melanoma as a phenotypic read-out, I demonstrate that the zebrafish *menhir*/human *CASC15* lncRNA reduces melanoma aggressiveness, thus reporting the first melanoma tumor suppressor lncRNA characterized *in vivo*. I also show that the disruption of the two zebrafish *PVT1* lncRNA syntenic homologs leads to developmental defects and demonstrate a compensatory mechanism for their action. In Chapter 2, I detail the multiple approaches I developed to inactivate lncRNA genes in zebrafish and elaborate on the innovation of a minimally invasive method to inactivate lncRNAs *in vivo*.

Together, my work has deepened our functional knowledge of a set of syntenic lncRNAs, revealing the first *in vivo* vertebrate lncRNA tumor suppressor, and has shown that synteny can indeed be used to inform conserved function among vertebrate lncRNAs.

Chapter 1: Conserved functions of rapidly evolving long noncoding RNAs

INTRODUCTION

I. Long noncoding RNAs as key players of gene regulation

a. Noncoding RNAs challenge the central dogma of molecular biology

Five years after establishing the double helix structure of DNA, Sir Francis Crick postulated the basis for molecular biology's central dogma: "DNA makes RNA makes protein" (Crick 1966). Together with his Sequence hypothesis, Crick suggested that nucleic acids serve no purpose other than to code for Amino-Acids and that proteins are the essential component of life. RiboNucleic Acids (RNA) had been suggested to be functional by Jacob and Monod in the early 60s, even though the RNA molecule was not yet biochemically identified (Jacob and Monod, 1961). The emergence of the G paradox (the inconsistency between the number of protein-coding genes and the level of organism complexity) and the discovery that a significant part of the genome does not code for proteins ("originally termed junk DNA"), together suggested an additional, nonprotein-coding purpose for nucleic acids (Riddihough, 2005). More recent advances in whole genome sequencing methods have revealed that although 70 to 90% of the genome is pervasively transcribed, only 1,5-2% of the human genomic sequence code for exons of protein-coding genes, (Derrien et al., 2012; Necsulea et al., 2015; Ulitsky et al., 2011; Glusman et al., 2006), delineating clearly the breadth of the noncoding RNA world.

Noncoding RNA genes (ncRNAs) produce a diversity of transcripts that can be grouped into several families according to their size (Amaral and Mattick, 2008; Pauli et al., 2011) : the small ncRNAs (<50 nucleotides), the intermediate ncRNAs (between 50 and 200 nucleotides) and the long noncoding RNAs (lncRNAs) that are longer than 200 nucleotides (nt). The biological importance of many small and intermediate ncRNAs has been shown in fundamental cellular processes, such as the role of ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) in protein translation and the gene regulatory function of microRNAs (miRNAs). However, the role of the majority of lncRNAs remains elusive.

b. What are long noncoding RNAs?

Long noncoding RNAs (lncRNAs) are a large and heterogeneous family of trancripts that, similar to messenger RNAs, are transcribed by RNA polymerase II, capped, spliced and poly-adenylated (Ulitsky and Bartel, 2013). They, however, have little or no protein-coding potential (Iyer et al., 2015) and, as such, are often not predicted as genes through classic genomic transcription footprinting investigations (Glusman et al., 2006).

It has been proposed that lncRNAs originate from multiple mechanisms such as pseudogenization of protein-coding genes, gene duplication, bidirectional transcription and *de novo* formation derived from Transposable Elements (TE) (Kapusta and Feschotte, 2014; Ulitsky, 2016). Transposable Elements are characterised by their mobility and their ability to modify the genetic environment of their insertion locus, making them a major source of evolutionary variability (Hua-Van et al., 2011). Indeed, a large proportion of human lncRNAs are TE derived (Ulitsky, 2016). This phenomenon has been observed in other vertebrates, as TEs are present in the majority of vertebrate lncRNAs, with some lncRNAs such as UCA1 being entirely composed of TEs (Ganesh and Svoboda, 2016; Wang et al., 2008). TEs influence lncRNAs in multiple ways, from acting as tissue-specific promoters (Davis et al., 2017) to contributing to secondary structure formation (Kapusta et al., 2013), or promoting lncRNA plasticity and sequence diversification (Ganesh and Svoboda, 2016).

A common feature of lncRNAs across a variety of organisms is their relative low expression levels and high organ and tissue specificity (Kaushik et al., 2013; Necsulea et al., 2015). The differences in genetic and epigenetic signatures of protein-coding and lncRNA genes, such as differential CpG enrichment and transcription factor binding sites (TFBSs) in the gene promoters (Alam et al., 2014), contribute to the low and tissue-specific expression of lncRNAs. Despite their low expression, lncRNA transcription is actively regulated. Indeed, there are several reports of regulatory loops between lncRNAs and important transcription factors (such as the SOX or HOX families) (Alam et al., 2014). Moreover, TFBS sequences are generally more conserved throughout evolution in lncRNA promoters compared to those in intergenic regions and even in some protein-coding sequences (Necsulea et al., 2015).

Although no formal functional classification has been established for lncRNA loci, they can be subdivided in categories according to their genomic location (bidirectional; sense or antisense to protein-coding genes; intronic; intergenic) (Mallory and Shkumatava, 2015). In addition, it has been observed that lncRNAs are often transcribed as multiple isoforms, with alternative Transcription Start Sites (TSSs), several splicing variants and alternative

cleavage and polyA signals (ATAs) (Ziegler and Kretz, 2017) from scattered genomic loci that overlap with other coding/noncoding genes or DNA regulatory motifs. Thus, the complexity of the genomic locus and the identity of the transcripts produced both have to be considered when aiming to inactivate a lncRNA.

c. LncRNAs show multiple layers of conservation throughout evolution

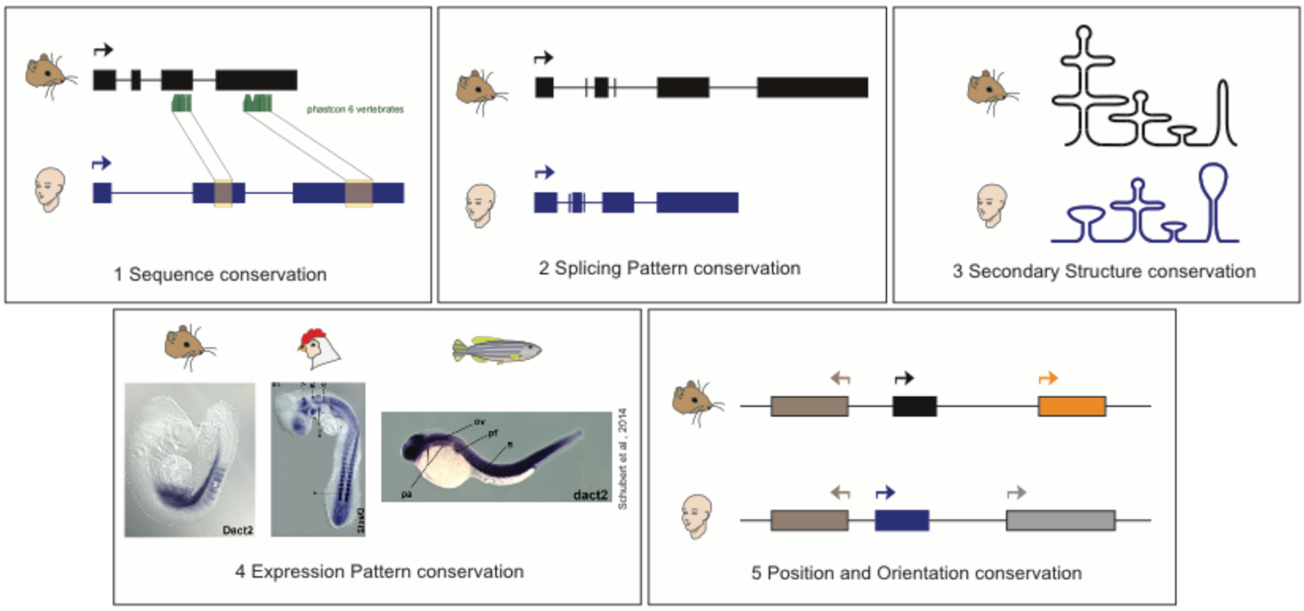


Figure 1: The multiple layers of lncRNAs conservation

LncRNAs can be conserved at different levels including sequence conservation, Phastcons plots (green) relative to the mouse locus are based on a 6-genome alignments and indicate the location of the conserved sequences (1), RNA splicing pattern and exon/intron proportion preservation (2), similar RNA expression pattern in different organisms (Pictures from Schubert et al, 2014) (3), conservation of the secondary structure of the mature transcript (4), and syntenic conservation (lncRNA genomic position and transcriptional orientation relative to adjacent genes indicated in brown, orange and grey) (5).

For protein-coding genes, functionality is often predicted and confirmed through the analysis of conserved evolutionary features. Although lncRNAs have been shown to evolve more rapidly than protein-coding genes (Kapusta and Feschotte, 2014), the same predictive strategies can be informative for noncoding transcripts. Indeed, both genic and transcriptomic analyses from multiple vertebrates have highlighted the conserved features of several lncRNAs illustrated in Figure 1 (Amaral et al., 2018; Hezroni et al., 2015; Necsulea et al.,

2015; Ulitsky et al., 2011). For example, the transcription status and splicing pattern of lncRNAs are often conserved at levels higher than for intergenic DNA (Hezroni et al., 2015; Ulitsky, 2016). In addition, despite low sequence conservation throughout the length of lncRNAs, although rare, they can have short patches of sequences conserved between organisms of the same clade, principally localised to exons and splice sites (Ulitsky, 2016). This observation together with the sequence polymorphisms and genetic drift observed for long intervening noncoding RNAs within the human population, indicate that lncRNA sequences are under generally low evolutionary constraints compared to protein-coding genes (Haerty and Ponting, 2013; Kapusta and Feschotte, 2014). These results suggest that additional features contribute to lncRNA function and lend to the notion that a fraction of lncRNAs has species-specific functions or may not be functional.

It is reasonable to hypothesize that the biochemical activity of a fraction of lncRNAs may depend on their 3D structure rather than their nucleic acid sequence. Indeed RNAs are highly dynamic and motile molecules, and some lncRNAs display conserved secondary structure. For example, the *Drosophila roX* transcript displays secondary structure preservation, as both *Drosophila melanogaster* isoforms *roX1* and *roX2* share loop structures ((Ilik et al., 2013) that are conserved with other *Drosophila* species separated by 40 million years (Quinn et al., 2016). However the level to which conserved lncRNA structural features contribute to their function is largely unknown primarily because existing RNA structure prediction tools are noisy and often generate false positives. In addition the *in vitro* techniques to analyse RNA secondary structure, such as SHAPE (Selective 2'-hydroxy acylation analysed by primer extension), lack power when nothing is known about the transcript 3D structure (Kapusta and Feschotte, 2014)..

As discussed earlier, many lncRNAs have highly specific expression patterns with a predominant expression in the germ line (Ulitsky, 2016). It has also been reported that the specific expression pattern of some lncRNAs is conserved throughout evolution (Chodroff et al., 2010), particularly among primates in which almost half of the investigated lncRNAs display tissue expression conservation (Necsulea et al., 2015).

Another level of lncRNA conservation is the preservation of genomic position, also known as synteny, and relative transcriptional orientation. The first syntenic lncRNA described was mammalian *XIST*, which is conserved from human to mouse (Brockdorff et al., 1991). Twenty years later, prevalent positional conservation from zebrafish to humans was reported for a set of lncRNAs (Ulitsky et al., 2011). This set was further refined by the Ulitsky lab (Figure 2), showing that 35% of zebrafish lncRNAs were conserved to human at

strictly the syntenic level (compared to only 2% at the sequence level) (Hezroni et al., 2015; Ulitsky et al., 2011). Moreover, it has been reported that mammalian syntenic lncRNAs have their position frequently associated with chromatin loop anchors and CTCF binding sites (Amaral et al., 2018), suggesting that these loci may contribute to genome organizational features. Furthermore, additional examples of syntenic lncRNA conservation have been reported among plants (Mohammadin et al., 2015) and among different *Drosophila* species (Quinn et al., 2016).

Altogether, these five conservation layers can be used to inform lncRNA functional conservation predictions between organisms. These predictions can then be tested at the organismal or cellular level by examining the phenotypic and molecular consequences of lncRNA depletion or loss and by performing rescue experiments with syntenic orthologs. Conservation can also be assayed by deciphering the molecular mechanism of action and target gene repertoire of lncRNAs among different organisms.

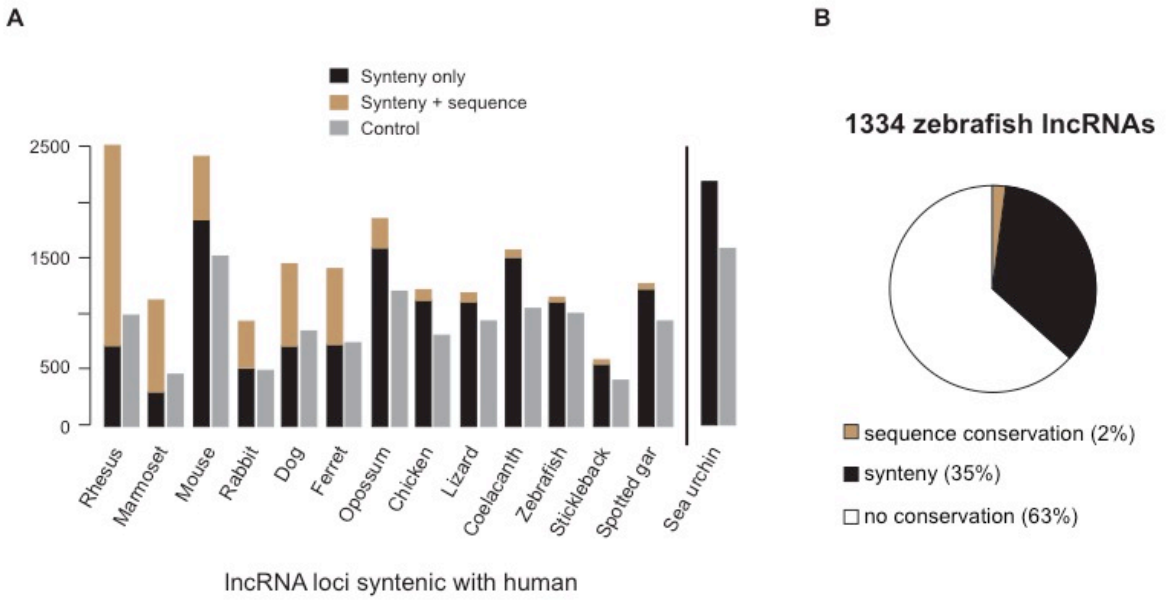


Figure 2: lncRNA are more conserved at the synteny than sequence level (adapted from Hezroni et al, 2015)

A Number of lncRNA loci from different species syntenic (with or without sequence conservation) with human lncRNA loci. Control numbers were generated by randomly placing the human lncRNAs in intergenic regions, repeating the analysis ten times, and averaging the number of observed syntenic relationships. **B** Number of lncRNAs annotated in zebrafish, and the conserved proportion between zebrafish and human.

d. The functions of lncRNAs in gene regulation

Although the role of most lncRNAs remains elusive, their initial functional investigation started in the 90s and early 2000s with the analysis of the X inactivation related lncRNAs *Xist* and *Tsix* (Maharens et al., 1997, Sado et al., 2001) and the imprinted genomic loci associated lncRNAs *Airn* and *Kcnq1ot1* (Fitzpatrick et al., 2002; Sleutels et al., 2002), establishing these chromatin-associated lncRNAs as "a rosetta stone for recent long noncoding RNAs" (Nakagawa, 2016). Numerous possible functional mechanisms in gene regulation inherent to the RNA molecule exist (Figure 3), each varying according to their DNA, RNA or protein interacting partners. For example, when binding directly to DNA, lncRNAs can participate to the formation of chromosome looping to induce short-range interactions, such as R-loops (RNA-DNA hybrid genomic structures formed during transcription) that regulate positively or negatively the adjacent protein-coding genes (Villegas and Zaphiropoulos, 2015); or long-range interactions such as the formation of RNA/DNA triplexes facilitating the recruitment of chromatin modifier complexes (Grote et al., 2013; Joung et al., 2017; Lai et al., 2013; Werner et al., 2017). Indeed lncRNAs can also guide chromatin modifiers or DNA methylase proteins to activate (Grote et al., 2013) or repress (Ding et al., 2016; Marín-Béjar et al., 2013; Yap et al., 2010) specific genes. For example, *ANRIL* interacts with PRC1 to promote H3K27 tri-methylation of the adjacent INK4 locus, inhibiting major tumor suppressor genes (Yap et al., 2010), whereas *Fendrr* recruits methyl-transferase to promote cardiac gene activation in embryogenesis (Grote and Herrmann, 2014; Grote et al., 2013).

By interacting with other RNA molecules, lncRNAs can alter mRNA structure, like in the case of the lncRNA *MALAT1,* which modulates mRNA alternative splicing (Tripathi et al., 2010a), methylation status and subcellular localization (Yoon et al., 2013). LncRNAs can also form mRNA-lncRNA duplexes to prevent mRNA decay by protecting the transcript from miRNA mediated cleavage (Faghihi et al., 2008; Guil and Esteller, 2012). *SPRIGHTLY* pseudoknotted secondary structure has also been reported to stabilize cancer related genes by interacting with their premature mRNA intronic sequences (Lee et al., 2017). However certain RNA/RNA interactions can rather promote mRNA destabilization as the STAUFEN1 protein degrades RNA duplexes (Kim et al., 2005). LncRNAs can also interact with other noncoding RNAs such as microRNAs (miRNAs). For example, the muscle specific linc-*MD1* presents multiple miRNA binding sites to miR133/miR135 and has been reported to compete with miR133/miR135 target genes and to attenuate miRNA repression mechanisms (Cesana et al.,

2011). In addition, recent publications report that the noncoding portion of the transcript *Nrep* and the lncRNA *Cyrano* modulate levels of mature miR-29 and miR-7, respectively, via target-directed miRNA degradation (Bitetti et al., 2018; Kleaveland et al., 2018).

lncRNAs have also been shown to interact with proteins and modulate many of their biological properties, such as their localization, stabilization, decoy, scaffolding and enzymatic activities (Figure 3). Indeed, lncRNAs like *SAMMSON* (Leucci et al., 2016) and *MALAT1* (Tripathi et al., 2010) regulate subcellular localization of mitochondrial associated p32 protein and splicing regulators, respectively. *CASC15* has recently been reported as a modulator of USP36 intranuclear localization, consequently affecting the ubiquitination and degradation of the CHD7 transcriptional co-factor (Mondal et al., 2018). Moreover, lncRNAs often guide, recruit and/or promote the scaffolding of transcriptional regulators or chromatin modifyers (Ding et al., 2016; Marín-Béjar et al., 2013; Mohammad et al., 2010; Schmidt et al., 2016). LncRNAs can also cover protein functional domains, accountable for proteins phosphorylation-mediated inactivation (*PVT1*) or binding of specific DNA target (*GAS5*), therefore mediating protein stabilization (Tseng et al., 2015) or decoy (Kino et al., 2010). Protein enzymatic activity and biological function can also be impacted by its interaction with lncRNAs, such as IGF2BP1, which mRNA stabilization properties are promoted by the oncogenic lncRNA *THOR* (Hosono et al., 2017); and DNA helicase DDXII, which presents increased enzymatic and DNA binding capacities upon interacting with the *CONCR* transcript (Marchese et al., 2016).

Thus, lncRNAs, through their highly dynamic and tissue specific properties, act at several levels of gene regulation to promote precise, fast and low energetic cellular answers to diverse stimuli or biological processes (Marchese et al., 2017).

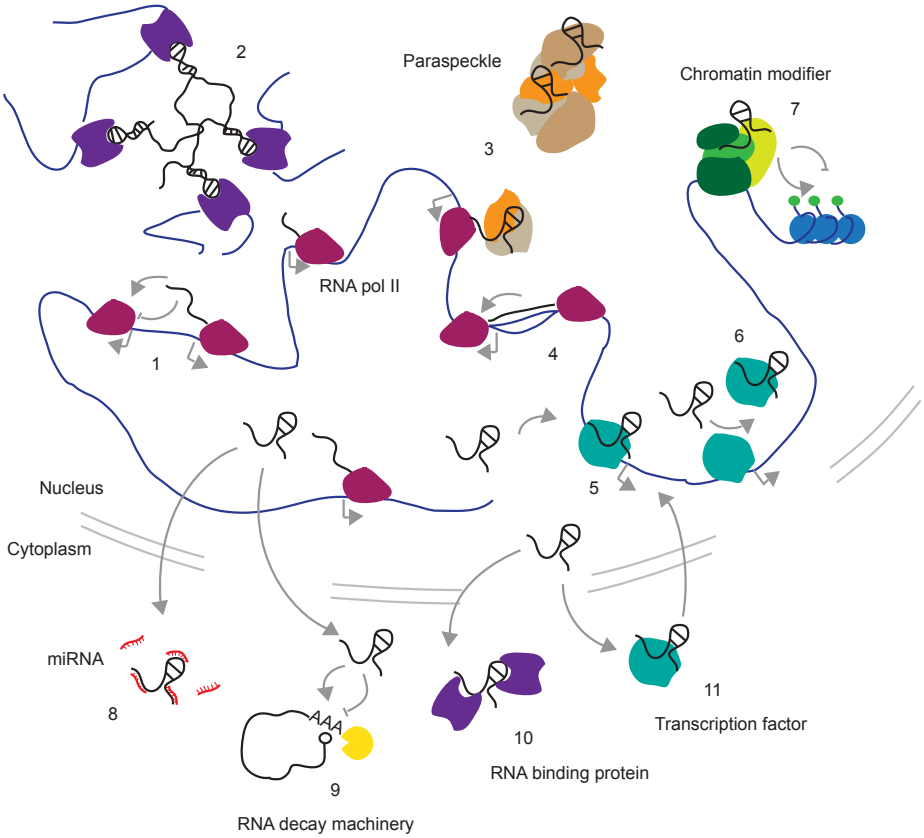


Figure 3: Diverse mechanisms employed by lncRNAs to regulate gene expression. (adapted from Marchese, Raimondi and Huarte, 2017)

Nuclear specific mechanisms such as lncRNA transcription-dependent activation or repression of neighbour genes (1), lncRNA-mediated inter-chromosomal interactions (2), formation of paraspeckles/nuclear speckles (3) or R-loops (4), lncRNAs as guides (5) or decoys (6) of transcription factors or as scaffolds for chromatin modifying complexes (7), and cytoplasmic specific mechanisms such as lncRNAs acting as miRNA sponges (8), regulating post-transcriptional mRNA decay (9), or impacting the cellular localization of RNA-binding proteins (RBPs) (10) or DNA-binding proteins (DBPs) (11).

II. The *in vivo* function of lncRNAs

a. LncRNAs have multiple developmental and physiological roles

Recently, more than 7 million diseases associated with single nucleotide polymorphisms (SNPs) have been reported to be mapped within lncRNA loci (Miao et al., 2017). There is however a weak correlation between abundant expression and phenotype related to lncRNA inactivation (Ana Rita Amândio, 2016; Eißmann et al., 2012; Nakagawa et al., 2012; Nakagawa, 2016; Zhang et al., 2012). Indeed, due to their highly restricted expression, global high throughput analysis of lncRNA mutants might not be informative enough to decrypt subtle phenotypes (Bassett et al., 2014), such as the retinal vascularisation defects in *Malat1* knock-down (Michalik et al., 2014) or the hormonal defects and decrease in milk producing cells in *Neat1* knock-out (Nakagawa et al., 2014; Standaert et al., 2014).

Examples of animal models presenting deleterious developmental, physiological or behavioural phenotypes due to lncRNA inactivations are limited (Bassett et al., 2014; Perry and Ulitsky, 2016, Bitetti et al 2018). A set of lncRNAs has been reported to be essential for mouse development, although lethality is sex specific for two of them (*Xist* and *Tsix* (Maharens et al., 1997; Sado et al., 2001)) and partially penetrant for *Peril*, *Mdgt* (Sauvageau et al., 2013) and *Neat1* (Standaert et al., 2014). Growth defects were also observed in mouse mutants for lncRNAs *Kcnq1ot1* (Mohammad et al., 2010), *Airn* (Sleutels et al., 2002), *Neat1* (Standaert et al., 2014), *Lincpint* (Lai et al., 2015), and *Mdgt* (Sauvageau et al., 2013).

Fendrr (Fetal lethal non-coding developmental regulatory RNA) is a mammalian syntenic lncRNA whose loss leads to embryonic or perinatal fully penetrant lethality. In 2013, Grote et al. characterised *Fendrr* expression in lateral plate mesoderm and its impact on embryogenesis through generation of knock-down and knock-out mouse mutants. Although shRNA knock-down individuals do not present any noticeable developmental defects, *Fendrr* null allele mutants obtained with insertion of premature polyA signal resulted in fully penetrant lethality at stage E13.75, due to ventricular and intra-ventricular myocard hypoplasia along with omphalocele (ventral body wall thickness defect) (Grote et al., 2013). Importantly, the lethality and developmental defects could be rescued through the introduction of a *Fendrr* containing BAC clone, confirming that the observed phenotypes were due to loss of *Fendrr* (Grote et al., 2013). Concomitantly, Sauvageau et al described *Fendrr* broad expression pattern in adult tissue and observed perinatal lethality of knock-out

mutants obtained through genomic replacement with lacZ reporter gene (Sauvageau et al., 2013). *Fendrr* homozygous mutant pups died of respiratory failure due to lung hypoplasia and intraventricular septal heart defects, but Sauvageau et al did not observed any protruding omphalocele (Sauvageau et al., 2013). If *Fendrr* appears to be vital for mouse embryonic and perinatal survival, its biological function in human has mainly been investigated in the cancer field. Indeed, *FENDRR* expression appears to be correlated with good prognosis and reduced malignancy of several cancers (Li et al., 2018; Kun-Peng et al., 2017; Xu et al., 2014).

LncRNAs have been described to be often expressed in neuronal tissues (Chodroff et al., 2010; Kaushik et al., 2013), however only a few examples present characterization at the organismal level. Lnc-*Brnb1* and *Pnky* are among the few transcripts for which precise neuronal defects have been described in mutants; lnc-*Brnb1* knock-out being characterised by a reduction of neuronal intermediate progenitor cells and abnormal cortical lamination (Sauvageau et al., 2013), whereas *Pnky* knock-down is characterised by increased neuronal differentiation and a depletion of neural stem cells (Ramos et al., 2015). Other lncRNAs mutants, such as *Peril* knock-out, present differential expression of several neurogenesis markers, however precise lncRNA *in vivo* functions have not yet been identified. Nevertheless, *Peril* partially penetrant lethality is supposed to be associated with brain development (Sauvageau et al., 2013). Finally, lncRNAs do not only impact organism morphology, as it has been reported that zebrafish *libra* lncRNA mutants present altered behavior, which is conserved in mouse ortholog *Nrep* mutants (Bitetti et al., 2018).

b. Misregulation of lncRNAs in cancer

In addition to their known roles in development and physiological conditions, lncRNAs are extensively investigated in oncology. Indeed, multiple lncRNAs are misregulated in cancer and 11.75% of cancer SNPs are localised in lncRNA loci (Yan et al., 2015). Several reviews report multiples lncRNAs associated with several types of cancer (Aftab et al., 2014; Bhan et al., 2017; Richtig et al., 2017). LncRNAs even have a high potential as cancer biomarkers or predictive outcome when detected in different patient samples (urine, plasma, serum, whole blood, gastric juice, bone marrow and tumor) (Prensner and Chinnaiyan, 2011; Bhan et al., 2017). In 2013, the expression of 10207 transcripts was profiled in 1300 tumors including prostate and ovarian cancers, glioblastoma and lung squamous cell carcinoma. These profiles are now used to predict lncRNAs tumor promoting

or suppressing functions (Du et al., 2013). The cancer treatment field is always seeking for new therapeutic targets to propose the best personalized treatment possible to each patient. LncRNAs have a high potential for therapeutics, as it has been reported that changes in lncRNA expression or sequence can alter cancer cells sensitivity to chemotherapy (Bhan et al., 2017). Their therapeutic potential is currently tested through transcript destabilizations (by siRNA, ASO, Gapmer of RNA destabilisation elements), alterations of lncRNA promoter activity, or lncRNA-small molecule interactions and functional disruptions by aptamer binding (Lavorgna et al., 2016; Matsui and Corey, 2016; Bhan et al., 2017).

c. LncRNAs associated with melanoma progression

i. Description of cutaneous melanoma

Human skin, which represents 15% of the human body weight, can be affected by three different types of cancer: squamous cell carcinoma and basal cell carcinoma that affect skin keratinocytes, and melanoma (Hombach and Kretz, 2013). Melanoma is the cancer of melanocytes, neural crest derived cells localised in the basilar epidermis, which produce and distribute melanin pigments (see Figure 4). Melanoma can be divided in three subtypes: mucosal melanoma, uveal melanoma and cutaneous melanoma (either chronically sun damaged or not). Melanocytic neoplasm ranges from benign lesion (melanocytic naevus), Stage I dysplasic naevus, stage II melanoma *in situ*, and invasive melanoma characterised by stage III metastasis deposit in local lymph node and stage IV distant metastasis (Shain and Bastian, 2016). If stage I/II melanoma can be easily treated by surgical approaches, metastatic melanoma is highly lethal and its incidence increases by 3.7% each year (Sandru et al., 2014). Invasive melanoma severity and outcome is variable according to the metastatic site, patients affected by brain melanoma metastasis have a 7-fold higher risk of death than patients with lung or lymph node metastatic melanoma.

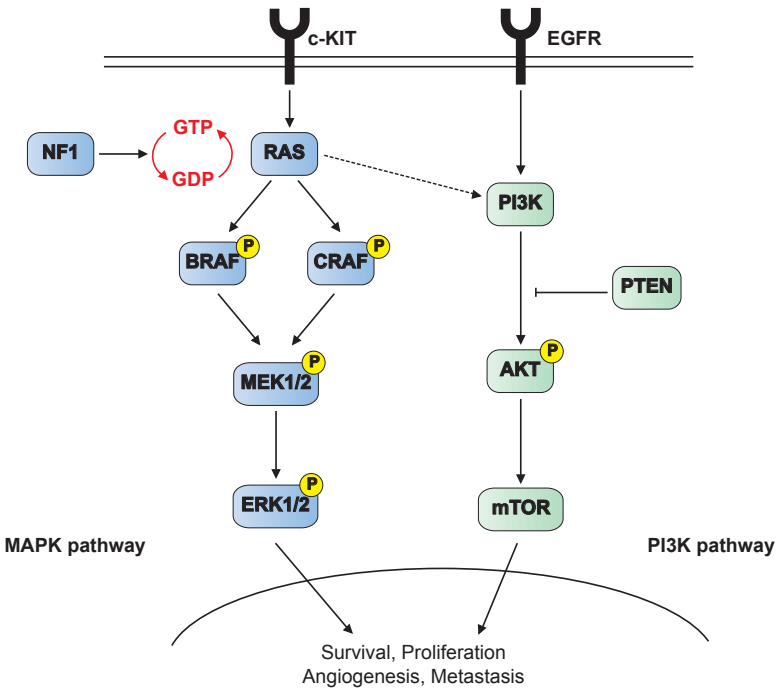
The great majority of cancers originates from an ectopic activation of the MAPK pathway, characterised by (1) activation of RAS (NRAS, KRAS, HRAS) proteins by the growth receptor c-met or c-kit or by activating mutations, (2) activating phosphorylation of the downstream mediators BRAF or CRAF, (3) phosphorylation and activation of MEK, then (4) ERK proteins leading to (5) transcriptional activation of pro-growth signals (Figure 4A) (Sullivan and Flaherty, 2012). MAPK activating mutations that drive melanoma present different incidences (Figure 4B). Melanoma mutational status (Figure 4B) has been described in The Cancer Genome Atlas database (TCGA), which regroups the RNAseq analysis of 333

melanoma primary (20%) and metastatic tumors (80%) with their associated peripheral blood and 14 other tissue (Akbani et al., 2015).

NRAS was the first melanoma oncogene identified in 1984 (Albino et al., 1984). NRAS and BRAF mutant melanoma represent almost 80% of melanoma (Figure 4B), even though intra-tumoral mutation status are often heterogeneous (Fedorenko et al., 2012). Stage IV BRAF mutant melanoma and stage I to IV NRAS mutant melanoma present poorer prognosis than Triple wild types melanoma (Sullivan and Flaherty, 2012). If BRAF mutant melanomas are sensitive to MAPK and MEK inhibiting drugs. NRAS mutant melanomas present unpredictable effects when treated with MAPK inhibitors , as ¾ of patients develop resistance to treatment and over-activation of the MAPK pathway (Sullivan and Flaherty, 2012).

Among NRAS mutations, there is a majority of alterations of the glutamine 61 (88,1%) impairing NRAS intrinsic catalytic activity, or of the guanine 12 (4,9%) where NRAS become insensitive to GTPase deactivation (Grill and Larue, 2016). It has been recently reported that $NRAS^{Q61}$ and $NRAS^{G12}$ present differential phosphorylation and activation of target proteins. In cell lines, $NRAS^{G12}$ clusters with PIM2 and PI3K/AKT pathways, whereas $NRAS^{Q61}$ interacts with CK2α and MAPK pathway (Posch et al., 2016). Indeed $NRAS^{Q61}$ melanoma cell lines are more sensitive to CX4945 treatment, a drug targeting CK2α proteins. This differential interaction could not be reported in human patients, probably because of low number of $NRAS^{G12}$ patients data available (Posch et al., 2016).

A



B

Melanoma mutational status	Incidence
BRAF (D594, L597, V600)	40-52%
NRAS (Q61, G12 G13)	15-28%
KRAS (G12)	2%
HRAS (G12)	1%
NF1	14%
Triple WT	14%

Figure 4: MAPK and PI3K molecular pathways and melanoma mutational status and incidence

A Schematic of the Mitogen-Activated Protein Kinase (MAPK) and PhosphoInositide-3-Kinase (PI3K) and their activation of transcriptional program driving melanoma progression. **B** Melanoma mutation status and incidence in human patients (Akbani et al., 2015).

ii. LncRNAs as a new target for melanoma understanding and treatment

One of the major characteristics of lncRNAs is their highly tissue-specific expression profile (Kaushik et al., 2013; Necsulea et al., 2015) that can be also observed at the cellular level within a tissue. Indeed skin lncRNA repertoires are different in the two cell types composing the epidermis. LncRNAs *ANCR* and *TINCR,* respectively, regulate epidermis maintenance and differentiation of keratinocytes progenitor cells (Kretz et al., 2012; Kretz et al., 2012). However, these two transcripts are not reported to be expressed in melanocytes nor to be misregulated in human melanoma patients.

Currently, 17 lncRNAs have been reported to be associated with melanoma in human patients (Table 1). Melanoma lncRNA functions were mainly analysed in human cell lines (CM Cutaneous Melanoma, UM Uveal Melanoma), except for *ANRIL* (Cunnington et al., 2010; Yap et al., 2010), *SNGH5* (Ichigozaki et al., 2015) and *SPRIGHLTY* (Khaitan et al., 2011; Liu et al., 2016; Mazar et al., 2014; Zhao et al., 2016; Lee et al., 2017) that were assessed in patient samples and plasma. Among lncRNAs with a potential function in

melanomagenesis, *THOR* is the only one analysed in model organisms through genetic deletion (Hosono et al., 2017), whereas genetic deletions have been also generated for *SLNCR1* (Schmidt et al., 2016) and *SPRIGHTLY* (Mazar et al., 2014) in human cell lines. All other melanoma associated-lncRNAs (Flockhart et al., 2012; Li et al., 2014) (Leucci et al., 2016; Chen et al., 2016; Tang et al., 2013; Goeder et al., 2016; Wu et al., 2018) were inactivated by transient knock-down approaches. Among the 17 reported candidates, 14 lncRNAs are upregulated in melanoma. *PAUPAR* is the only melanoma lncRNA reported to be downregulated in melanoma (Ding et al., 2016), whereas *CASC15* expression level differs between cutaneous (up-regulated) and uveal melanomas (down-regulated) (Lessard et al., 2015; Xing et al., 2017). Recently reported to be associated with melanoma, *EMICERI* expression profile has not yet been characterised in melanoma human patients (Joung et al., 2017). Among the lncRNAs listed in the Table 1, only four show conservation to zebrafish, *THOR* (Hosono et al., 2017) and *MALAT1* (Tian et al., 2014; Luan et al., 2017) presenting extended patches of sequence conservation, and *PVT1* (Chen et al., 2017) and *CASC15* (Lessard et al., 2015; Xing et al., 2017) are syntenic lncRNAs with short stretches of sequence conservation (Ulitsky lab; unpublished data).

Table 1: LncRNAs currently reported to be associated with melanoma in human patients (updated from (Aftab et al., 2014); (Richtig et al., 2017))

lncRNA	Model of investigation	Mutagenesis	Expression	Conservation	Role of lncRNA	References
ANRIL	Patient samples, Cell line	Antisens transcript	Upregulated	Sequence conservation with mouse	SNP in 9P21 region increases melanoma risks. *ANRIL* represses the transcription of CDKN2A/B which leads to cell cycle, migration and colony formation perturbation.	Yap et al, 2010; Cunnington et al, 2010; Xu et al, 2016
BANCR	human CM cells A-375, 293, SK-MEL-5, 1205Lu, UACC903, CHL-1 mouse xenograft	shRNA	Upregulated	No conservation	High levels of BANCR lead to increased migration (by targeting CXCL11) and proliferation. BANCR expression directly correlates with tumor stage and indirectly with patient survival.	Flockart et al, 2012; Li et al, 2014
CASC15	human and mouse CM cells WP, M16, pMel NRAS(G12D), RKTJ-CB1 mouse xenograft human UM cells MUM2B, CCM1	siRNA, overexpression	Upregulated Downregulated	Syntenic with zebrafish	CASC15 expression correlates with tumor stage in cutaneous melanoma. It appears to regulate melanoma phenotype switch between proliferative and invasive stages. In uveal melanoma, CASC15 is anticorrelated with cancer progression through XIST activation	Lessard et al, 2015; Yue et al, 2017
EMICERI	human CM cells A375	X	X	X	EMICERI down-regulation leads to melanoma cells increased resistance to BRAF inhibitor drug vemurafenib through *cis* regulation of MOB3B kinase activator of the Hippo pathway	Joung et al, 2018
GAS5	human CM cells A375, SK-MEL-110, SK-MEL-28, M21	overexpression	Cell line specific	X	Malignant melanomas are due to special break points at 1p36 and at several sites throughout Ap22-q21. GAS5 overexpression leads to reduced melanoma migration and invasiveness (reduced MMP2 levels).	Chen et al, 2016
HOTAIR	human CM cells A375	siRNA	Upregulated in metastasis	syntenic with mammals	HOTAIR favors motility, invasion and metastatic potential of melanoma through over-activation of MMP2/MMP9 basement membrane degradation	Tang et al, 2013
LLME23	human CM cells YUSAC mouse xenograft	RNA interference	Upregulated	X	LLME23 promotes the expression of the proto-oncogenic RAS-related small GTPase Rab23 through regulation of PSF binding	Wu et al, 2013
MALAT1	human UM cells MUM2C human CM cells A375	siRNA, overexpression	Upregulated	Sequence conservation with zebrafish	Possibly involved in cell proliferation and invasion. MALAT1 acts by targetting miR-202 in cutaneous melanoma and miR-140 in uveal melanoma (stabilising SLUG and ADAM10 expression)	Tian et al, 2014; Luan et al, 2016; Sun et al, 2016
PAUPAR	human UM cells MUM2B, CCM1, OM431, 293T mouse xenograft	overexpression, siRNA	Downregulated	Sequence conservation with mouse	PAUPAR is a tumor suppressor which reduces cell migration and metastasis. Its overexpression in cell culture and mouse xenograft leads to reduced invasive capacities through modulation of HES1 expression by inhibiting histone H3KA methylation	Ding et al, 2016
PVT1	hCM cells A375	shRNA, Overexpression	Upregulated	Syntenic with zebrafish	PVT1 level is associated with tumor presence in melanoma tissue and in patient serum. It increases cancer invasion potential through MYC protein stabilisation. PVT1 is a potential biomarker	Chen et al, 2016
RMEL3	human CM cells A375, MEL624 WM1356	siRNA	Upregulated	Syntenic with mouse	RMEL3 depletion leads to decreased cell survival and proliferation in BRAFV600E melanoma cell lines. Its strong correlation with PI3K genes alters cell cycles and apoptosis regulation levels	Goeder et al, 2016
SAMMSON	human CM cells 501Mel, SK-MEL-28 mouse xenograft	LNA, GapmeRs, siRNA, overexpression	Upregulated	Syntenic with rabbit and sheep	Promotes cell viability and growth irrespective of melanomas mutational status. SAMMSON interacts with p32 to increases mitochondrial targetting and pro-oncogenic function	Leucci et al, 2016
SLNCR1	human CM cells A375, WM1976, WM1575	siRNA, Genetic deletion	Upregulated	Syntenic and sequence conserved with mouse	SLNCR1 is associated with poor melanoma survival. It increases melanoma invasion by transcriptionnally upregulating MMP9	Schmidt et al, 2016
SNGH5	Patient serum	X	Upregulated	X	Serum level reflects tumor bearing status of the patient but not the progression of the cancer	Ichigozaku et al, 2016
SPRIGHTLY	human CM cells WM1552C human melanocyte HEM-1 Patient plasma, mouse xenograft	overexpression, siRNA, Genetic deletion	Upregulated	Sequence conservation with mouse	Regulation of cells viability apoptosis and melanoma cell motility. SPRIGHTLY is associated with melanoma genesis, with patient higher tumor stage, worse prognosis and low survival	Khaitan et al, 2011; Mazar et al, 2014; Zhao et al, 2016; Liu et al, 2016
THOR	human CM cells MM603, SK-MEL-5 human LC cells NCI-H1437, NCI-H1299 mouse xenograft, zebrafish	siRNA, overexpression, Genetic deletion	Upregulated	Sequence conservation with zebrafish	Ultra-conserved lncRNA associated with cancer proliferation phenotype due to THOR-IGF2B1 interaction	Hosono et al, 2017
UCA1	human CM cells A375	siRNA	Upregulated	X	Promotes invasion and cell proliferation	Tian et al, 2014; Wei et al, 2016

d. Zebrafish model to investigate melanoma progression

Zebrafish, is an established model to investigate vertebrate embryogenesis due to its external fertilization, its well-defined developmental stages and its transparent embryos (Kimmel et al., 2004) Despite its whole genome duplication and 450 million years of evolutionary distance with human (Hoegg et al., 2004), its anatomical, biochemical, physiological and genetic properties are conserved to mammals, making zebrafish a good model to investigate vascular, inflammatory and neurodegenerative human diseases (Schmid and Haass, 2013) but also defining zebrafish as a useful oncological model (Etchin et al., 2011). The first example of cancer induction in zebrafish was reported by Langenau et al. in 2003, which reproduces human T cell acute leukemia histopathology through overexpression of the human c-myc oncogene (Langenau et al., 2003). Unlike xiphophorus fish (MeierJohann and Schartl, 2006; Wellbrock et al., 2002), zebrafish do not develop spontaneously melanoma, even though the inactivation of the chromatin modifier kdm2aa seems to be sufficient to drive spontaneous tumorigenesis of different cancer subtypes, including melanoma in a low number of zebrafish (8,4% on 7 to 28 months old individuals) (Scahill et al., 2017).

i. Similarities between zebrafish and human skin

In human and in zebrafish, melanocytes are melanin producing cells originating from the neural crest lineage (John, 1997; Shain and Bastian, 2016). In human, melanocytes (around 1500/mm^2 of human epidermis) divide once or twice a year and are localised in other tissues than epidermis, such as skin hair, eye uveal tract, meninge and anogenital tract (Shain and Bastian, 2016).

Zebrafish skin, like human skin, is divided into three layers: the stratified epidermis, the dermis and the hypodermis (Figure 5). Unlike in human skin, zebrafish pigmentation is located in the hypodermis and composed of three different types of pigmented cells, or chromatophores, originating from the neural crest lineage (John, 1997). The melanophores are the zebrafish melanin producing cells orthologous to the human melanocytes. The iridophores are responsible for the silver-iridescent aspect of the zebrafish skin due to cytoplasmic reflective platelets in different orientation (S and L types). The xantophore cells form pterinosomes vesicles containing pteridin and carotenoid pigment of yellow-orange colors. Several pigmentation mutants have been reported in zebrafish, involving single or several chromatophores (John, 1997). Pigmentation deficient zebrafish lines, such as Casper, were

engineered to create transparent adult zebrafish embryos to facilitate the *in vivo* imaging of cellular processes at a whole organism scale (White et al., 2008).

Chromatophore cellular organisation of the zebrafish skin's striped pattern has been investigated with electron microscopy (Hirata et al., 2005; Hirata et al., 2003). Zebrafish skin presents 4 longitudinal stripes composed of a single layer of 6 melanophores, associated with thin iridophores stripes and scattered xantophores. Its interstripes are mainly composed of iridophores and xantophores cells, and zebrafish non-striped skin (ventral) of iridophores with scattered melanophores. Presence of pigmented cells (melanophores and xantophores) has also been observed on the zebrafish mineralised collagenous cycloid scales, indeed scales analysis and transplant are common techniques to investigate melanin-producing cells behaviour in zebrafish melanoma research (Kaufman et al., 2016; Michailidou et al., 2009).
Despite differences in the organisation of pigmentation, human and zebrafish melanocytes/melanophores present multiple histological and physiological similarities due to their common neural crest cell origin and transcriptional programs (Barriuso et al., 2015; Etchin et al., 2011; Yen et al., 2014)

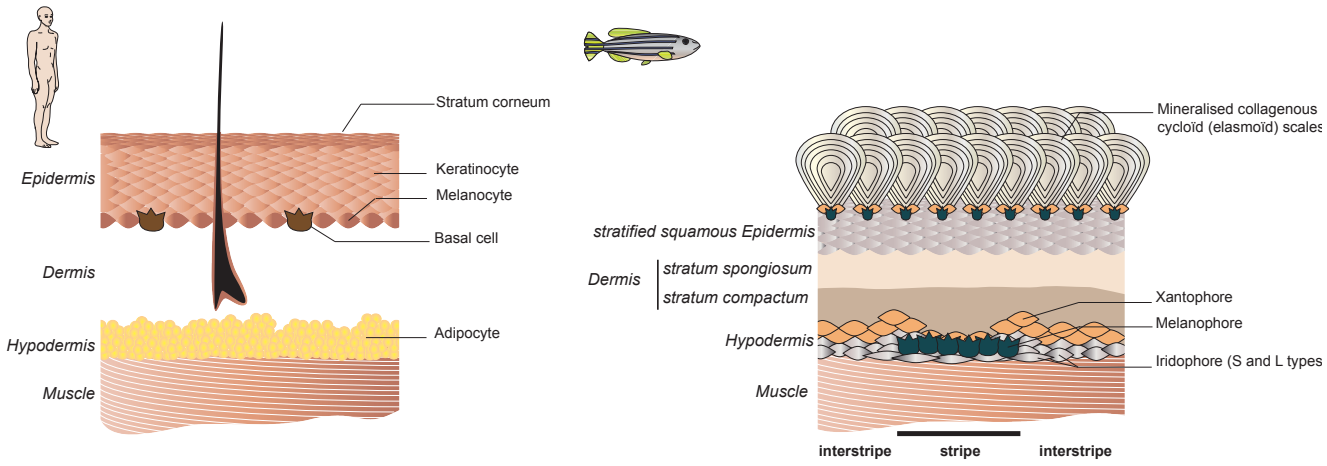


Figure 5: Schematics of the human and zebrafish skin

Human and zebrafish skin are structurally similar and are composed of three layers: epidermis, dermis and hypodermis. Human skin is composed of only one type of pigmented cells, the melanocytes, whereas zebrafish skin pigmentation is composed of xantophores, iridophores and melanophores.

ii. Strategies to investigate melanoma in zebrafish

The conserved melanocytes/melanophores properties make zebrafish a great model to investigate melanoma at the organismal level. Four different methodologies have been

reported in the zebrafish oncology field: (1) Forward mutagenesis due to N-ethyl-N-nitrosourea mutagen agent (ENU) or retroviral and transposon insertion (Amatruda and Patton, 2008; Amsterdam et al., 1999), which induce formation of leukemia, pancreatic cancer, melanoma and rhabdomyosarcomas in zebrafish; (2) Reverse mutagenesis through the creation of transgenic lines expressing human oncogenes (see below); (3) xenografts of human/mice cancer primary cells or cell lines (see below) and (4) therapeutic/drug screens (Etchin et al., 2011).

- Transgenic melanoma models in zebrafish

Several transgenic constructs have been reported to induce melanoma in zebrafish (see Table 2). These constructs are composed of a zebrafish melanophore specific promoter, such as microphtalmia associated transcription factor A (mitfa) or proto-oncogene receptor tyrosine kinase A (kita), that will drive the expression of human oncogenes NRAS or BRAF (see Table 2). In 2005, Patton et al reported the first example of the transgenesis inducing melanoma in zebrafish (Patton et al., 2005). Insertion of the mitfa::$BRAF^{V600E}$ in the zebrafish genome lead to alteration of the pigmentation pattern in embryos and development of benign naevi in 10% of adults. When inserted in a p53 defective background (missense mutation (Berghmans et al., 2004)), mitfa::$BRAF^{V600E}$ leads to the formation of neoplastic naevi in 13,6% and malignant melanoma in 50% of 4 months old fish. Zebrafish melanoma tumors were characterised as MAPK pathway over-activated tumors and were histologically similar to human tumors (Patton et al., 2005).

Among other reported constructs, mitfa::$NRAS^{Q61K}$ requires p53 inactivation to induce melanoma development (Dovey et al., 2009). $HRAS^{G12}$ contruct is reported to be sufficient to induce melanoma in zebrafish and activate MAPK pathway, even though $HRAS^{G12}$ does not drive melanomagenesis in human patients. Interestingly, the zebrafish kita promoter appears to be more efficient than the mifta promoter to induce $HRAS^{G12}$ melanoma (Michailidou et al., 2009; Santoriello et al., 2010). Michailidou et al have described the evolution of melanoma in injected zebrafish, from neoplastic lesion disturbing the striped pattern to Radial Growth Progression (RGP) where melanoma cells expand in the connective tissue and epidermis and Vertical Growth Progression (VGP) with formation of nodules, tumoural hypoplasia and tumoural dysplasia (Michailidou et al., 2009).

Once melanoma driving potential of the different constructs had been characterised, melanoma neural crest origin, its initiation process (Kaufman et al., 2016; White et al., 2011) and the impact different genes were described in melanoma $BRAF^{V600E}$ p53 deficient individuals. SETDB1 has then been described to accelerate melanoma formation, and KIT

tyrosine kinase receptor appears to modulate and slow-down $BRAF^{V600E}$ melanoma (Ceol et al., 2011; Neiswender et al., 2017). Recently, a zebrafish melanoma transgenic model has also been used to characterize the conserved oncogenic function of the lncRNA *THOR* (Hosono et al., 2017).

- Xenografts of melanoma cells in the zebrafish

Metastasis is a significant cause of death in melanoma patients (Izraely et al., 2011). The process of metastasis shares multiple cell motility mechanisms (cell migration and adhesion behaviour, cortex rigidity, protrusion and actomyosin contractions) with the embryogenesis of multicellular organisms. Therefore, several developmental models can support investigation of precise mechanism reported in melanoma (Stuelten et al., 2018). The metastatic process consists in (1) Loss of tumor cell epithelial polarity, (2) Breakdown of the tissue architecture by highjacking tumor environment, (3) Breach of the basement membrane by Matrix Metalloproteinases (MMPs, often associated with over-activation of PI3K pathway), (4) Intravasation of tumor cells in the blood or lymphatic vessels, (5) Extravasation, (6) Migration in the new tissue, (7) Expansion of the metastatic colony according to environment permissiveness which can be associated with facultative (8) Angiogenesis (Stuelten et al., 2018).

Numerous i*n vitro* tests have been developed to analyse metastasis (wound healing, transwell motility, invasion assay or hanging drop assay). However, they cannot assess complex processes such as vessel intravasation/extravasation or take into account the intrinsic biological properties of a living organism (Teng et al., 2013). Mouse metastatic models have been used as an *in vivo* model to investigate metastasis through xenografts, but they present several technical caveats such as the low size of cohorts, as immune-compromised mice are expensive and require special facilities. Moreover, metastasis analysis in mice is a long process which last several months and mouse models present inherent biological hindrances to metastasis, such as their ability to fight metastatic process and their tissue deepness (making it difficult to investigate early metastatic stages) (Teng et al., 2013).

With its transparency in adults (Akhter et al., 2016) and in embryos, zebrafish is an optimal model to observe metastatic process using *in vivo* microscopy through transplantations or xenografts of mammalian tumor cells. Moreover, while adult fish requires preliminary irradiations of 20-25Gy to eliminate immune response to the xenograft (Taylor and Zon, 2009), zebrafish embryos do not develop immune system until 11 dpf (day post-fertilisation) and dispose of several alternative grafting sites (yolk, abdominal perivitelline

space, pericardium) (Mimeault and Batra, 2013). Together with the high number of zebrafish cohorts, transgenesis and chemical treatment properties, zebrafish has many advantages as a metastatic melanoma model.

Several subtypes of cell lines and primary tumors (liquid or solid) have already been reported to conserve their transcriptional patterns and invasive properties in zebrafish xenografts (Teng et al., 2013; Fornabaio et al., 2018). Various metastatic processes have been investigated and reported in zebrafish xenograft, such as angiogenesis where zebrafish transplant appears to have similar potential as mouse model (Nicoli et al., 2007), hyperplasia/malignant growth differentiation (Taylor and Zon, 2009), identification of cancer stem cells and intra-tumoral heterogeneity (Fior et al., 2017), investigation of post-transplantation tumor dissemination, migration and homing properties (Mimeault and Batra, 2013; Fornabaio et al., 2018), and cancer cell metastatic cooperation to enhance metastatic potential of poorly invasive cells (Chapman et al., 2014).

iii. Human melanoma-associated lncRNAs with conserved evolutionary features

Four melanoma-associated long noncoding RNAs are conserved between human and zebrafish (Table 1). Their mechanisms and functions in melanoma and other cancers are summarized below.

- *Testis associated Highly conserved Oncogenic long noncoding RNA (THOR)*

The lncRNA *THOR* is a sequence-conserved transcript preferentially expressed in vertebrate testis. *THOR* is overexpressed in multiple human cancers including lung carcinomas and has been investigated in human melanoma cancer cell lines through knock-down or knock-out, but also *in vivo* through xenograft in mouse metastatic model or melanoma $NRAS^{Q61}$ transgenic zebrafish (Hosono et al., 2017). The *THOR* lncRNA interacts with IGF2BP1 to regulate transcriptional levels of several genes, and presents an oncogenic function conserved from human to zebrafish. *THOR* is the first example of a lncRNA with conserved oncogenic function (Hosono et al., 2017).

- *Metastasis Associated Lung Adenocarcinoma Transcript 1 (MALAT1)*

MALAT1 is one of the most abundant and broadly expressed syntenic and partially sequence conserved lncRNA (Ulitsky et al., 2011). Several mouse null allele of *Malat1* were generated (Eißmann et al., 2012; Nakagawa et al., 2012; Zhang et al., 2012), however the only phenotypic defect reported so far in *Malat1*$^{-/-}$ mouse is abnormal retinal vascularisation (Michalik et al., 2014). In human melanoma patients, *MALAT1* level increases with cancer progression and its knock-down impairs melanoma invasive properties, suggesting an oncogenic function of *MALAT1* (Tian et al., 2014) that is conserved in other cancer subtypes (Arun et al., 2016).

- *Plasmacytoma Variant Transcript 1 (PVT1)*

PVT1 is a known lncRNA oncogene owing to its many association and proximity to *MYC* proto-oncogene genomic locus (Tseng et al., 2015). Several studies reported *cis*-regulatory function of the *PVT1* transcript on adjacent *MYC* transcription (Werner et al., 2017) or on stabilization of *MYC* protein levels (Tseng and Bagchi, 2015). Recently, it has also been reported that sequences in the promoter of *PVT1* interact with 58kb distant *MYC* promoter, promoting *MYC* expression and cell growth (Cho et al., 2018; Liu et al., 2016). Human *PVT1* locus generates almost 25 different RNA isoforms with distinct TSSs, splicing patterns and ATAs, covering a large DNA locus and overlapping with several DNA regulatory motifs (Werner et al., 2017). *PVT1* transcripts show conserved features, such as sequence conservation of the first 3 exons in mammals and synteny in vertebrates (Hezroni et al., 2015; Ulitsky et al., 2011). *PVT1* has been reported to act as a human oncogene in several cancer subtypes, and recently as a melanoma progression biomarker (Chen et al., 2017; Colombo et al., 2015; Liu et al., 2016; Zhou et al., 2016; Tseng et al., 2014; Zheng et al., 2016).

- *CAncer SusCeptibility 15 (CASC15)*

Human *CASC15* is a mammalian sequence conserved (64% between human and mouse) and vertebrate syntenic lncRNA locus. *CASC15* spans a large DNA locus adjacent to the developmentally important transcription factor *SOX4*, where multiple lncRNA isoforms are transcribed (with different TSSs, splicing patterns or ATAs). *CASC15* has been involved in cancer progression under different names (*CANT1*, *lnc00340*, lnc-*sox4* or *TLINC*), however its putative role is distinct according to cancer subtypes.

CASC15 expression has been reported to decrease with neuroblastoma severity (Russell et al., 2015), its knock-down and overexpression in neuroblastoma cell lines supporting its role as a tumor suppressor through its interaction with ubiquitinase USP36 (Mondal et al., 2018). Concomitantly, *CASC15* levels have been shown to increase in cutaneous melanoma patients where it has been proposed to play a role in the cancer proliferation to invasion switch (Lessard et al., 2015), and has then been reported as a tumor suppressor in uveal melanoma (Xing et al., 2017).

The tumor suppressor function of *CASC15* has yet been contested by several studies reporting oncogenic function of the lncRNA when acting in cis to control adjacent *SOX4* expression (Chen et al., 2016; Fernando et al., 2017; Merdrignac et al., 2018; Wu et al., 2018). Therefore *CASC15* might present distinct functions related to alternative isoforms or cancer subtypes, thus its functionality remains elusive.

III. Project objectives

Taking advantage of the conserved properties of the zebrafish model, the purpose of my thesis is to investigate the *in vivo* function of syntenic lncRNAs and to determine if their functions are conserved among vertebrates. Toward this end, I have generated zebrafish mutants for a set of syntenic lncRNAs identified in Ulitsky et al., 2011 using CRISPR-Cas9 genome editing technology (Hwang et al., 2013) (detailed in Chapter 2). I have observed, like reported in numerous studies on mammalian lncRNAs, that zebrafish lncRNAs are expressed at low levels and primarily appear to be dispensable for the viability and fertility in standard laboratory conditions. I have therefore investigated lncRNA functionality using specialized phenotypic tests in zebrafish and by combining lncRNA mutations to attenuate putative compensatory mechanisms. Generating a double knock-out for both isoforms of *PVT1* orthologs, I have observed partially penetrant embryonic defects that will require deeper characterization (detailed in Chapter 1, part II).

As zebrafish is an oncology model and lncRNAs appear to be frequently misregulated in cancer, I have also assessed the role of six zebrafish lncRNAs in $NRAS^{G12}$ driven melanoma development (Chapter 1, part I). My results have brought into focus the tumor suppressor function of lnc-*sox4a* or *menhir (MElaNoma HIndrance long noncoding RNA)* in zebrafish $NRAS^{G12}$ melanoma. Indeed, zebrafish *menhir* is conserved to human at several levels: the *menhir* long isoform presents a 20bp stretch of sequence conservation with human *CASC15* (Ulitsky lab; data not shown), is syntenic and shows a conserved expression pattern in healthy

and tumoral tissues. The observed sequence conservation being specific to the first exon, it is highly probable that the conserved 20bp motifs illustrate DNA regulatory element conservation rather than RNA. My results also show that the *menhir* tumor suppressor function is conserved throughout evolution in the absence of sequence conservation, as its syntenic human ortholog *CASC15* seems to be sufficient to rescue *menhir* mutant phenotype.

RESULTS

I. Assessing the impact of lncRNAs on melanoma development: the lncRNA *menhir/CASC15* acts as a tumor suppressor in cutaneous melanoma

Because several of the selected lncRNAs have been reported to be associated with melanoma in human patients, we investigated their oncogenic or tumor-suppressor potential in melanomagenesis. The function of lncRNAs in cancer is often investigated in cell culture and rarely in animal models, or in very specific physiological mechanisms (development, behaviour, etc.) (Bassett et al., 2014). In order to assess the impact of lncRNAs on melanoma aggressiveness at the level of the entire organism, we investigated the function of our candidate genes in melanomagenesis in the zebrafish oncology model.

a. A small-scale reverse genetic screen for melanoma progression

While human and zebrafish share multiple similarities in melanocytes physiology and genetics, zebrafish do not develop melanoma spontaneously (Patton et al., 2005; Scahill et al., 2017). Melanoma development can be induced by expressing a melanoma oncogene in zebrafish melanophores. Several constructs have been reported to drive melanoma development in zebrafish, each having advantages and caveats (Table 2). We used a construct expressing the G12D oncogene form of human NRAS under the control of the zebrafish melanophores specific promoter mitfa. The G12D mutation in NRAS is associated with 3-4% of human melanomas (Table 2), and has been reported to efficiently drive melanoma in a p53 wild type genetic background in zebrafish (unpublished data from Adam Hurlstone lab, UK). The plasmid containing the mitfa::$NRAS^{G12D}$ transgene was kindly provided to us by Adam Hurlstone (Manchester, UK). NRAS melanomas present fewer therapeutic solutions than BRAF mutant melanoma (Johnson and Puzanov, 2015), as BRAF melanoma are more sensitive to MAPK inhibitor (Sullivan and Flaherty, 2012). Thus, investigating the impact of lncRNAs on $NRAS^{G12D}$ induced melanoma progression has the potential to identify new therapeutic targets.

Table 2: Different constructs used to induce melanoma in zebrafish

Transgene	p53 status	Prevalence in human	References
mitfa::BRAF^{V600E}	mutant	34-44%	Patton et al, 2005
mitfa::GNAQ^{Q209P}	mutant	20% in uveal and mucosal melanoma	Mouti et al, 2016
mitfa::HRAS^{G12V}	wild-type	driving melanoma in mice	Michailidou et al, 2009
kita::HRAS^{G12V}	wild-type	driving melanoma in mice	Santoriello et al, 2010
mitfa::NRAS^{Q61K}	mutant	12-23%	Dovey et al, 2009
mitfa::NRAS^{G12D}	wild-type	1-2%	unpublished, Hurlstone lab

So far, only one lncRNA called *THOR* has been reported to impact melanoma aggressivity in zebrafish (Hosono et al., 2017), with the oncogenic effect on melanoma. To investigate the impact of lncRNAs on melanoma progression, I have injected the human NRAS^{G12D} expression construct (Figure 6A) into the one-cell stage wild type and lncRNA mutant embryos. Embryos with Venus expression in the lens (encoded by the cyaa:venus cassette in the expression construct; Figure 6A) were selected and the development of melanoma was examined from week 5 to 21 (Figure 6B). As this procedure generates mosaic animals by random insertion triggered by Tol2 transposase, the assay was conducted on a minimum of 50 fish per condition.

We selected six lncRNAs to assess their oncogenic or tumor-suppressor potential, and I monitored melanoma tumorigenesis in lncRNA deletion mutant fish (see Chapter 2, Table 5) and their wild type siblings (Figure 6C). Neither lnc-*myca*$^{-/-}$ nor lnc-*mycb*$^{-/-}$(see Figure 17 for description of the mutants), showed significant differences in melanoma tumorigenesis triggered by NRAS^{G12D} compared to wild type. Likewise, no differences in tumorigenesis were observed between wild type and *malat1*$^{-/-}$ fish (generated through insertion of an early SV40 polyA signal, see Figure 22 D&F). The same results were observed for two non-melanoma related lncRNAs: lnc-*ppm1bb*$^{-/-}$ (see mutant characterisation in Figure 20) and *megamind*$^{-/-}$ (a CNS-specific lncRNA, lncRNA full deletion was generated by the Wolfe lab, (Kok et al., 2015)). By contrast, the lnc-*sox4a* mutant showed a significant difference in tumorigenesis when compared to wild type animals (Figure 6C). Indeed, there is a strong increase of 22% of fish presenting tumors in the lnc-*sox4a*$^{-/-}$ background in early adulthood (5 weeks) compared to wild type, suggesting that this lncRNA regulates melanoma initiation. Melanoma progression also appears to be more aggressive in lnc-*sox4a*$^{-/-}$ background, as the

difference in tumorigenesis between wild type NRAS^{G12D} and lnc-*sox4a*$^{-/-}$ NRAS^{G12D} increases with time from 22% to 33%. Indeed, only 27% of wild type NRAS^{G12D} fish developed tumors before week 21 in comparison to almost 60% of lnc-*sox4*$^{-/-}$ fish (Mantel Cox statistical test *P*=0,0002)

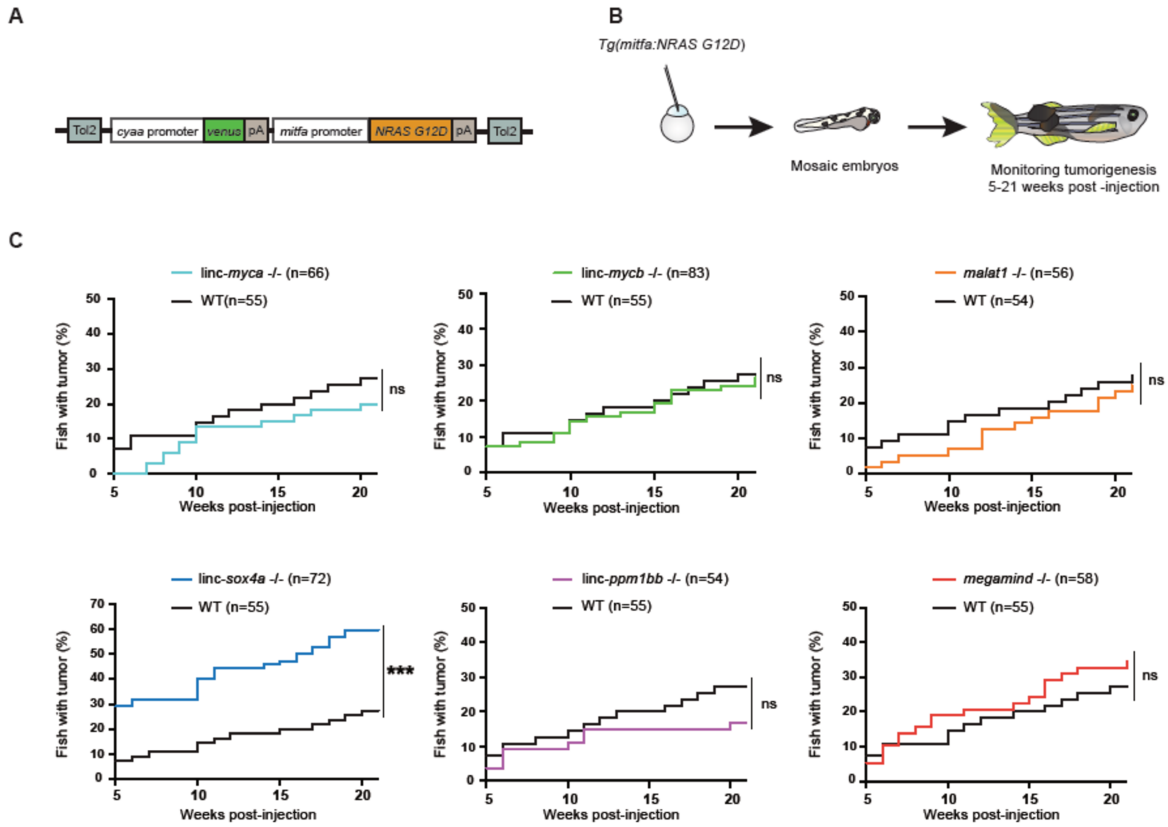


Figure 6: lnc-*sox4a*$^{-/-}$ zebrafish display increased melanoma tumorigenesis

A Schematic diagram of the Tol2 construct used to induce melanoma in zebrafish. It is composed of Venus marker controled by the cyaa promoter (lens specific) and of human oncogene NRAS^{G12D} under the control of the mitfa promoter. This construct (unpublished) was furnished by Adam Hurlstone (Manchester, UK) **B** Schematic of the injection procedure to analyse impact of lncRNA mutation on melanoma tumorigenesis. A minimum of 50 individual with Venus lens were generated for each lncRNA mutant and WT (wild type) cibling line. **C** Percentage of fish with tumor between week 5 and 21 in mutant lncRNA (lnc-*myca*, lnc-*mycb*, *malat1*, lnc-*sox4a*, lnc-*ppm1bb*, *megamind*) and their WT ciblings. Data were analysed with Mantel-Cox test, lnc-*sox4a* and WT presents a *P* = 0,0002

b. Characterisation of lnc-sox4a/menhir mutants and their melanoma-aggressiveness phenotype

i. Characterisation of lnc-*sox4a/menhir* mutants

Based on shared transcriptional orientation and genomic position, the zebrafish lnc-*sox4a* is predicted to be the syntenic ortholog of the human *CAncer SusCeptibility 15* (*CASC15*) lncRNA (Figure 7A) (Glusman et al., 2006). The lnc-*sox4a* gene is located 3' of the *SOX4* (*SRY box 4*) protein-coding gene, a transcription factor with crucial functions in development (Bhattaram et al., 2010; Cizelsky et al., 2013; Huang et al., 2013; Nissen-Meyer et al., 2007; Zhao et al., 2017) and oncogenesis (Huang et al., 2012; Tiwari et al., 2013; Zhang et al., 2013). Recently, *CASC15* (also named *CANT1, linc00340, lnc-sox4, TLINC*) has been reported to play a tumor suppressor role in various human cancers (Chen et al., 2016; Fernando et al., 2017; Merdrignac et al., 2018; Mondal et al., 2018; Russell et al., 2015; Xing et al., 2017) and to be associated with the melanoma proliferation/invasion switch (Lessard et al., 2015). In addition to synteny, lnc-*sox4a* and *CASC15* share conserved expression in ovaries (Figure 7 B, C&D). Interestingly the adjacent protein-coding genes do not show a similar expression pattern (Figure 7 E&F).

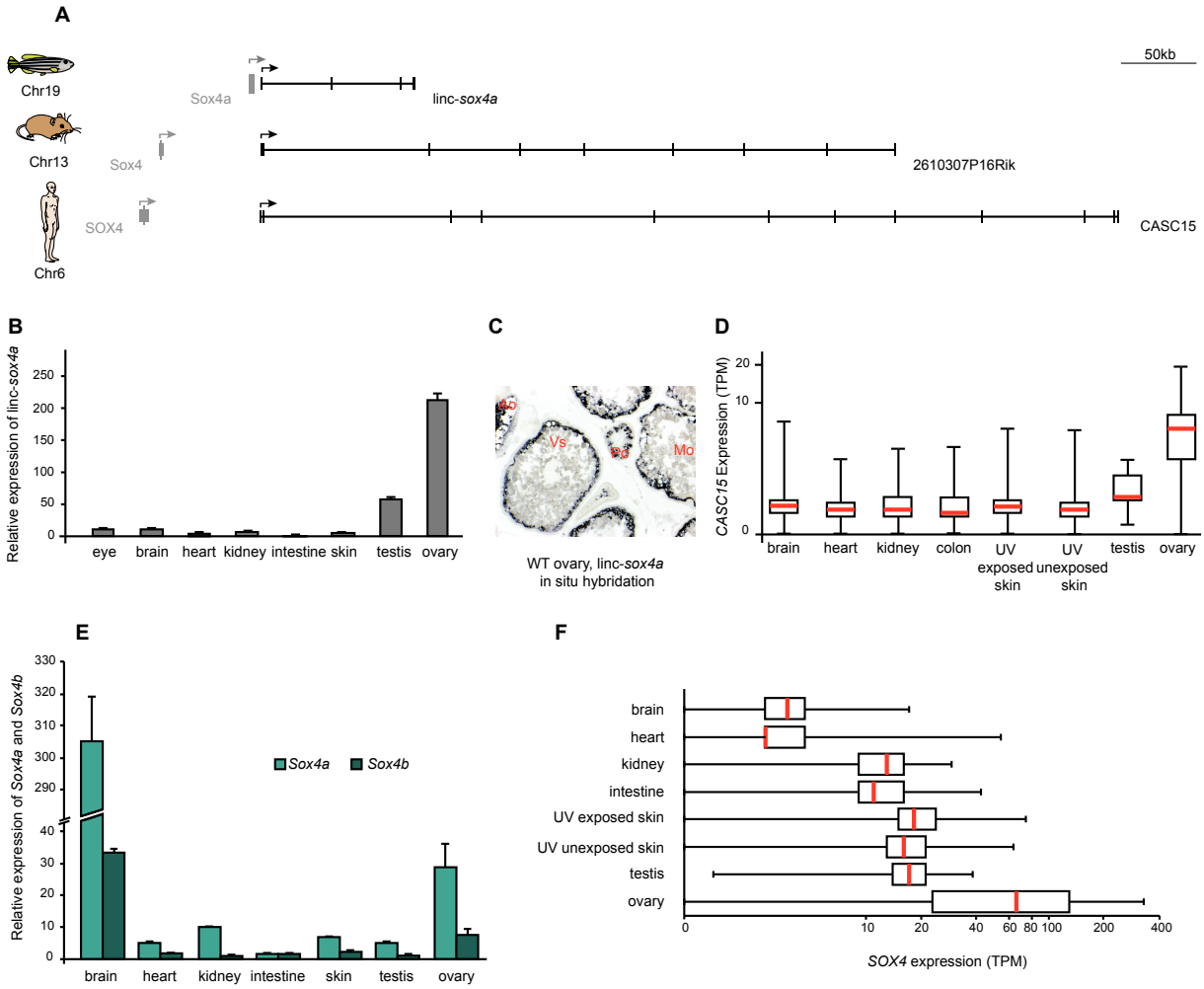


Figure 7: lnc-*sox4a/menhir* and *CASC15* share conserved genomic position and expression pattern

A Genomic locus of the lnc-*sox4a* and the adjacent *sox4a* protein-coding gene in zebrafish, mouse and human. **B** qRT-PCR analysis of lnc-*sox4a* (exon 3 to 4) in zebrafish adult organs (fold change relative to intestine). **C** lnc-*sox4a* in situ hybridation on paraffin section of zebrafish ovary. lnc-sox4a is expressed at each stage of ovary maturation. Its intracellular location is pushed toward the oocyte cortex with the maturation and accumulation of the yolk granules. Po: Primary oocytes; Vs: Vitellogenic stage; Mo: Mature oocyte; Ao: Atretic oocyte **D** Expression of human *CASC15* in adult organs in transcripts per million (TPM). Data were obtained through the Genotype-Tissue Expression (GTEx) project available on EBI Expression Atlas. **E** qRT-PCR analysis of protein-coding genes *sox4a* and *sox4b* expression in wild type adult organs (fold change relative to *sox4b* kidney). **F** Expression of human *SOX4* protein-coding gene in adult organs in transcripts per million (TPM). Data were obtained through the Genotype-Tissue Expression (GTEx) project available on EBI Expression Atlas.

To inactivate lnc-*sox4a* located on Chromosome 19 in zebrafish, I generated a 250bp deletion of the TSS (Figure 8A, see Appendix). This deletion is sufficient to fully inactivate lnc-*sox4a* expression in both the ovary and skin but not in the brain (Figure 8B). Indeed, this deletion induces a usage of a brain specific TSS (Figure 8C) and leads to the production of a 1,7kb lnc-*sox4a* isoform (Figure 8A), expressed at the same level as the endogenous transcript in wild type animals. Therefore, the usage of the alternative TSS in the brain suggests that this transcript might be required for the zebrafish brain function. Moreover, despite exhaustive attempts, I could not produce homozygous mutants for a lnc-*sox4a* 3' deletion, supporting the functional importance of the 3' end of the transcript. In human, several alternative isoforms of *CASC15* have been reported, including a brain specific "*CASC15-S*" isoform composed of the last exonic block, which has been proposed to have a tumor suppressor function in neuroblastoma (Russell et al., 2015). Interestingly, *CASC15-S* has been reported in Russel and al., to be exclusively expressed in the human brain. Therefore, *CASC15-S* and lnc-*sox4a* alternative isoform share similar splicing and tissue expression pattern.

In human, *CASC15* has been reported to act both *in cis* or *in trans* depending on the cancer subtype. In zebrafish ovaries, the tissue where the canonical isoform of lnc-*sox4a* is preferentially expressed, depletion of lnc-*sox4a* induces up-regulation of both the adjacent *sox4a* transcript and the *sox4b* transcript located at the Chromosome 16 (Figure 8D). A similar upregulation was observed in lnc-*sox4a*$^{-/-}$ skin samples. Through modulation of the expression of *sox4a* and *sox4b* transcription factors, our results suggest that lnc-*sox4a* acts both *in cis* and *in trans* on target genes. In the brain, where lnc-*sox4a* deletion is "rescued" by the brain specific alternative TSS, *sox4a* transcriptional level presents high inter-individual variations (Figure 8D).

While zebrafish ovary is structurally different from the mammalian ovary, the oocytes growth and maturation show multiple similarities (Hoo et al., 2016). I compared ovaries isolated from wild types and lnc-*sox4a*$^{-/-}$ fish, stained them with Eosin-Hematoxylin (Figure 8E&F) and did not observe any structural differences. In addition zebrafish lnc-*sox4a*$^{-/-}$ mutants do not show any viability, fertility or morphological defects under standard laboratory conditions.

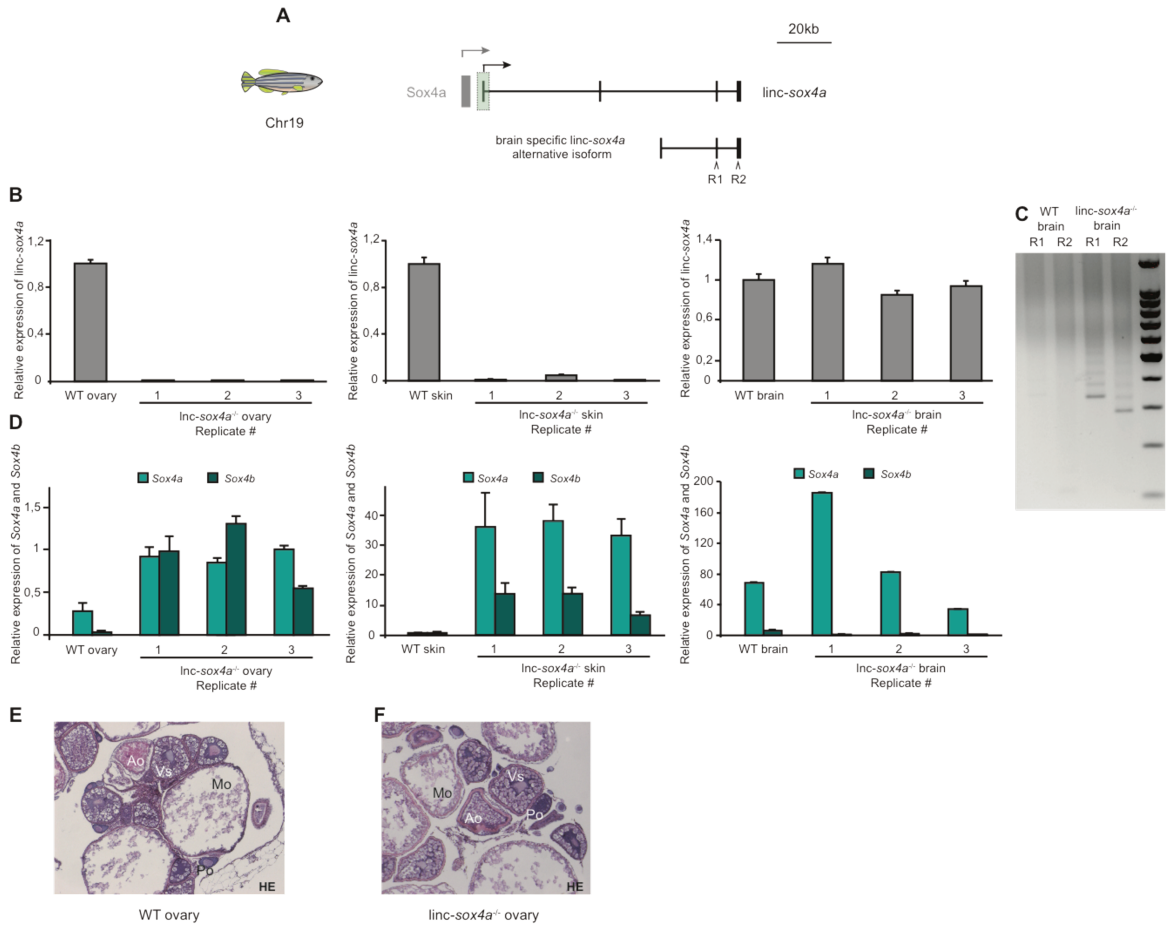


Figure 8: TSS deletion is not sufficient to fully inactivate lnc-*sox4a/menhir* due to brain specific alternative isoform

A Genomic locus of zebrafish lnc-*sox4a*. lnc-*sox4a* is in 3' and in the same orientation as the protein-coding gene *sox4a*. Zebrafish lnc-*sox4a* mutants were generated through deletion of the lncRNA transcription start site (TSS) represented in a green box. The lnc-*sox4a*$^{-/-}$ brain specific alternative isoform is illustrated under the wild type isoform with R1 and R2 corresponding to the 5' Rapid Amplificaiton of CDNA Ends (RACE) primers **B** qRT-PCR analysis of lnc-*sox4a* (exon 3-4) expression in WT (single organ) and lnc-*sox4a*$^{-/-}$(single individual) ovary, skin and brain (fold change relative to WT). **C** Ethidium Bromide of the 5' RACE on WT (single organ) and lnc-*sox4a*$^{-/-}$ brain with two RACE primers R1 and R2. **D** qRT-PCR analysis of protein coding genes *sox4a* and *sox4b* in WT and lnc-*sox4a*$^{-/-}$ (individual) ovary, skin and brain (fold change respectively relative to *sox4a* ovary #3, *sox4a* WT skin and *sox4b* brain #1). **E, F** Eosin and hematoxylin staining of wild type and lnc-*sox4a*$^{-/-}$ ovary. Po: Primary oocytes; Vs: Vitellogenic stage; Mo: Mature oocyte; Ao: Atretic oocyte

ii. Characterisation of lnc-*sox4a/menhir* mutants in $NRAS^{G12}$ melanoma

As demonstrated in Figure 6C, inactivation of lnc-*sox4a* promotes aggressiveness of $NRAS^{G12D}$ melanoma. To further characterize the role of lnc-*sox4a* in melanoma development, I analysed lnc-*sox4a* expression in $NRAS^{G12D}$ zebrafish tumors. Compared to healthy zebrafish skin isolated from non-NRAS mutants, lnc-*sox4a* is highly upregulated in $NRAS^{G12D}$ tumors (Figure 9A). It should be noted that lnc-*sox4a* level is highly varying between individual fish. As highlighted in Figure 7, lnc-*sox4a* and its syntenic human counterpart *CASC15* share similar expression pattern in healthy tissues. Expression datas obtained from the Pan Cancer Analysis of Whole Genome (PCAWG) indicate that, similar to lnc-*sox4a*, *CASC15* is up-regulated in tumor tissues compared to UV-exposed and unexposed skin (Figure 9B), corroborating conserved expression pattern. Likewise, zebrafish protein-coding genes *sox4a*/*sox4b* and human *SOX4* are also up-regulated in melanoma tissues (Figure 9 C&D). However, expression data from PCAWG and previous work on *CASC15* in neuroblastoma or melanoma do not associate lncRNA cancer functionality with the adjacent protein-coding gene *SOX4* (Figure 10) (Lessard et al., 2015; Mondal et al., 2018; Russell et al., 2015), suggesting that *CASC15* acts independently of *SOX4* in melanoma and neuroblastoma.

In addition to increased tumorigenesis observed in lnc-$sox4a^{-/-}$(Figure 6C), lnc-$sox4a^{-/-}$ zebrafish also present significant lower survival at 21 weeks post-injection (Figure 9E) and increased severity of melanoma (Figure 9F) resulting in a significantly increased proportion of fish with Vertical Growth Progression between week 5 and 12 (Figure 9G). Melanoma reached a plateau in stage severity proportion at week 13 and is stable until week 21. All statistical analyses and *P*-values are detailed in Table 3.

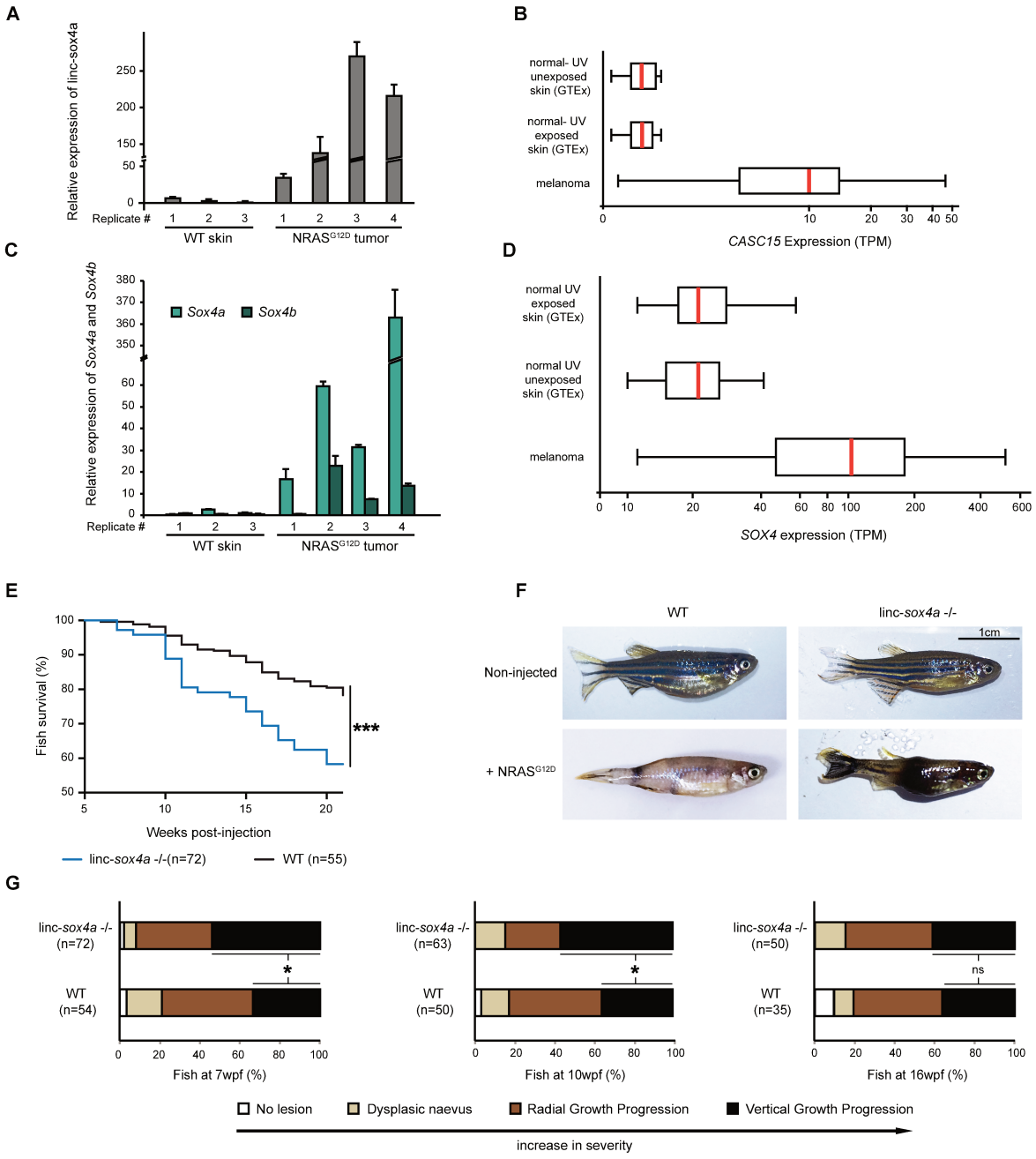


Figure 9: Zebrafish lnc-*sox4a*/*menhir*$^{-/-}$ presents increased melanoma aggressiveness

A qRT-PCR analysis of lnc-*sox4a* in zebrafish individual wild type ovary and skin and wild type NRAS^{G12D} tumor (fold change relative to skin#3) **B** Expression of human *CASC15* in melanoma (n=36), UV exposed (n=394) and unexposed (n=265) skin in transcripts per million (TPM). Data were obtained through the Pan Cancer Analysis of Whole Genome (PCAWG) project available on EBI Expression Atlas. **C** qRT-PCR analysis of *Sox4a/Sox4b* in zebrafish individual wild type ovary and skin and wild type NRAS^{G12D} tumor (fold change relative to *sox4a* skin#1) **D** Expression of human *SOX4* protein-coding gene in melanoma (n=36), UV exposed (n=394) and unexposed (n=265) skin in transcripts per million (TPM). Data were obtained through the Pan Cancer Analysis of Whole Genome

(PCAWG) project available on EBI Expression Atlas. **E** Survival of NRAS^{G12D} WT and lnc-*sox4a*$^{-/-}$ fish between week 5 and 21. Data was analysed with Mantel-Cox test and presented a *P*= 0,0002. **F** Pictures of representative fish at 7 weeks post-fertilization (wpf) non-injected WT and lnc-*sox4a*$^{-/-}$, and injected NRAS^{G12D} WT and lnc-*sox4a*$^{-/-}$. **G** Illustration of NRAS^{G12D} wild type and lnc-*sox4a*$^{-/-}$ fish at different melanoma stages (no lesion, dyplasic naevus, Radial Growth Progression, Vertical Growth progression) at 7, 10 and 16wpf. Decreased n is due to sacrifice of fish reaching the limit level of melanoma severity. Data were analysed with unpaired t-test with a P-value=0,0223 for Vertical Growth Progression (VGP) at 7wpf and *P*=0,0390 at 10wpf.

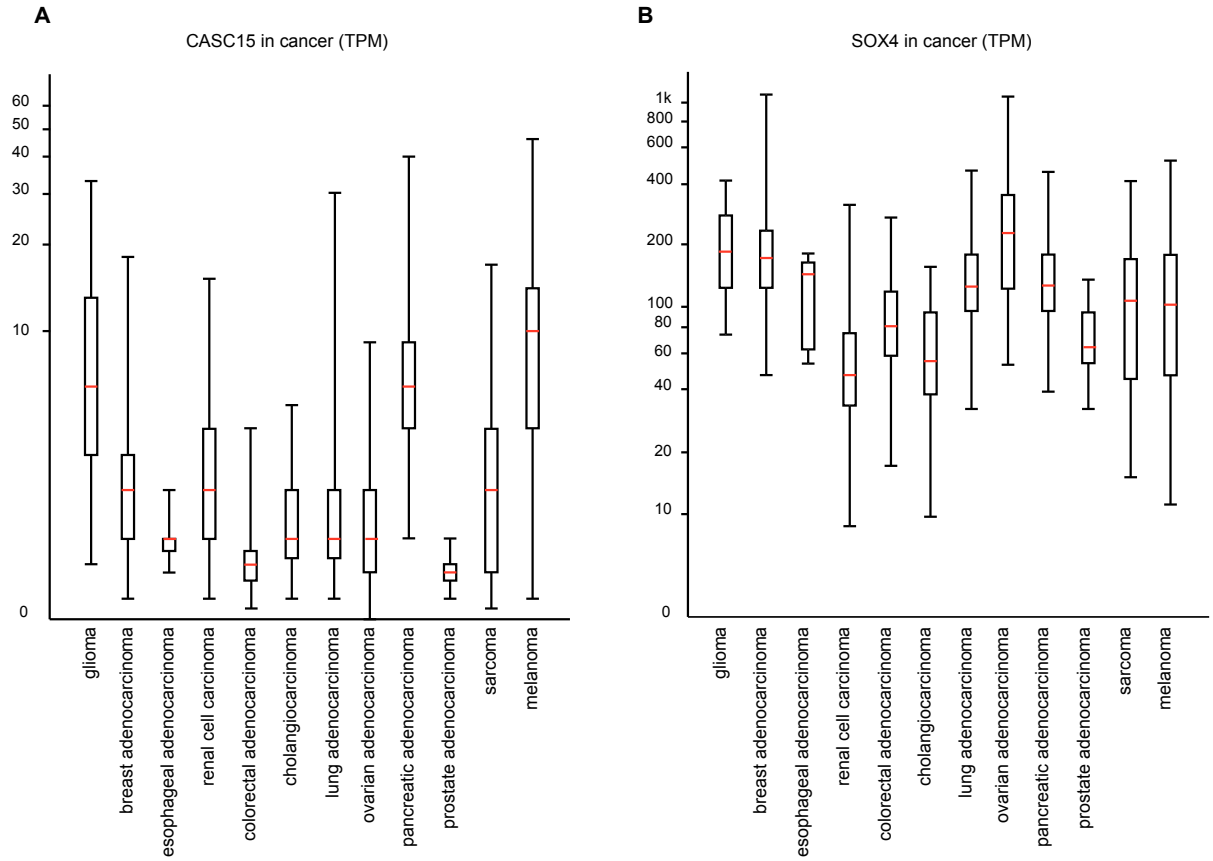


Figure 10: CASC15 and SOX4 expression in various human cancers

A, B Expression of human *CASC15* lncRNA and *SOX4* protein-coding gene in glioma (n=18), breast adenocarcinoma (n=85), esophageal adenocarcinoma (n=7), renal cell carcinoma (n=117), colorectal adenocarcinoma (n=51), cholangiocarcinoma (n=18), lung adenocarcinoma (n=37), ovarian adenocarcinoma (n=110), pancreatic adenocarcinoma (n=75), prostate adenocarcinoma (n=19), sarcoma (n=34) and melanoma (n=36). Data were obtained through the Pan Cancer Analysis of Whole Genome (PCAWG) project available on EBI Expression Atlas.

While examining cancer progression during the 15 week period, a rare event was detected: unlike wild type and other lncRNA mutants, lnc-*sox4a*$^{-/-}$ NRAS^{G12D} fish showed development of highly proliferative internal tumors (Figure 11 A&B). This internal tumorigenesis affected 1,3% (2 individuals out of 169) lnc-*sox4a*$^{-/-}$ NRAS^{G12D} injected fish while it was not observed in wild type nor in other lncRNA mutant NRAS^{G12D} injected fish (0 out of 667; hypergeometric statistical test with a *P*-value of 0,041). The histological characterisation of these tumors is in progress (Figure 11 C&D), yet primary results showed that these tumors have melanoma signatures. Development of these fast-growing internal tumors is correlated with NRAS^{G12D} expression in zebrafish, as this event was never observed in the uninjected lnc-*sox4a*$^{-/-}$ mutant (n = >200).

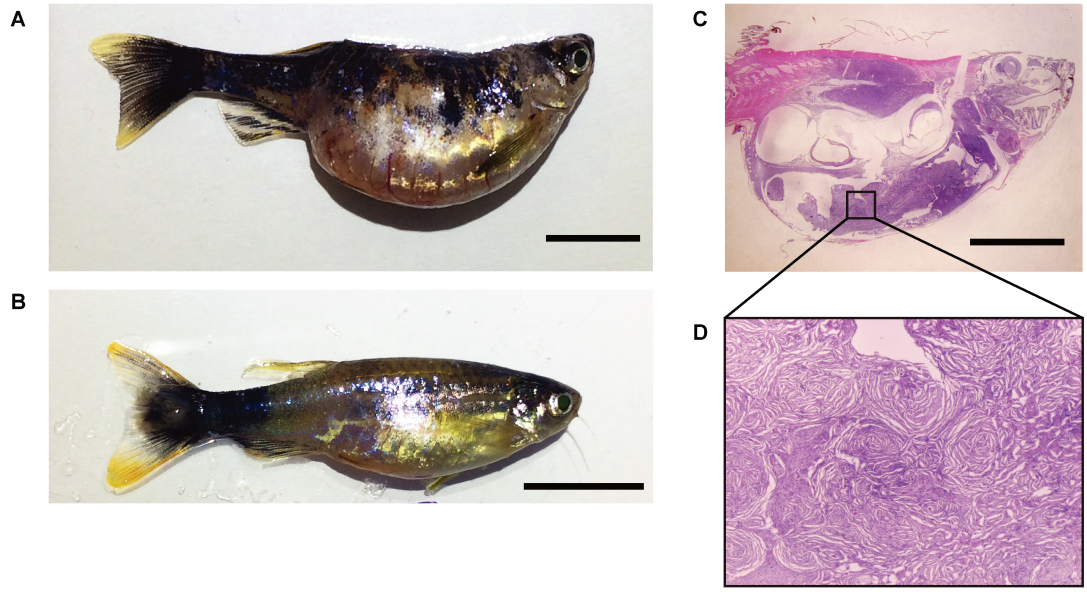


Figure 11: Zebrafish lnc-*sox4a*/*menhir*$^{-/-}$ shows additional internal tumorigenesis

A Picture of a zebrafish lnc-*sox4a*$^{-/-}$ NRAS^{G12D} at 19 weeks post fertilisation. **B** Picture of a zebrafish WT NRAS^{G12D} at 19 weeks post fertilisation at VGP stage (caudal nodule). Black bar represents 1cm **C** Longitudinal section stained with Hematoxylin, Eosin and Safran. **D** Zoom in of internal tumor. Section and staining were performed by the Histim platform in Cochin Institute (Paris, France).

c. Determining the lnc-sox4a/menhir metastatic potential

Our results show that $NRAS^{G12D}$ lnc-*sox4a*$^{-/-}$ zebrafish display increased melanoma aggressiveness characterised by increased tumorigenesis, lower survival, faster cancer development and associated internal tumorigenesis. In human, melanoma is lethal in more than 80% of cases once metastasised (Sandru et al., 2014). To investigate the extrinsic impact of lnc-*sox4a* in metastasis, I performed xenograft experiments using 3 different human melanoma cell lines: (1) WN-266.4 (a $BRAF^{V600E}$ cell line previously shown to be invasive in zebrafish xenograft (Chapman et al., 2014)); (2) SK-MEL-2 (a $NRAS^{Q61R}$ cell line reported to be invasive in nude mouse (Pollack et al., 2007)) and (3) 501-mel (a $NRAS^{G12D}$ a non-invasive cell line in zebrafish (Chapman et al., 2014)). These cells were GFP labelled through viral infection and grafted in the yolk of 2-day-old wild type and lnc-*sox4a*$^{-/-}$ zebrafish (Figure 12A), and the invasive potential of each cell line was monitored up to 4 days post-grafting. To visualise the GFP positive grafted melanoma cells moving in zebrafish blood vessels we used a transgenic kdrl::mcherry zebrafish line in which the blood vessels of all endothelial cells are labelled by RFP. An example of a transgenic 6 dpf (day post-fertilization) zebrafish embryo with GFP cells in the yolk and an invading melanoma cell indicated by a white arrow is illustrated in Figure 12A.

For the invasive cells lines WN-266.4 and SK-MEL-2, no significant differences in the number of cells outside the graft site were detected between wild type and lnc-*sox4a*$^{-/-}$ (Figure 12 B&C). By contrast, a significant increase in the number of 501-Mel cells outside of the graft site was detected in 6 dpf lnc-*sox4a*$^{-/-}$ larvae (Figure 12D), suggesting that lnc-*sox4a*$^{-/-}$ embryos are more permissive than wild type to metastatsis of non-invasive $NRAS^{G12}$ melanoma cells. These results also suggest that lnc-*sox4a* is involved in maintaining the integrity of the basement membrane, consistent with the role of other lncRNAs already reported to be associated with basement membrane integrity (Schmidt et al., 2016; Tang et al., 2013). Taking together, our results are consistent with a tumor suppressor function for lnc-*sox4a*, thus hereafter we refer to zebrafish lnc-*sox4a* as *menhir* (*ME*la*N*oma *HI*ndrance *long noncoding RNA*).

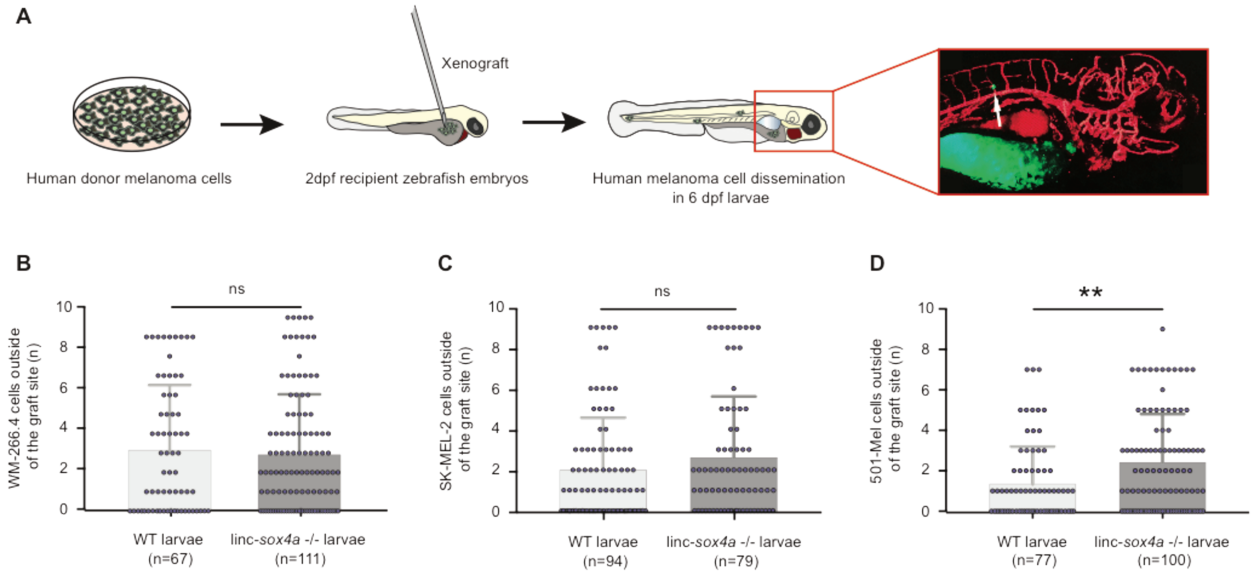


Figure 12: *menhir* mutant presents increased metastatic potential

A Procedure to investigate metastasis in wild type and mutant (left). Example of a 6 day post-fertilization larvae transgenic for kdlr::mcherry grafted with GFP positives human melanoma cells (right), the arrow points to a metastased cell that has penetrated the zebrafish blood circulatory system. **B** Number of human WN-266.4 GFP melanoma cell observed outside of the injection site in wild type and *menhir*$^{-/-}$ mutant at 6dpf. **C** Number of human SK-MEL-2 GFP melanoma cell observed outside of the injection site in wild type and *menhir*$^{-/-}$ mutant at 6dpf. **D** Number of human 501-Mel GFP melanoma cell observed outside of the injection site in wild type and *menhir*$^{-/-}$ mutant at 6dpf. Data are presented as mean with standard deviation, each dot representing individual larvae. Significance of Mann-Whitney test for 501-Mel cells is P=0.0016

d. Rescue of the menhir phenotype with the human lncRNA ortholog CASC15

Having established the melanoma aggressiveness phenotype of the *menhir*$^{-/-}$ NRAS^{G12D} zebrafish, I investigated the conserved functionality between the human and zebrafish orthologs. Indeed, despite limited sequence conservation, *menhir* and *CASC15* share (1) genomic position and transcriptional orientation, (2) expression pattern in healthy and cancer tissue, (3) brain-specific splicing isoform expression and (4) a correlation with the modulation of melanomagenesis. Given these characteristics, we tested whether the *menhir* and *CASC15* lncRNAs have a common function in vertebrates using rescue experiments.

To this end, I first modified the Tol2 NRAS^{G12D} construct by inserting a cassette containing either a control Venus or the *CASC15* gene under the control of the zebrafish mitfa

promoter (Figure 13A). The 21-week melanoma assay in both WT and *menhir*$^{-/-}$ injected zebrafish is currently ongoing. The selected human *CASC15* isoform was amplified from human brain RNAs and does not contain any sequence conserved with the zebrafish *menhir* ortholog. I have also attempted to rescue *menhir*$^{-/-}$ melanoma aggressiveness phenotype injecting a Tol2 construct expressing zebrafish *menhir* transcript, which appeared to be toxic for mutant and wild type embryos (data not shown).

Our results show that *menhir*$^{-/-}$ NRAS^{G12D} fish expressing *CASC15* show a less aggressive progression of melanoma than *menhir*$^{-/-}$NRAS^{G12D} fish (Figure 13B) characterised by a significant decrease of tumorigenesis (Figure 13C) and higher survival (Figure 13D) than either *menhir*$^{-/-}$ NRAS^{G12D} or WT NRAS^{G12D}. This reduced melanoma aggressiveness was accompanied by the ectopic expression of *CASC15* transcript in melanoma tumors (Figure 13E, individual tumors). I also observed a decrease of melanoma severity between *menhir*$^{-/-}$ and its human rescue characterised by a significantly reduced number of fish in the Vertical Growth Progression stage (Figure 13I). All p-values for Figure 13I are detailed in Table 3. The monitoring of *menhir*$^{-/-}$ NRAS^{G12D} + venus also show attenuation of the melanoma aggressiveness (Figure 13F), although the *menhir*$^{-/-}$ NRAS^{G12D} + venus develop more tumors than the *menhir*$^{-/-}$ NRAS^{G12D} fish expressing *CASC15* (P=0,0412) (Figure 13G). Moreover, despite of decreased melanoma aggressiveness when *CASC15* is expressed, we also observe a drastic increase of NRAS^{G12D} oncogene expression in the *CASC15* rescue construct compared to the NRAS^{G12D} + venus (Figure 13H).

My results so far demonstrate that human *CASC15* rescues *menhir*$^{-/-}$ melanoma aggressiveness (characterised by melanoma tumorigenesis, survival and severity), and suggest that syntenic zebrafish *menhir* and human *CASC15* have conserved functions in melanoma development despite absence of sequence conservation.

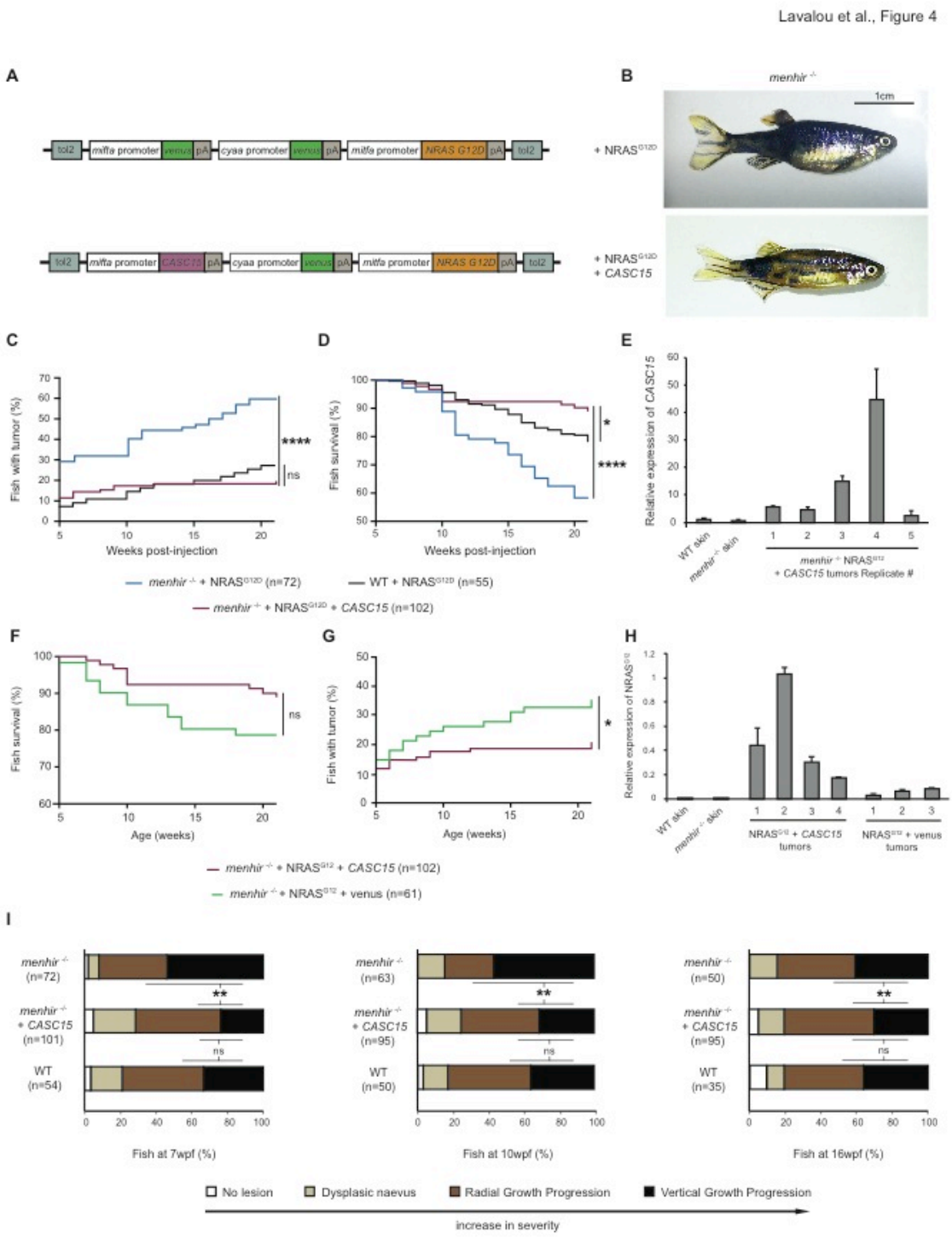


Figure 13: Expression of human *CASC15* in zebrafish melanocytes rescues increased melanomagenesis of *menhir* zebrafish mutants

A Schematic diagram of the modified Tol2 construct used to rescue melanoma aggressiveness in zebrafish *menhir*$^{-/-}$. Human *CASC15* transcript with an inserted artificial intron between exon 3 and 4 is expressed under the control of melanophores specific promoter mitfa. To control that insert size does not impact melanoma aggressiveness, a mitfa::venus cassette has been integrated in the

$NRAS^{G12D}$ control construct. **B** Representative pictures of *menhir*$^{-/-}$ injected with $NRAS^{G12D}$ (at 10 wpf) or $NRAS^{G12D}$ + *CASC15* construct (at 9 wpf). **C** Percentage of fish with tumor between five and twenty-one weeks for WT $NRAS^{G12D}$, *menhir*$^{-/-}$ $NRAS^{G12D}$, *menhir*$^{-/-}$ $NRAS^{G12D}$ + *CASC15*. Data was analysed using Mantel-Cox statistical test and four-star signify a P<0,0001. **D** Percentage of fish survival between five and twenty-one weeks for WT $NRAS^{G12D}$, *menhir*$^{-/-}$ $NRAS^{G12D}$, *menhir*$^{-/-}$ $NRAS^{G12D}$ + *CASC15*. Data was analysed using Mantel-Cox statistical test, with a P= 0,0287 between wild type and *menhir*$^{-/-}$ + *CASC15*, and a P<0,0001 between *menhir*$^{-/-}$ and its rescue. **E** qRT-PCR analysis of human *CASC15* expression in *menhir*$^{-/-}$ $NRAS^{G12D}$ + *CASC15* rescue injected zebrafish (fold change relative to *menhir*$^{-/-}$ skin). **F** Percentage of fish with tumor between five and twenty-one weeks for *menhir*$^{-/-}$ $NRAS^{G12D}$ + *CASC15* and *menhir*$^{-/-}$ $NRAS^{G12D}$ + *venus*. Data was analysed using Mantel-Cox statistical test and with a P=0,0412. **G** Percentage of fish survival between five and twenty-one weeks for *menhir*$^{-/-}$ $NRAS^{G12D}$ + *CASC15* and *menhir*$^{-/-}$ $NRAS^{G12D}$ + *venus*. Data was analysed using Mantel-Cox statistical test. **H** Human $NRAS^{G12}$ expression in non-injected skin, $NRAS^{G12}$ + *CASC15* melanoma tumors and $NRAS^{G12}$ + *venus* melanoma tumors (biological replicates) detected by qRT-PCR. *eef1α1* was used as a reference gene. **I** Illustration of $NRAS^{G12D}$ wild type, *menhir*$^{-/-}$ and *menhir*$^{-/-}$ + *CASC15* fish at different melanoma stages (no lesion, dyplasic naevus, Radial Growth Progression, Vertical Growth progression) at 7, 10 and 16wpf. Data were analysed with unpaired t-test. Decreased n is due to sacrifice of fish reaching the limit level of melanoma severity. Data were analysed with unpaired t-test. All *P*-value are reported in Table 4.

Table 3: Statistical analyses of $NRAS^{G12D}$ zebrafish melanoma stage by age and mutation status

7wpf	*menhir* $^{-/-}$ vs WT	*menhir* $^{-/-}$ vs *menhir* $^{-/-}$ h*CASC15*	WT vs *menhir* $^{-/-}$ h*CASC15*
No lesion	ns	ns	ns
Melanocytic lesion	ns	0,02	ns
Radial Growth Progression	ns	ns	ns
Vertical Growth Progression	0,0223	0,0034	ns

10wpf	*menhir* $^{-/-}$ vs WT	*menhir* $^{-/-}$ vs *menhir* $^{-/-}$ h*CASC15*	WT vs *menhir* $^{-/-}$ h*CASC15*
No lesion	ns	0,0172	ns
Melanocytic lesion	ns	ns	ns
Radial Growth Progression	ns	ns	ns
Vertical Growth Progression	0,0390	0,0058	ns

16wpf	*menhir* $^{-/-}$ vs WT	*menhir* $^{-/-}$ vs *menhir* $^{-/-}$ h*CASC15*	WT vs *menhir* $^{-/-}$ h*CASC15*
No lesion	ns	0,015	ns
Melanocytic lesion	ns	ns	ns
Radial Growth Progression	ns	ns	ns
Vertical Growth Progression	ns	0,00337	ns

e. CASC15 expression profile in human cancer

To test the correlation between expression of the selected lncRNA and melanoma, Nicolas Servant (Institut Curie, Paris) collected human patient data from The Cancer Genome Atlas (TCGA) database and analysed *MALAT1* (Figure 14A), *PVT1* (Figure 14B) and *CASC15* (Figure 14C) expression in different melanoma driving mutations ($BRAF^{G469}$, $BRAF^{V600}$, $KRAS^{G12}$, $NRAS^{G12}$, $NRAS^{Q61}$). *MALAT1* expression does not show a significant correlation with any specific melanoma mutation, even though *MALAT1* appears to have slightly reduced expression in $NRAS^{G12}$ melanoma (Figure 14A), whereas the *PVT1*

expression level is stable across different mutations (Figure 14B). Interestingly, *CASC15* is the transcript with the most variation in its expression across different melanoma mutations (Figure 14C). *CASC15* is significantly lower expressed in $BRAF^{V600}$ and its expression is significantly higher in $KRAS^{G12}$ compared to wild type KRAS. It also appears to be slightly up-regulated in $NRAS^{G12}$ compared to $NRAS^{WT}$, however the upregulation is not significant. When *CASC15* expression data was organised according to the melanoma clinical stage (Figure 14D), it displayed no significant change between stages I to III-IV. It is important to note that only a low number of samples was available for rare melanoma mutations (such as $BRAF^{G469}$, $KRAS^{G12}$ and $NRAS^{G12}$) and thus, might impact the reliability of the statistics.

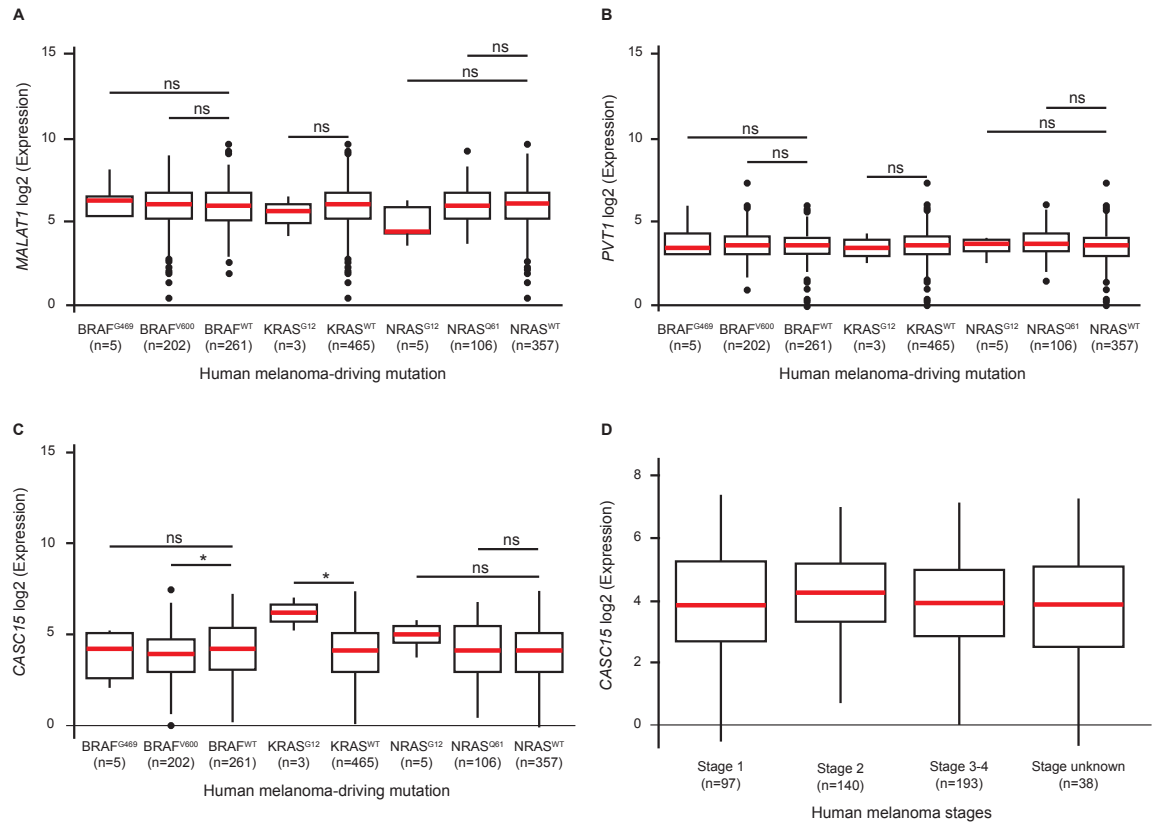


Figure 14: *CASC15*, *PVT1* and *MALAT1* expression in human patient expression database

A Human *MALAT1* expression level according to melanoma mutation status. **B** Human *PVT1* expression level according to melanoma mutation status. **C** Human *CASC15* expression level according to melanoma mutation status. **D** Human *CASC15* expression level according to melanoma clinical stage. All Data were collected on The Cancer Genome Atlas (TCGA) and analysed by Nicolas Servant. Statistics were analysed with Wilcoxon test and one star correspond to $P<0,01$.

f. Melanoma induction triggers changes in skin pigmentation cells

In addition to melanoma development and aggressiveness, I observed abnormalities in the pattern of xantophore pigmented cells (Hirata et al., 2003; Hirata et al., 2005; John, 1997). Indeed, some of mitfa::NRAS^{G12D} fish (independently of their lncRNA mutant status) harboured patches of orange pigmentation resembling those of goldfish. This orange colour is due to xantophore cells that produce carotenoide pigment and could be due to (1) proliferation of the cell type and/or (2) over-production of the pigment. By back-crossing these NRAS^{G12D} fish with unpigmented Casper fish (White et al., 2008) (Figure 15C), I amplified the appearance of orange pigmentation (referred to as Spritz; Figure 15A) compared to melanocyte deficient reported fish lines as Nacre (Figure 15B) or Casper (Figure 15C). As the obtained progeny is lacking melanocytes but still present increased orange pigmentation, it appears that NRAS^{G12} ectopic expression can also alter proliferation or pigment production of other pigmented cell lines.

Figure 15: Associated melanoma induced xantophore proliferation

A Spritz zebrafish in Casper (top) or Nacre (bottom) background (22 weeks post fertilization). **B** Nacre zebrafish (22 weeks post fertilization). **C** Casper zebrafish (22 weeks post-fertilization). Black bar represent 1cm.

II. Determining the function of additional syntenic lncRNAs: linc-*myc* mutants exhibit multiple developmental defects.

The lncRNA Plasmacytoma Variant transcript 1 (*PVT1*) represents an ensemble of 25 different lncRNA isoforms (with multiple TSSs, splicing patterns and ATAs). Located downstream of the protein-coding gene *MYC, PVT1* has been reported to be up-regulated in many cancers and to play oncogenic and tumor-suppressor functions (Cho et al., 2018; Dang, 2012; Kalkat et al., 2017; Nilsson and Cleveland, 2003). In zebrafish, both the *myca* and *mycb* protein-coding genes have genomically adjacent lncRNAs that display conserved position, orientation and splicing patterns to *PVT1* (Figure 16A). To target the *PVT1* zebrafish orthologs linc-*myca* and linc-*mycb* for genomic disruption, two different strategies were followed: 869bp/674bp TSS deletions (linc-*myca* and linc-*mycb* respectively), for which I only obtained heterozygous individuals, and 1,1kb/2,2kb 3' end deletion (respectively linc-*myca* and linc-*mycb*), for which I obtained individual homozygous deletion mutants. This strategy allowed me to delete 50% of 631 nucleotides of the lnc-*myca* transcript and 75% of 1403 nucleotides of the lnc-*mycb* transcript (Figure 16A).

To investigate a potential co-regulation, I analysed the expression of the zebrafish *PVT1* orthologs and their adjacent protein-coding genes across several tissues by qRT-PCR. Zebrafish lnc-*myc* and adjacent *myc* are both enriched in the brain (Figure 16B). My results also indicate that lnc-*myca* shows very similar expression pattern with the transcription factor *myca* (Figure 16 B&D), suggesting that the two genes are co-regulated. These similarities in expression are however not present between *mycb* and lnc-*mycb,* which are enriched in the ovary and testis respectively. Concerning human lncRNA and protein-coding gene orthologs, expression data from GTEx shows that *PVT1* (precise isoform was not indicated) and *MYC* also show similar expression pattern (*PVT1* being less expressed than adjacent protein-coding gene), suggesting conservation of shared DNA regulatory motifs (Figure 16C&E). These results correlate well with previous observations showing that *PVT1* regulates *MYC* RNA and protein levels (Tseng and Bagchi, 2015).

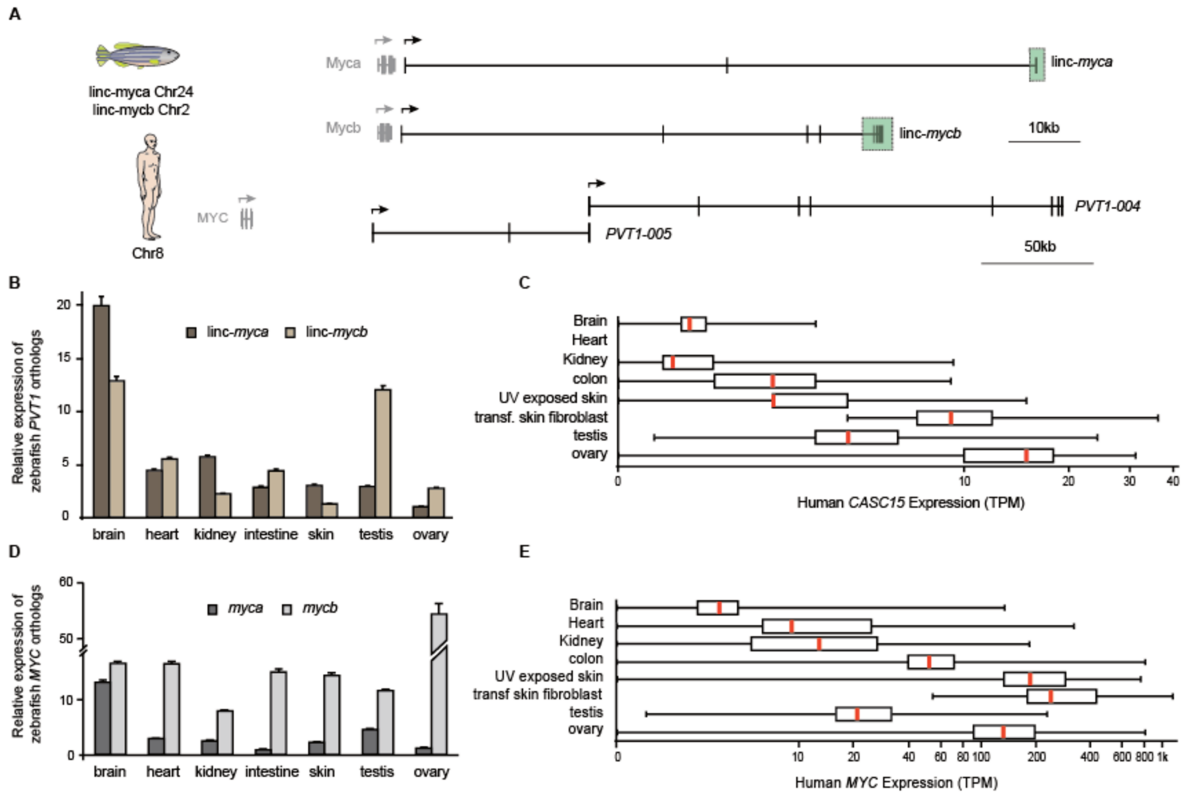


Figure 16: Characterisation of lnc-*myc* and their adjacent protein-coding genes expression patterns.

A Genomic locus of zebrafish *myca*/lnc-*myca*, *mycb*/lnc-*mycb* and human *MYC/PVT1* locus. The three lncRNAs are localised downstream of *MYC* and in the same orientation than the protein-coding gene. Zebrafish lnc-*myca* and lnc-*mycb* mutant were generated through genetic deletion of the 3' part of the transcript illustrated by green boxes. **B** qRT-PCR analysis of lnc-*myca* (exon 1 to 2) and lnc-*mycb* (exon 4 to 5) expression in zebrafish adult organs (fold change relative to lnc-*myca* ovary). **C** Expression of human *PVT1* in adult organs in transcripts per million (TPM). Data were obtained through the Genotype-Tissue Expression (GTEx) project available on EBI Expression Atlas. **D** qRT-PCR analysis of *myca* (exon 2 to 3) and *mycb* (exon 1 to 2) expression in zebrafish adult organs (fold change relative to *myca* intestine). **E** Expression of human *MYC* in adult organs in transcripts per million (TPM). Data were obtained through the Genotype-Tissue Expression (GTEx) project available on EBI Expression Atlas.

To exclude a compensation mechanism mediated by lnc-*myca* and lnc-*mycb,* I generated lnc-*myc*$^{-/-}$ double mutants and identified only three double homozygous fish among 120 screened. A substantial part (Chi-square statistical test *P-value*<0.0001) of their progeny presented developmental defects, characterised by growth delay, heart oedema, eye and brain hypoplasia (Figure 17 A&B). The phenotype of lnc-*myca*$^{-/-}$, lnc-*mycb*$^{-/-}$ double mutants do not show full penetrance, but a prevalence of 65% of embryos presented developmental defects

(Figure 17C). Moreover, among the 35% wild type looking progeny, only 22% survived to adulthood (defined as 1 month post-fertilization). To characterise the impact of the single and double mutants on adjacent protein-coding gene expression, I analysed lncRNAs and adjacent protein-coding genes levels by qRT-PCR (Figure 17 D&E). The analysis of lnc-*myc* expression profile shows upregulation of the lnc-*mycb* in the lnc-*myca*$^{-/-}$ background and vice-versa, revealing a compensatory mechanism and redundant functions despite the absence of sequence conservation between the lnc-*myc* isoforms (Figure 17D). Furthermore, changes in lncRNA expression do not noticeably alter adjacent *mycb* oncogene transcript levels, whereas *myca* is consistently upregulated in lnc-*myca* deletion (Figure 17E).

Such striking developmental phenotypes have rarely been reported in lncRNA mutants, even on highly expressed lncRNAs such as *MALAT1* (Nakagawa et al., 2012; Zhang et al., 2012). To ensure that the inactivation of both lnc-*myca* and lnc-*mycb* is the only responsible for the observed phenotype (no undesired secondary mutation) and to increase the genetic background diversity (as we identified only three double mutant founders), I have backcrossed each individual mutant line with a neuroD1::GFP line and generated a new population of double lnc-*myca*$^{-/-}$ lnc-*mycb*$^{-/-}$ with a different genetic background (screening for double mutants is in progress). If the partially penetrant developmental defects are confirmed in the new generation, it will support the presence of a compensatory mechanism between lnc-*myca* and lnc-*mycb,* and the morphological phenotype will be characterised more precisely focusing on brain hypoplasia with the aid of the neuroD1::GFP marker integrated as a result of backcrossing.

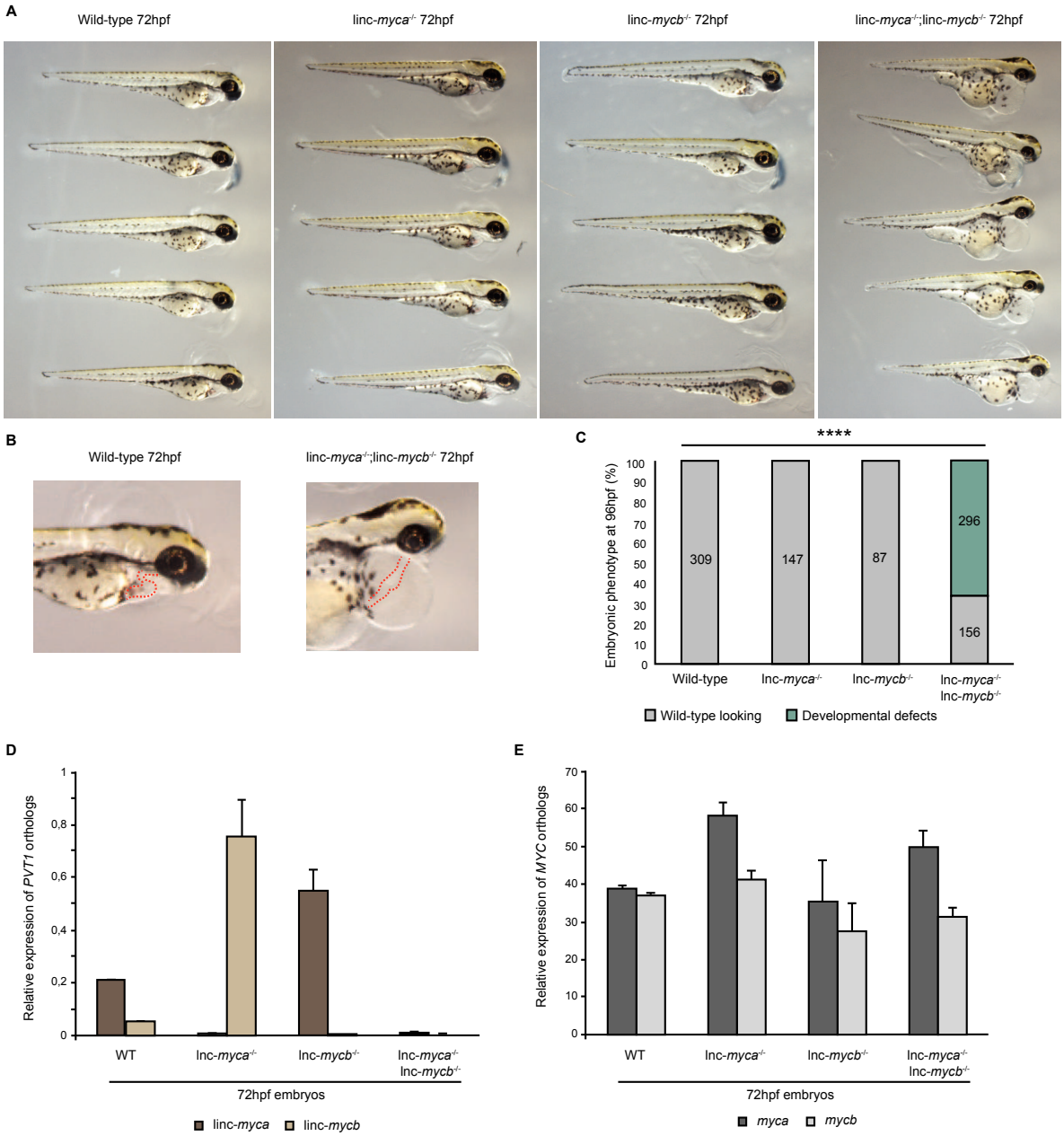


Figure 17: Double mutant lnc-*myca*$^{-/-}$ lnc-*mycb*$^{-/-}$mutants but not single mutants presents developmental defects

A Picture of 72hpf wild type, lnc-*myca*$^{-/-}$, lnc-*mycb*$^{-/-}$ and double lnc-*myca*$^{-/-}$ lnc-*mycb*$^{-/-}$ mutants. **B** Zoom in of 72hpf wild type and double lnc-*myca*$^{-/-}$ lnc-*mycb*$^{-/-}$ mutants, dotted read line define embryonic heart **C** Proportion of embryos presenting pleiotropic developmental defects in wild type, individual lnc-*myca*$^{-/-}$ and lnc-*mycb*$^{-/-}$, and double lnc-*myca*$^{-/-}$ lnc-*mycb*$^{-/-}$ (P<0,0001 according to Chi-square statistical test). **D** qRT-PCR analysis of zebrafish *PVT1* orthologs in wild type, lncRNA single and double mutants 72hpf embryos. **E** qRT-PCR analysis of *myca* and *mycb* in wild type, lncRNA single and double mutants 72hpf embryos

DISCUSSION & PERSPECTIVES

I. Using zebrafish to investigate the *in vivo* function of lncRNAs in melanoma

There are numerous examples of diseases/cancer being associated with mis-expression of lncRNA, supporting the hypothesis that lncRNA have crucial functions in cell stress conditions. The role of a handful of lncRNAs in melanomagenesis has been investigated by lncRNA knock-down or overexpression in human cancer cell lines (Table 1). Several *in vitro* and *in vivo* models are available to investigate melanoma formation. Endogenous melanocytes localization pattern of both mouse and zebrafish model are different from that of humans (surrounding the hair follicle in mice (Chudnovsky et al., 2005) and localized in the hypoderm with to other pigmentation cell types in zebrafish (John, 1997)) and can alter accurate reproduction of melanoma progression. Therefore mouse models are mainly used for xenograft experiments (see Table 1). Zebrafish melanoma tumors (induced by expression of human oncogenes) have been reported to be histologically highly similar to human tumors (Patton et al., 2005). I have also observed alteration of the xantophores expression pattern (Figure 15) in the mitfa::$NRAS^{G12}$ fish. This result suggests that the $NRAS^{G12}$ oncogene itself or melanocytes interactions with adjacent cells impact inner properties of xantophores, inducing over-proliferation and/or over-pigmentation of this particular cell type. Therefore, zebrafish melanoma could present differential tumoural environment and/or cellular interactions compared to humans.

a. *Do rapidly evolving lncRNAs have conserved functions in cancer?*

i. *menhir,* a zebrafish melanoma tumor suppressor

My thesis reports the first investigation of the putative ortholog of the *CASC15* transcript through *in vivo* knock-out. Besides being syntenic, *menhir* and *CASC15* show conserved expression profiles across organs and melanoma tumors. Ideally, we would compare lncRNA expression patterns in skin specific cell types, nevertheless we cannot isolate melanophores from zebrafish skin.

My results also show increased $NRAS^{G12}$ melanoma aggressiveness in *menhir* mutants characterized by (1) increased external tumorigenesis, (2) lower survival, (3) advanced melanoma severity and (4) internal high proliferative melanoma tumorigenesis. These melanoma internal tumors are still under histological characterizations, and could correspond

to mucosal melanoma, tumoural proliferation of a metastasis originating from cutaneous melanoma or melanoma transformation of meninx/fascia melanocytes. $NRAS^{G12}$ expression combined with *menhir*$^{-/-}$ genetic background could promote expression of this internal melanoma development. It is also possible that internal melanoma appears in wild type or other lncRNA mutants (even though unreported in other melanoma driving constructs (Michailidou et al., 2009; Patton et al., 2005; Santoriello et al., 2010)). However tumor size would be too small to be externally perceptible before 21 weeks post fertilization, *menhir* would then promote tumor proliferation.

I have also observed (5) increased metastasis potential of human non invasive 501-Mel cell line in *menhir*$^{-/-}$, suggesting that *menhir* is necessary to decrease tumor environment permissiveness. We did not observe significant changes for WN-266.4 and SK-MEL-2 cell types. These two cell lines have reported invasive properties, it is then highly possible that *menhir* deletion effects on tumor environment are too subtle for such aggressive tumor cell lines. Like reported in human *GAS5, HOTAIR* or *SLNCR1* (Chen et al., 2016; Schmidt et al., 2016a; Tang et al., 2013), *menhir* extrinsic role on melanoma tumor invasion properties could impact the basement membrane sensitivity to MMPs, permissive tumor surrounding tissue architecture or other mechanism facilitating tumor cell intravasation or extravasation.

To rescue *menhir* mutant melanoma aggressiveness, I have ectopically expressed a human *CASC15* isoform in mitfa positive cells (marker for neural crest cell lineage that include melanocytes) (Hosono et al., 2017). Our results show slow-down of melanoma progression in the *menhir*$^{-/-}$ + $NRAS^{G12}$ + *CASC15* compared to *menhir*$^{-/-}$ $NRAS^{G12}$ and wild type $NRAS^{G12}$. However, it is likely that increased size of the insertion cassette (increase of 4,1kb) will affect insertion efficiency and therefore melanoma aggressiveness. I have then generated a second construct containing an mitfa::venus cassette (2,8kb) and injected it in the *menhir* mutants. Our results shows that indeed construct size could have an impact on melanoma aggressiveness, although our *CASC15* rescue construct shows a significant decreased in tumorigenesis despite of an overexpression of the melanoma driving oncogene compared to the $NRAS^{G12D}$ + venus construct. As *CASC15* and $NRAS^{G12}$ are expressed under the control of the same promoter mitfa in the same transgenic construct. It will then be necessary to confirm that the duplication of the promoter does not impair on $NRAS^{G12}$ expression, even though the two promoters are separated by 2,5-4,2 kb. I have also attempted to rescue *menhir*$^{-/-}$ melanoma aggressiveness phenotype with a transgenic construct including a mitfa::*menhir* cassette, but this construct appeared to be toxic for zebrafish embryos as I observed high lethality in early developmental stages. I was still able to obtain a few fish with

this integration, suggesting that the toxicity is not coming from *menhir* cDNA, although the number of survivors was not high enough to draw any conclusion.

ii. *CASC15/menhir* conserved properties

In physiological conditions, my results show conserved expression of *CASC15*/*menhir* in the ovary. Unlike human *SOX4*, zebrafish *sox4a/b* protein-coding transcripts are not enriched in the ovary, probably due to structural/molecular differences between human and zebrafish ovary/oocytes. In human cancers, *CASC15* has been described to play *cis* or *trans* regulatory function depending on the cancer subtype, several papers reporting that *CASC15* activates *SOX4* expression (Chen et al., 2016; Fernando et al., 2017; Merdrignac et al., 2018). In zebrafish, deletion of *menhir* main isoform TSS leads to mis-regulation of both *sox4a* and *sox4b* that is located on a different chromosome than lncRNA, up-regulating both isoforms in ovary, and mainly *sox4a* in skin. However, lncRNA-triggered disruptions of *sox4a/b* protein levels in ovary do not alter ovarian structure or fish fertility in standard laboratory conditions. In zebrafish brain, where *menhir* deletion results in production of alternative isoforms, *sox4a* was the only protein being up or down-regulated according to the biological replicate. To conclude, zebrafish *menhir cis* and/or *trans* mode of action still remains to be characterized.

iii. *The CASC15* function in human cancer

Previous publication on human patient samples and cancer cell lines propose differential function of the *CASC15* transcripts. These studies were mainly carried out using knock-down and overexpression of *CASC15* in different cancer subtypes. *CASC15* has been reported as a tumor suppressor independent from *SOX4* in neural crest lineage cancer neuroblastoma and uveal melanoma (Mondal et al., 2018; Russell et al., 2015; Xing et al., 2017). In addition, it has been shown to be involved in the proliferation/invasion switch in $NRAS^{G12}$ melanoma (independently from *SOX4*) (Lessard et al., 2015) and as an oncogene, regulating adjacent *SOX4* protein in hepatocellular carcinoma, acute myeloid leukemia, intra-hepatic cholangiocarcinoma and gastric cancer (Chen et al., 2016; Fernando et al., 2017; Merdrignac et al., 2018; Wu et al., 2018). Taken altogether, it appears that *CASC15* regulates multiple targets and could act *in cis* or *in trans* dependent on the cancer subtype.

Mondal et al. were the first one to report a clear mechanism of action for *CASC15*, through the analysis of *CASC15* protein interactors. In human neuroblastoma cell lines, *CASC15* regulates ubiquitin protease USP36 intra-cellular localization through direct interaction. This binding prevents formation of the CHD7-USP36 complex, leading to

degradation of the transcription factor and down-regulation of SOX9 target gene (Mondal et al., 2018). Similar mode of action of *CASC15* has not yet been described in other cancer cell lines, nor *in vivo*. Therefore, *CASC15* mechanism promoting oncogenic or tumor suppressor function still remains to be uncovered.

b. *$NRAS^{G12}$ driven melanoma in zebrafish to analyse cancer-related functions of lncRNAs*

$NRAS^{G12}$ melanoma induction in zebrafish *malat1* and *PVT1* orthologs mutants did not alter melanoma severity and progression compared to wildtype $NRAS^{G12}$ fish. Zebrafish *malat1* is a lncRNA with patches of sequence conservation, and the mutant was generated through insertion of a non-invasive polyA that was sufficient to induce complete inactivation of the transcript (see annexe). In human melanoma patients, *MALAT1* level increases with cancer progression and its knock-down in cell lines impairs melanoma invasive properties, suggesting an oncogenic function of *MALAT1* (Tian et al., 2014). If the function if this lncRNA is conserved in vertebrates, *malat1* should conserve its oncogenic role and promote cancer progression (Arun et al., 2016; Sun et al., 2016; Luan et al., 2017). To our knowledge, mouse *Malat1* null mutant were never used to analyse the oncogenic role of the lncRNA. It is therefore possible that *Malat1* null mutants do not show noteworthy defects in standard laboratory conditions because the lncRNA is primarily necessary during stress conditions such as cancer development. A reasonable explanation for the absence of decreased melanomagenesis in the tested zebrafish *$malat1^{-/-}$* mutant might be linked to the melanoma mutational status. Indeed *MALAT1* appears to be slightly less expressed in $NRAS^{G12}$ melanoma patient data collected from TCGA, suggesting that *MALAT1* oncogene potential may be less associated with $NRAS^{G12}$ driven melanoma. However, this reduction in *MALAT1* was considered as non-significant due to the low number of patient samples with this specific NRAS mutation in the TCGA database, therefore additional data will need to be collected to conclude if *MALAT1* could have a role in $NRAS^{G12}$ driven melanoma.

In human cancerogenesis, *PVT1* locus has been reported to have differential functions (Cho et al., 2018; Tseng and Bagchi, 2015; Werner et al., 2017). We did not observe any changes in melanoma progression in lnc-*myc* individual mutants. *PVT1* expression appears to be stable among melanoma driving mutations, therefore oncogene mutational status should not affect lnc-*myc* putative role in melanoma aggressiveness. Our results show a compensation mechanism between lnc-*myca* and lnc-*mycb* isoforms, this redundancy could

interfere with the putatively conserved role of the transcript in melanomagenesis. Because of the low survival of lnc-*myc* double mutants, we will not be able to assess melanoma progression in lnc-*myca*$^{-/-}$ lnc-*mycb*$^{-/-}$ genetic background.

Inactivation of *CASC15* ortholog *menhir* is the only lncRNA mutant where I observed differential NRAS^{G12} melanomagenesis. *CASC15* seems to be correlated with the NRAS^{G12} mutational status, as we observed an increase of the invasion potential in the *menhir* mutant of the 501-Mel NRAS^{G12} cells and that *CASC15* first investigations in melanoma were also conducted in NRAS^{G12} cell line (Lessard et al., 2015). Looking at the TCGA database, we observed an increase of *CASC15* expression in RAS^{G12} mutations, however not significant for NRAS^{G12} (probably because of the low number of data about this mutation). To assess if *menhir*$^{-/-}$ melanoma aggressiveness can be observed in another human oncogene background, I have injected mutant zebrafish with a mitfa::NRAS^{Q61E} construct (kindly provided by Adam Hurlstone, Manchester, UK) and generated more than 100 transgenic individuals. However, NRAS^{Q61E} mutation was not sufficient to efficiently promote aggressive melanoma (data not shown). Supporting our results, this construct was previously reported to be aggressive only in a p53 inactivated background (Dovey et al., 2009), although the Zon lab demonstrated the invasive potential of this mutation (McConnell et al., 2018). This recent report suggests that deeper investigations in the relationship between mutations oncogenic potential and zebrafish genetic backgrounds are necessary. Indeed, it has been reported that the different activating mutation of NRAS drive melanomagenesis differently in cancer cell lines (Grill and Larue, 2016; Posch et al., 2018). It was however not confirmed in patients as the cohort of human melanoma NRAS^{G12} sample is low (Posch et al., 2018).

c. Combining zebrafish therapeutic devices and RNA-based treatment to target cancer

i. Zebrafish as a devices to assess drug efficiency

As reported previously, zebrafish can be used to investigate melanoma through forward and reverse genetics or xenografts. Zebrafish embryo is also an ideal model to perform therapeutics and drug screening due to its small size and ex-utero development. It is indeed possible to perform chemical screening on 96 well plates, each well containing 2-5 embryos. Moreover, zebrafish embryos are permeable to soluble active substances and present similar sensitivity to drug as mice (Fior et al., 2017) and similar concentration/lethality ratio as human (Mimeault and Batra, 2013). The possibility to perform

xenografts also makes possible comparisons of the drug impact on the organism before/after the cancer cell implantation.

Several drug tests have been conducted on zebrafish embryos to investigate cancer cell properties such as melanoma cell angiogenesis (Nicoli et al., 2007) or the impact of neural crest suppressors (White et al., 2011). Zebrafish embryos have also been used to identify suitable treatments and sensitivy of multiple cancer subtypes (Fior et al., 2017; Ju, 2015). At the time of patient customized therapy, zebrafish embryos rise as an effective low cost model to assess individual patient tumor sensitivy to chemotherapy and/or alternative treatments *in vivo*, to design the best treatment combination for each tumor and propose the optimal therapeutic strategy to each patient (Astone et al., 2017).

ii. LncRNA molecules as innovative therapeutics

In therapeutics, RNA molecules are either taken in account as gene target or as drug device to treat several diseases. Indeed, the possibility to target mRNAs of proteins considered as "undruggable" molecules presents high potential for treatment design. With the advances in description of their roles in human diseases, lncRNAs became notorious drug targets (Lavorgna et al., 2016; Matsui and Corey, 2016). For example, targeting of *SAMMSON* in patient-derived melanoma xenograft model has been reported to reduce cancer progression (Leucci et al., 2016). *In vivo,* successfully reported targeting of mRNAs and lncRNAs have been achieved through ASO-mediated inhibition (Matsui and Corey, 2016). As the delivery of RNA inhibitory molecules has been one of the major challenges of the field, multiple strategies (lipid nanoparticles, polymers, cell penetrating peptides, etc.) are actually under investigation (Lavorgna et al., 2016).

The RNA molecule itself can also act as a drug. RNA aptamers and mRNA vaccines have been reported to efficiently modulate disease progression (Vinores, 2006) Li et al., 2013; Reautchnig et al., 2017). Pegaptanib, a RNA aptamer molecule, is a current therapeutic strategy to treat age-related macular degeneration in humans (Vinores, 2006).

If *menhir/CASC15* tumor suppressor function relies on specific domains or secondary structures, such RNA molecule could be considered as potential therapeutic device. Its efficiency could be assessed *in vivo*, monitoring cancer progression in zebrafish embryos.

d. Outlook

i. *menhir* melanoma invasiveness analysis

We described and characterized melanoma a *menhir* mutant and the rescue of external melanoma lesion phenotypic aggressiveness over 16 weeks through human *CASC15* expression. We are now investigating *menhir* mutant and its rescue on melanoma invasion at a vertical level through histological analysis. To do so, we are generating coronal and longitudinal sections of melanoma-induced fish and we will quantify cancer invasiveness in wild type and *menhir* mutants (as reported in (Neiswender et al., 2017). In human, brain is one of the most frequent and aggressive metastatic locus (Sandru et al., 2014; Shain and Bastian, 2016). We have extracted the brain of wild type and *menhir*$^{-/-}$ and we will assess cancer invasiveness looking for metastasis, amplifying human NRAS^{G12} transcript by qRT-PCR.

We have also assessed if *menhir* mutant embryos present higher permissiveness for human melanoma 501-Mel invasion. The next step is to look into *CASC15* intrinsic function on melanoma cells invasiveness. Human 501-Mel NRAS^{G12} cells express two isoforms of the *CASC15* lncRNA. To ensure transcript inactivation, I will target *CASC15* isoforms common 3' ends. I will then perform 501-Mel and *CASC15*$^{-/-}$ 501-Mel cells xenografts on wild type zebrafish 2 dpf embryos, according to the protocol used previously. I will characterize *CASC15* mutant cells metastatic behavior *in vivo* along with cell proliferation properties *in vitro*. I will also attempt to rescue any observed phenotype by introducing zebrafish *menhir* full or partial gene in *CASC15* mutant cells, confirming conserved function and identifying lncRNA functional domains.

ii. Identification of *CASC15/menhir* protein interactors to decipher lncRNA functionality

LncRNA's molecular mechanisms are often deciphered through analysis of transcript partners (Engreitz et al., 2016). As described in the introduction, lncRNA's functionality in gene regulation is most of the time conferred by its protein interactors. Indeed, even when binding to DNA or RNA molecules, lncRNAs often interact with proteins to carry out their functions (Engreitz et al., 2016; Marchese et al., 2017).

Recently, ChIRP-MS (Chromatin-Immuno Precipitation followed by Mass Spectrometry) in human neuroblastoma cell lines revealed 20 potential protein interactors of *CASC15-003* isoform (Mondal et al., 2018). In these conditions, *CASC15* was reported to

interact mainly with nucleolar and RNA binding proteins, including messenger and ribosomal RNA related proteins. Indeed, *CASC15* described interactors are NOP14 (Nucleolar protein 14) and POP1 (Ribonucleases P/MRP protein subunit) that are respectively involved in rRNAs and tRNAs catalytic processes (Mondal et al., 2018).

To analyze *CASC15/menhir* function, I will compare protein interactome of *menhir* and *CASC15* to identify their common partners. To do so, I will use a high-throughput technic developed in the Shkumatava lab: in cell protein-RNA interaction or incPRINT (Graindorge et al., in revision). incPRINT method consists in the measurement of RNA-protein *in vivo* interaction through quantifiable luminescence reading and has proven its efficiency, unveiling known and novel interactors of the *Xist* long noncoding RNA. Unlike classical immune-precipitation technics, incPRINT has the potential to identify transient RNA-protein interactions and protein partners of lowly expressed lncRNAs in cellular conditions. The identification of *menhir* and *CASC15* common interactors would confirm the observed functional conservation of syntenic lncRNAs with no sequence conservation and highlight their potentially preserved molecular mechanisms.

iii. Incorporate *CASC15/menhir* role to cancer-related pathways

Once we have deciphered lncRNAs mechanism of action through identification of their proteins partners, we will investigate *menhir* and *CASC15* target genes by analyzing transcriptome in healthy and tumor tissues of zebrafish mutants (qRT-PCR, RNAseq). Indeed, conservation of *menhir*/*CASC15* molecular targets would reinforce our hypothesis of syntenic lncRNA's conserved function. Moreover, the description of *CASC15* interactions with melanoma inducing mutation and its integration in melanoma driving pathway will present high value for the cancer lncRNA field.

iv. Determine *CASC15/menhir* differential function in cancer subtypes

As described earlier, *CASC15* appears to have differential functions according to the cancer subtypes. If its conserved functionality with zebrafish *menhir* is confirmed, we would like to assess *CASC15* function *in vivo* in other cancer subtypes.

As melanoma is a neural crest cell derived cancer, we would like to investigate *menhir* functionality in other neural crest lineage cancers (Maguire et al., 2014) such as neuroblastoma for which *CASC15* has already been described as a tumor suppressor (Mondal et al., 2018; Russell et al., 2015). Expressing human oncogene under the control of tissue specific promoters, it is possible to induce several *CASC15* related cancer: such as

neuroblastoma (Mondal et al., 2018; Russell et al., 2015) induced by dbh:MYCN (Corallo et al., 2016), uveal melanoma (Xing et al., 2017) induced by mifta::GNAQ^{Q209} (Mouti et al., 2016), or leukemia (Fernando et al., 2017) induced by overexpression of c-Myc, Notch intracellular domain or TEL-AML1 (Shen et al., 2013; Teittinen et al., 2012; Langenau et al., 2003).

CASC15 multiple cancer mode of actions have never been investigated in the same assay. It would be important to compare *CASC15* functionalities in different cancer cell lines through analysis of similar inactivation strategies. We could also investigate mechanism of action of the different *CASC15* isoforms using incPRINT to assess isoform and cancer subtypes differential protein partners.

II. Deciphering phenotype of lncRNA mutants

Despite low expression levels, lncRNAs are often located in close proximity to developmentally important genes and involved in regulatory loop with major developmental transcription factors such as *HOX* or *SOX* family (Alam et al., 2014). Compared to protein synthesis, lncRNAs transcription is a fairly rapid process. Their inner properties and functions in gene regulation make lncRNAs an effective, highly tissue specific and rapid answer for cells to adapt to external or internal stimuli (Marchese et al., 2017) and to regulate physiological or pathological processes. In absence of stress conditions, inactivation of several lncRNAs have been reported to lack phenotypic consequences in living organism (Gouzardi et al., to be publish; Amândio et al., 2016; Eißmann et al., 2012; Nakagawa et al., 2012; Zhang et al., 2012), or to present low penetrant and/or subtle phenotypic defects (Nakagawa et al., 2014; Standaert et al., 2014), suggesting the existence of lncRNAs compensatory mechanism. *In vivo* description of lncRNAs stable inactivation-related morphological or behavioral defects are infrequently reported (Bitetti et al., 2018; Grote et al., 2013).

Even though my selected lncRNAs were reported to be expressed in zebrafish development (Ulitsky et al., 2011), their expression levels were insufficient to be detectable by in situ hybridization, limiting spatial characterization of their expression profile. Moreover, embryogenesis is a robust evolution-proof process involving several different cell types and multiple compensatory mechanisms (Rudel and Sommer, 2003), which could explain that lncRNAs-dependent developmental defect phenotypes have rarely been reported, the most

severe example so far being *Fendrr* (Grote et al., 2013; Grote and Herrmann, 2014; Sauvageau et al., 2013).

a. Lnc-myc functionality in zebrafish embryogenesis

Among our syntenic lncRNA mutants, only lnc-*myca*$^{-/-}$ lnc-*mycb*$^{-/-}$ displays partially penetrant embryonic defects (Figure 17). In zebrafish, duplicated *myc* and lnc-*myc* show different tissue-specific enrichment suggesting cell/tissue-specific function of the lncRNAs and the protein-coding gene isoforms. Zebrafish *myca* and lnc-*myca* highly similar expression pattern (same qRT-PCR Cts in adult tissues) indicates that the protein-coding and the lncRNA gene share DNA regulatory elements and enhancers, suggesting a possible *cis*-regulatory function of the *PVT1* ortholog on adjacent *myc*. Despite absence of sequence similarities between the two lncRNAs, our results show presence of a redundant compensatory mechanism between the two lnc-*myc* isoforms. Therefore, inactivations of the two lnc-mycs appear to be responsible for the observed developmental defects. The partial penetrance of the phenotype is possibly due to the setting up of an additional compensatory mechanism that we have not identified yet.

Interestingly, adjacent protein-coding genes transcript level do not seems to be modulated by inactivation of the lncRNAs, suggesting that the 3' part of lnc-*myc* is not necessary for *cis* regulation of *myc* protein. The slight up-regulation (1.2-1.5 fold) observed on *myca* protein in lnc-*myca*$^{-/-}$ context appears to be independent of the lncRNA, as lnc-*myca* up-regulation observed in lnc-*mycb*$^{-/-}$ context does not alter the transcriptional level of the protein. Then, the mutation we generated to inactivate lnc-*myca* possibly leads to deletion of a DNA regulatory element explaining the adjacent protein slight up-regulation. These results need to be confirmed, repeating the qPCR on biological replicates.

In human, both *PVT1* locus and transcript have been reported to differentially regulate *MYC* expression (Cho et al., 2018; Tseng and Bagchi, 2015; Werner et al., 2017). Up to now, no developmental defects have been reported for *PVT1* in mammals. In addition to its described role in cancer, *PVT1* has also been involved in cardiomyocytes physiology, down-regulation of *PVT1* being reported to induce hypertrophy of cultured cardiomyocytes (Yu et al., 2015), and in diabetes associated nephropathy, mediating extra-cellular matrix protein accumulation (Alvarez and DiStefano, 2011; Yu et al., 2015; Zhigui et al., 2016).

b. Outlook

Additional experiments are necessary to characterize lnc-*myc* developmental defects phenotype. In human, *PVT1* has been reported to stabilize *MYC* protein, interacting with *MYC* inner inactivation domain (Tseng et al., 2015). In zebrafish, *myca* and *mycb* isoforms do not show RNA sequence similarities but share several homology blocks at the protein level. Single *myc* mutants have been generated (CRISPR-Cas9 mediated integration for *myca* and small deletion for *mycb*) by the Zebrafish China Resource Center. No phenotypic consequences were reported for these mutants, however no informations were supplied about characterization of a double *myca/mycb* mutant. To characterize the impact of the lnc-*myc* deletion on adjacent protein-coding gene, we will perform Western blot on *myc* proteins in pooled embryos biological replicates.

To insure that the developmental phenotype observed in lnc-*myc* double mutant is actually due to lncRNAs 3' deletions and to increase the population of double mutant founders (initial screen allowed us to identify only two females and one male lnc-*myca*$^{-/-}$ lnc-*mycb*$^{-/-}$), I have backcrossed each single mutants with a neuroD1::GFP line, labeling zebrafish brain and eyes. This marker will allow us to investigate deeper the developmental defects focusing on observed brain and eye hypoplasia phenotypes. We will also be able to assess *myc* protein impact on the phenotype, as zebrafish *myca* and *mycb* mRNAs have been reported to be well expressed in these embryonic tissues (Kotkamp et al., 2014), Thisse et al., 2014).

The partial penetrance observed in lnc-*myc* double mutants suggests that there could be a second compensation mechanism attenuating the loss of lnc-*myc* on embryogenesis. Comparison of the transcriptome of single mutants, double lnc-*myc* mutant embryos and double lnc-*myc* escapees could help us to unveil such redundancy and identify the actors of the compensation.

To assess if lncRNA functionality is conserved throughout evolution, we will attempt to rescue double mutants embryonic phenotype using the Tol2 transgenesis system. To do so, we will insert a *PVT1* isoform we amplified from human brain cDNA (corresponding to *PVT1-004,* see Figure 16A), which will be expressed under the control of the zebrafish ubi (ubiquitous) promoter. If the developmental defects can be rescued by human *PVT1* isoform, we will deeply investigate the putatively conserved molecular mechanism of lnc-*myc*/*PVT1* lncRNAs. Such results would have a great impact on the lncRNA field, supporting synteny as a determinant of lncRNA function.

III. Synteny as a determinant of functional conservation?

An evolutionary pressure on lncRNA genomic positions has been reported in vertebrates and in plants (Hezroni et al., 2015; Mohammadin et al., 2015). Recent study associates synteny of lncRNAs with chromatin organization structures, suggesting that syntenic lncRNAs overlap with CTCF binding sites and participate to the determination topologically associating domains border (Amaral et al., 2018). Our results show a correlation between conserved position of lncRNAs and regulatory DNA elements. Indeed, lnc-*myca* and adjacent *myca* protein show a very similar expression pattern. This result suggests that lnc-*myca* is positionally conserved in order to share DNA regulatory motifs with the adjacent protein-coding gene. However, our analysis of lnc-*myc* mutants does not support *cis* regulatory mechanism of the lncRNA. As for *menhir*, its expression profile in healthy and cancer tissues is conserved in the evolution, unlike the one of adjacent protein-coding gene *sox4a.* Therefore *menhir* position could be conserved throughout evolution to preserve its adjacency to DNA elements regulating its expression pattern.

We identified a potentially conserved function between two syntenic lncRNAs, *menhir* and *CASC15,* that does not present sequence conservation. Indeed lncRNA sequence evolves faster than protein-coding gene, suggesting that long noncoding RNA sequence is not functionally determinant (Ulitsky, 2016). Indeed, lncRNA do not act on their own but through their interaction with proteins and/or nucleic acids (Marchese et al., 2017) that could be conserved despite lack of sequence conservation. Burge lab has actually report that RBP specific binding to RNA is highly dependent on RNA binding site flanking composition or secondary structure (Dominguez et al., 2018), that can be conserved for lncRNA in the evolution as reported in drosophilid *roX* (Quinn et al., 2016).

The main objective of my PhD was to determine if synteny could be a determinant of lncRNA's function conservation. The results I have obtained until now support this hypothesis as *menhir* and *CASC15* homologs appears to present conserved tumor suppressor functionality despite absence of sequence conservation. I have also uncovered unknown developmental role and redundancy of *PVT1* homologs, reporting the first example of lncRNA compensatory mechanism *in vivo* and featuring a second very interesting candidate to investigate for conserved functionality.

Chapter 2: A minimally invasive genome editing approach to inactivate lncRNAs in zebrafish

INTRODUCTION

I. Transient and stable strategies to inactivate lncRNAs *in vivo*

Several genetic strategies exist to inactivate protein-coding genes (exon replacement, in frame stop codon insertion, frameshifts by insertion/deletion, truncation or point mutation of functional domains). As most of these strategies directly impact the integrity of the protein product of the gene, they are not applicable to lncRNAs.

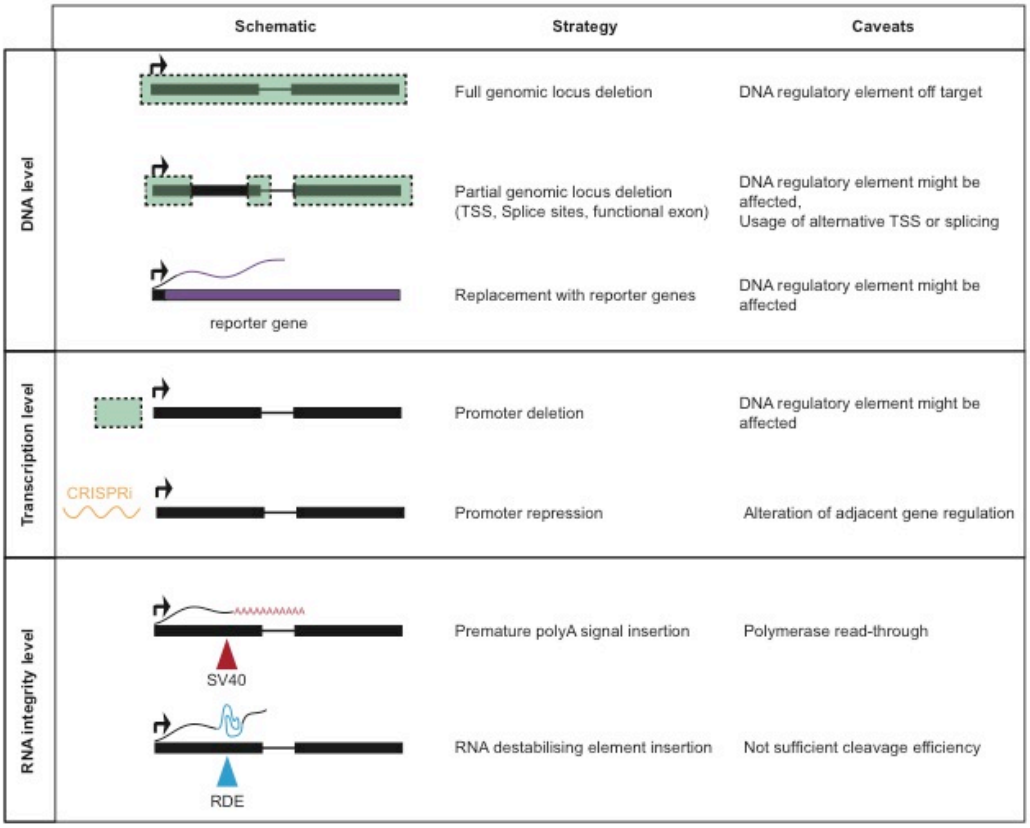
Multiple methods have been developed to inactivate lncRNAs in a transient or stable way. Transient inactivation of lncRNAs usually leads to a decreased level of the noncoding transcript (knock-down) and can be achieved with different techniques such as RNA interference (RNAi), Locked Nucleic Acids (LNAs), morpholinos (MOs), Allele Specific Oligonucleotides (ASOs) or Gapmers (Delás and Hannon, 2017; Schulte-Merker and Stainier, 2014). The common principle of these different technologies is to target the RNA molecule and induce its inactivation or degradation, however they all have several caveats, such as model-specific applications and transient off-target effects.

To decipher more robustly the *in vivo* molecular function of a lncRNA, it is preferable to generate stable genetic inactivation of the transcript. Several stable inactivation strategies (see Table 4) have been reported to efficiently target lncRNAs despite their complex locus architecture (alternative isoforms, hosting of other noncoding RNA, overlapping with adjacent protein-coding genes, spanning hundreds of kb) (Ziegler and Kretz, 2017). Advances in genome editing tools have lead to the development of multiple approaches for lncRNA inactivation including: (1) deletion of DNA sequence resulting in gene truncation (TSSs, SSs, ATAs, functional exon), deletion of the full gene locus or replacement of the lncRNA sequence by a reporter gene; (2) disruption of gene transcription by promoter/TSS deletion or promoter repression (Liu et al., 2017); and (3) RNA destabilisation through the insertion of premature polyA signals or RNA destabilizing elements. As for transient strategies, stable inactivation approaches have advantages and disadvantages (Table 4). Generation of lncRNA full or partial genetic deletions or its replacement with reporter gene is

a direct and efficient approach to inactivate any functional role of the lncRNA locus, although this method is highly invasive and presents caveats. Indeed, several examples of long noncoding locus internal DNA regulatory elements have been reported to be functional (Groff et al., 2016; Paralkar et al., 2016). It is thus essential to distinguish between lncRNA transcript and DNA element related functions (Haemmerle and Gutschner, 2015; Marchese et al., 2017). In contrast to genetic deletion, insertion of short sequences leading to transcript destabilization are promising, less invasive approaches that will likely have a reduced off-target impact on adjacent/internal DNA regulatory motifs. RNA destabilizing element (RDE) can have differential efficiency modulated by target gene structure or physiological characteristics respective to the investigation model.

To choose the best lncRNA inactivation strategy, it is essential to make the distinction between RNA molecule inner functionality, the act of transcription and the integrity of the genetic locus (Bassett et al., 2014; Haemmerle and Gutschner, 2015; Marchese et al., 2017). It is also crucial to analyse the various conserved features of the targeted lncRNA genomic locus (Figure 1) such as sequence, organization, structure and overlap with regulatory elements such as enhancers (Ziegler and Kretz, 2017).

Table 4: Strategies to genetically disrupt lncRNA expression



II. Insertion of RNA destabilizing elements as a minimally invasive approach to inactivate long noncoding RNAs

As creation of robust lncRNA loss of function mutants is a challenge in the ncRNA field, we have tested two minimally invasive approaches to inactivate lncRNAs in zebrafish. Premature polyA signal insertion has already been reported to be efficient to inactivate lncRNA (Eißmann et al., 2012; Grote et al., 2013). Although RNA polymerase has the potential to read through single polyA signal, integration of multiple premature polyA sequences have been reported to inactivate efficiently highly expressed lncRNA (Gutschner et al., 2011). To our knowledge, polyA signal premature insertions have never been used for endogenous gene inactivation in zebrafish.

An alternative strategy to destabilize RNA is the use of ribozymes: RNA sequences functioning as catalytic elements. Indeed ribozyme 3D structure induces protein-independent self-cleavage (Fedor and Williamson, 2005; Hartig et al., 2012). Several ribozyme subtypes have been characterised such as the Hammerhead and the Hepatitis Delta Virus (HDV) ribozyme (Figure 18). The hammerhead ribozyme was the first self-cleaving RNA sequence discovered in the 80s (Hutchins et al., 1986). Hammerhead ribozyme is composed of a 3 base paired helix (I, II, III as reported in Figure 18B) organised in an Y 3D structure (Fedor and Williamson, 2005; Hammann et al., 2007). Hepatitis Delta Virus (HDV) ribozyme is an 80 nucleotide RNA enzyme essential for Hepatitis virus replication (Lai et al., 2002). HDV ribozyme is a complex structure of 5 domains folding (Figure 18A) in the presence of divalent cation ($Mg^{2+,}$ $Mn^{2+)}$ and precise pH conditions, which lead to RNA cleavage occurring at the +1 sequence (Fedor and Williamson, 2005; Nishikawa et al., 2002) (Ke et al., 2004) (Hammann et al., 2007). The HDV ribozyme has been reported as the fastest naturally self cleaving RNA (Fedor and Williamson, 2005; Kapral et al., 2014). Self-cleaving ribozymes insertion has been proposed as an alternative approach for gene inactivation and has been reported to be efficient in different cellular models (Beilstein et al., 2014; Lee et al., 2016; Nomura et al., 2013; Asif-Hullah et al., 2007).

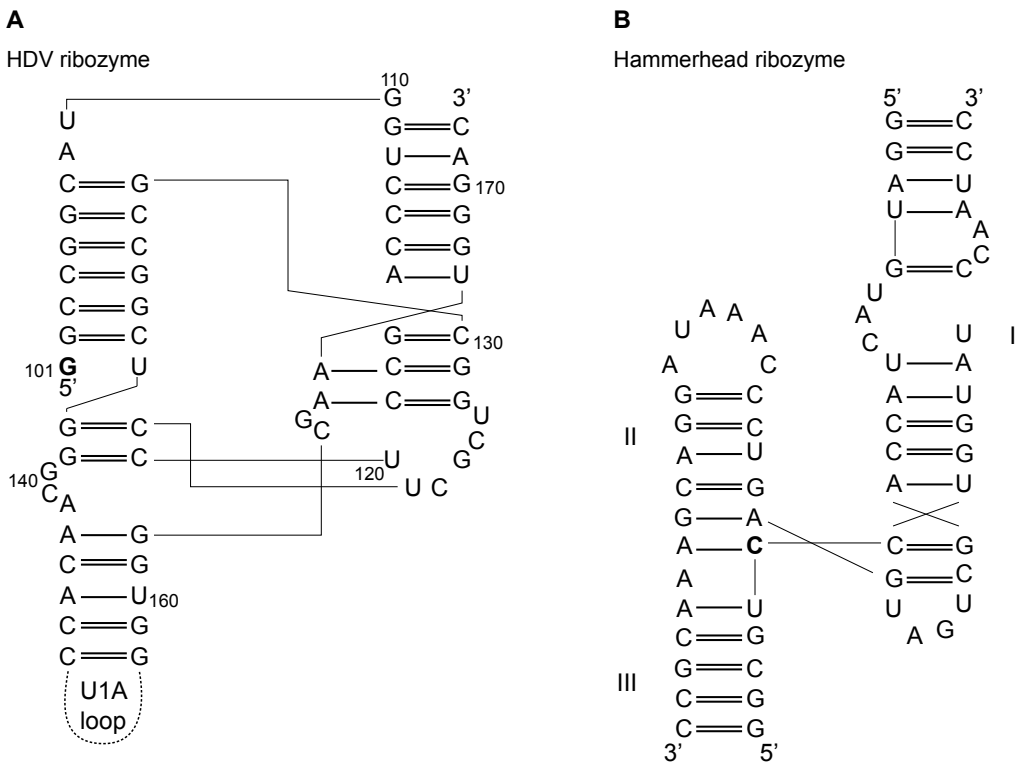


Figure 18: Self cleaving ribozyme differential architecture (adapted from (Hammann et al., 2007))

A The Hepatitis Delta Virus ribozyme has a convoluted pseudoknotted topology. The cleavage occurs in 5' of the first G **B** The active Hammerhead ribozyme is based on a three way junction. The cleavage occurs 3' in bold.

III. Achieving precise short sequence integration in the zebrafish genome

The advent of CRISPR-Cas9 genome editing technology has revolutionized the ability to create desired genetic mutants in many model organisms, including zebrafish (Auer et al., 2014; Hwang et al., 2013; Li et al., 2016) (Cong et al., 2013; Malina et al., 2013; Jinek et al., 2012) Mali et al., 2013). CRISPR-Cas9 is an essential component of bacteria and archea's adaptative immunity and consists of an RNA molecule that guides a non-site specific nuclease to a precise genomic locus (Garneau et al., 2010; Hale et al., 2009) where it induces target DNA double strand break (DSB) that will be primarily repaired by the error-prone Non Homologous End Joining (NHEJ) repair pathway.

DNA double strand breaks can be repaired by three different mechanisms (detailed in Figure 19): NHEJ that ligates damaged DNA ends often generating small insertions or deletions (in/del), The Micro-homology Mediated End joining (MMEJ) that uses short

homology sequence to repair DNA break, and homologous recombination (HR), which perform faithfull DNA repair employing a sister chromatid as a template (Branzei and Foiani, 2008; Hiom, 2010). These repair mechanisms are common to vertebrates, and zebrafish genome contains nearly all genes involved in DNA repair (Pei and Strauss, 2013).

Several examples of CRISPR-related genomic insertion in zebrafish have already been reported (Albadri et al., 2017). All three DNA repair pathway can promote targeted knock-in when CRISPR technology is combined with donor DNA (plasmid, double stranded oligonucleotide, single stranded oligonucleotide). However these insertions can be imprecise, mis-oriented (for NHEJ) and usually occur at low frequency (Albadri et al., 2017; He et al., 2015; Hisano et al., 2015; Hwang et al., 2013; Nakade et al., 2014; Zhang et al., 2016; Yanez et al., 1999) likely due to the rapid cell cycle of zebrafish early embryos (a cellular division every 15 minutes) (Gilbert, 2000).

Thus far, in other organisms, several strategies have been successful to promote HR and precise exogenous insertion in the genome, such as overexpression of the *RAD51* mRNA or depletion of NHEJ actors (Bertolini et al., 2007; Chu et al., 2015; Hoshijima et al., 2016; Maruyama et al., 2015). Because DSB DNA repair pathways are conserved in evolution, we have implemented these technologies in zebrafish to promote precise RNA destabilizing elements in lncRNAs.

IV. Overall objectives of the project

Together with Angelo Bitetti (PhD student in the Shkumatava lab), we have assessed (1) the integration potential of several strategies employed in other model organisms on a lowly expressed syntenic lncRNA lnc-*klf7b*. We also have (2) analysed the RNA destabilizing efficiency of single polyA signal and HDV ribozyme in zebrafish on differentially expressed lncRNAs. Our results show that knock-in (KI) is efficiently promoted in NHEJ-compromised conditions through the inactivation of *xrcc4* protein (component of the ligase IV complex, see Figure 20) and that polyA signal premature insertion is sufficient to deminish the highly expressed lncRNA *malat1*. However, we report that a HDV ribozyme insertion does not appear to direct efficient lncRNA knock-down in the case of a single tested lncRNA insertion.

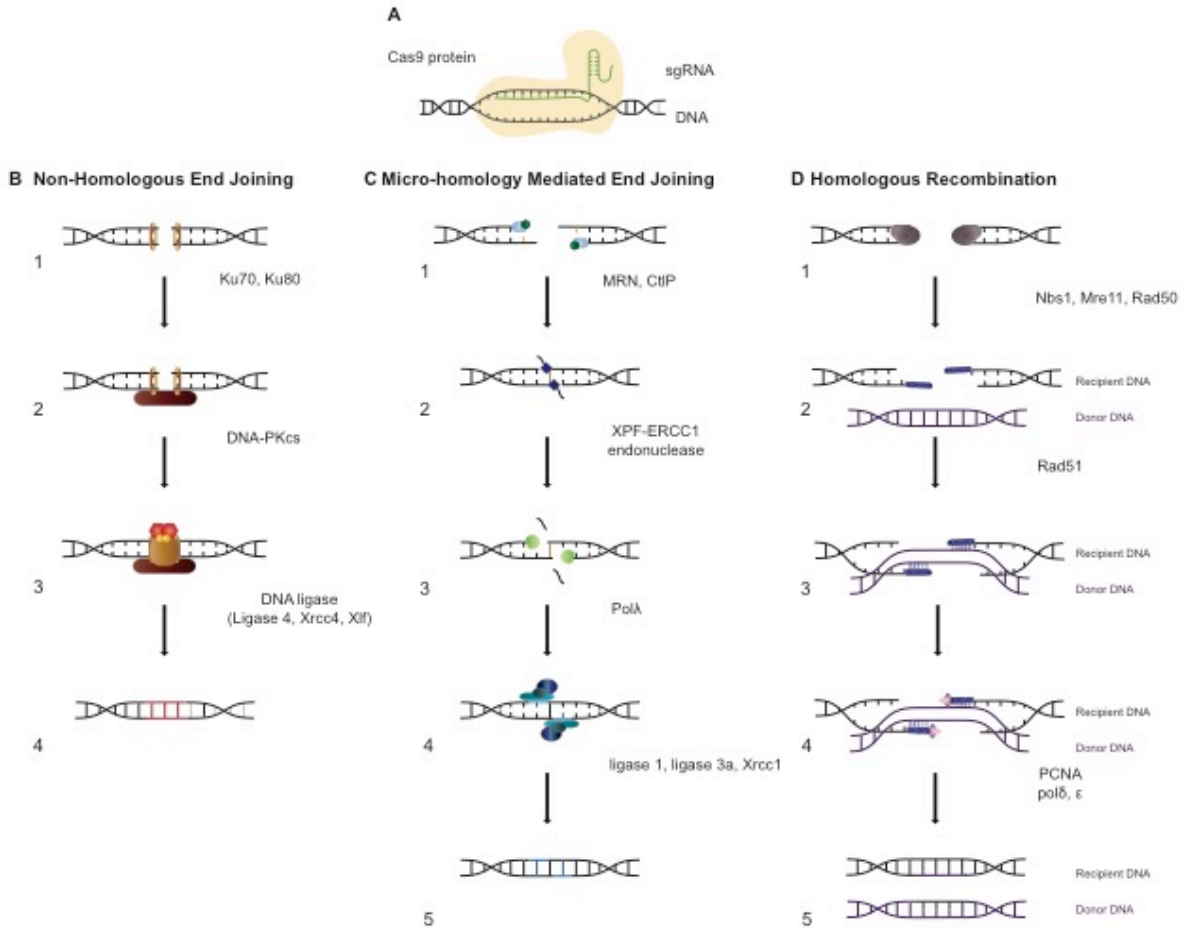


Figure 19: Schematic of various DNA repair pathways in zebrafish

A Illustration of CRISPR-Cas9 induced DNA double strand break (DSB). **B** Schematic of NHEJ DNA repair: Ku proteins slides on DNA ends to form a docking site for NHEJ proteins (1), DNA-PKcs bride the ends together and polymerases processes DNA ends (2) and DNA ligase IV complex achieve the ligation step (3) NHEJ can be an error prone DNA repair mechanism (4). **C** Schematic of the Micro-homology Mediated End Joining DNA repair: MRN aligned strands of 5-25 complementary bp with mismatch ends (1), XPF-ERCC1 endonuclease removes overhangs and mismatched bases (2), polT fill the missing base pairs (3) that are ligated by ligase1, 3a and xrcc1 (4). MMEJ repair can result in insertion/deletion mutations (5). **D** Schematic of Homologous Recombination DNA repair: DNA free ends are binded by Nbs1, Mre11 and Rad50 complex that will initiate a 5'-3' resection of the DNA (1), single stranded DNA associate with Rad51 protein (2), the newly formed nucleoproteic filament will then search for homologous DNA sequences on donor DNA (3) which will results in recipient DNA strand invasion, formation of Holliday junction and DNA synthesis (4). Homologous recombination is the most reliable DNA repair pathway (5).

RESULTS

I. CRISPR-Cas9 mediated genetic deletions as a strategy to inactivate lncRNAs in zebrafish

a. Generation of zebrafish mutant lines for syntenic lncRNAs

To generate genetic zebrafish mutants of syntenic lncRNAs, we took advantage of the CRISPR-Cas9 genome editing technology (Hwang et al., 2013). However, generating knock-out (KO) of syntenic lncRNA loci is not trivial because (1) they lack obvious sequence conserved domains to target, (2) they often overlap with their adjacent protein-coding genes and (3) they often are produced from long loci spanning several hundred base pairs (Ziegler and Kretz, 2017).

To overcome these challenges, I used several strategies to inactivate selected lncRNAs. For example, by targeting their TSSs (transcription start site) or the largest exonic region for deletion, or through the insertion of transcript destabilization elements. As described in Table 5, I encountered several technical limitations and obstacles while generating deletion mutants of my syntenic lncRNA candidates, such as absence of efficient guide RNAs (gRNAs) for lnc-*zfp64* and lnc-*gosr2*, lncRNA's TSS overlapping with adjacent protein-coding gene (lnc-*ppm1bb*), or large lncRNA genetic loci too long for whole-locus efficient deletion and including several DNA regulatory motifs (lnc-*sox4a*, lnc-*myca* and lnc-*mycb*). Despite several attempts, I did not manage to identify homozygous deletion mutants for lnc-*klf7b*, lnc-*myca* TSS, lnc-*mycb* TSS or lnc-*sox4a* 3' deletion. However, I did successfully generated four zebrafish deletion mutant lines: lnc-*ppm1bb* 3', lnc-*sox4a* TSS, lnc-*myca* and lnc-*mycb* 3', and pursued my investigation of syntenic lncRNA functionality in these lines.

Table 5: Generation of syntenic lncRNA deletion lines in zebrafish using CRISPR-Cas9 technologies

lincRNA	Locus	Deletion	Stage
linc-*klf7b*	Chr 9 :29432670-29436453	Full transcript	Heterozygous
linc-*zfp64*	Chr 23 :38,907,088-38,910,731	3' end	No efficient gRNA
linc-*ppm1bb*	Chr 12 :26,951,263-26,960,298	3' end	Homozygous
linc-*myca*	Chr 24 :10,218,039-10,312,078	TSS	Heterozygous
		3' end	Homozygous
linc-*mycb*	Chr 2 :31,737,208-31,812,588	TSS	Heterozygous
		3' end	Homozygous
linc-*sox4a*	Chr 19 :29,161,676-29,270,573	TSS	Homozygous
		3' end	Heterozygous
linc-*gosr2*	Chr 3 :37629355-37644171	3' end	No efficient gRNA

b. *Phenotypic characterisation of long noncoding RNA zebrafish mutants: Characterisation of lnc-ppm1bb$^{-/-}$*

Zebrafish lnc-*ppm1bb* and its human syntenic ortholog annotated as *ENST00000609837* are localised 5' and antisense to the *ppm1b* protein-coding gene (Figure 20A). While human and zebrafish lncRNAs do not share splicing pattern, both transcript have their TSS overlapping with the adjacent protein-coding gene. Therefore, I generated a zebrafish lnc-*ppm1bb* mutant through a 1,9kb deletion (green box) targeting the 3' largest exonic region. This strategy allowed me to delete 86% of a 903 nucleotides transcript, without disrupting the adjacent protein-coding gene promoter.

To investigate their potential for co-regulation, I analysed the expression of lnc-*ppm1bb* and the adjacent protein-coding gene in several adult zebrafish tissues. Using qRT-PCR, I determined the expression pattern of both lnc-*ppm1bb* and the protein-coding gene *ppm1bb* to be enriched in the zebrafish brain (Figure 20 B&C). However lncRNA and protein-coding genes appears to have different fold changes and patterns of expression in other tissues. Although the expression pattern of the human lncRNA is not reported, the human *PPM1B* protein-coding gene is enriched in the brain and gonads (GTEx database; Figure 20D). *PPM1B,* which stands for *Protein Phosphatase $Mg^{2+/}Mn^{2+}$ dependent 1B*, has been shown to dephosphorylate cyclin-dependent kinases (CDKs) (Prajapati et al., 2004).

This protein has been associated with early pre-implantation (Sasaki et al., 2007) and has been shown to negatively regulate necroptosis (Chen et al., 2015). Deletion of the 3' end of lnc-*ppm1bb* is sufficient to fully inactivate the transcript (Figure 20E). The level of *ppm1bb* protein appears to be affected by this genetic mutation but with high inter-individual variation (Figure 21F). The Lnc-*ppm1bb* deletion is 7kb upstream of the *ppm1bb* TSS, however this mutation did not permit us to distinguish between lncRNA molecule or DNA regulatory elements action of adjacent protein-coding gene.

My results show that lnc-*ppm1bb* is not required for zebrafish normal physiology. Indeed, under standard laboratory conditions, zebrafish lnc-*ppm1bb*$^{-/-}$ fish, despite alteration of *ppm1bb* transcript level, are viable, fertile and do not show gross morphological defects (Figure 20 G&H).

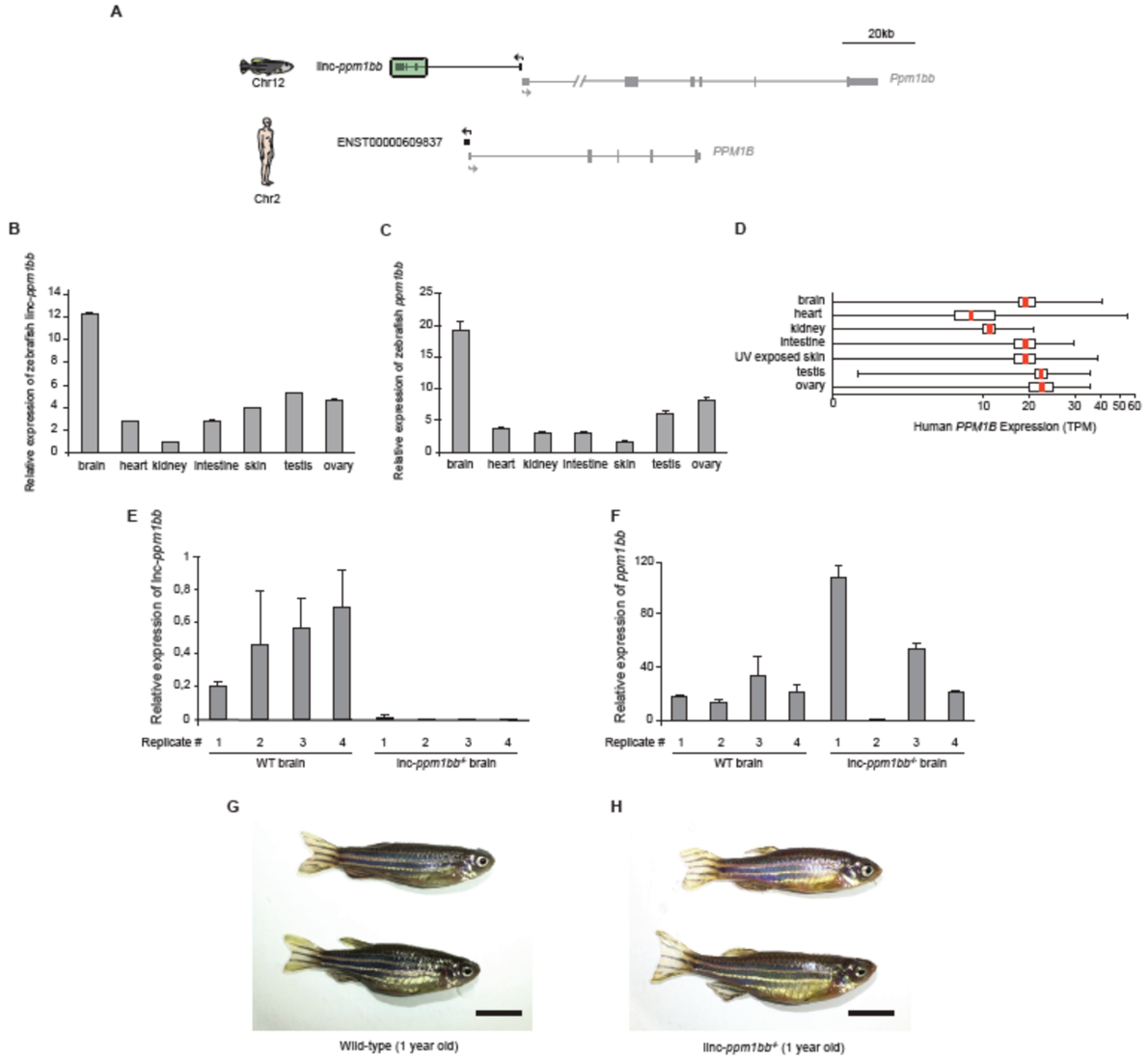


Figure 20: characterisation of lnc-*ppm1bb* zebrafish mutant

A lnc-*ppm1bb* transcript is conserved at the syntenic level between zebrafish and human, positioned in 5' and antisense of *ppm1b* protein-coding gene. Zebrafish lnc-*ppm1bb* mutants were achieved through deletion (green box). **B** qRT-PCR analysis of zebrafish lnc-*ppm1bb* (exon 3 to 4) expression in adult organs (fold change relative to kidney). **C** qRT-PCR analysis of zebrafish *ppm1bb* (exon 3 to 4) expression in adult organs (fold change relative to skin). **D** Expression of human *PPM1B* protein-coding-gene in adult organs in transcripts per million (TPM). Data were obtained through the Genotype-Tissue Expression (GTEx) project available on EBI Expression Atlas. **E** qRT-PCR analysis of zebrafish lnc-*ppm1bb* (exon 3 to 4) between WT and lnc-*ppm1bb*$^{-/-}$ adult organs (fold change relative to *ppm1bb* q-RT PCR lnc-*ppm1bb*$^{-/-}$ brain #2). **F** qRT-PCR analysis of zebrafish *ppm1bb* (exon 3 to 4) between WT and lnc-*ppm1bb*$^{-/-}$ adult organs (fold change relative to lnc-*ppm1bb*$^{-/-}$ brain #2). **G,H** Picture of one year old WT and lnc-*ppm1bb*$^{-/-}$ fish (male above, female under). Black bar represent 1 cm.

II. Insertion of RNA Destabilisation Elements as a minimally invasive approach to inactivate lncRNAs

The generation of lncRNA mutants is more challenging than that of protein-coding genes due to the absence of conserved domains and an open reading frame (ORF). Indeed, minimally invasive single nucleotide changes that lead to nonsense or frame-shift mutations in the ORF of protein-coding genes are not expected to impact lncRNA loci. Creating robust lncRNA loss-of-function mutant is a challenge of the field, and as such, we have tested two minimally invasive approaches to inactivate lncRNA in zebrafish: insertion of a premature polyA signal and insertion of a self-cleaving ribozyme, both of which are expected to lead to transcription destabilisation.

a. Inhibition of the Non Homologous End Joining pathway to promote precise genomic insertion in zebrafish

With the advances of CRISPR-Cas9 genome editing technology, generation of precise insertion in the genome became possible, even though this technique is highly inefficient in zebrafish. To improve precise genetic insertion rates, Angelo Bitetti (former PhD student in the lab) and I have designed a strategy to induce small insertions in a targeted manner by destabilising the Non-Homologous End Joining (NHEJ) pathway to promote Homologous Recombination (HR). To do so, we co-injected into the one-cell stage zebrafish embryos CRISPR-Cas9 components, DNA repair disturbing elements (translation inhibiting morpholinos (MOs) against *xrcc4* (DNA ligase IV) and *xrcc5* (Ku80), or *rad51* mature RNA; Figure 21A) and single stranded oligos (ss oligos) corresponding to the desired insert sequence. We designed two different types of ss oligos, corresponding to an insertion (Figure 21B) or a replacement strategy (Figure 21C), both with different homology arms sizes.

To assess integration events, we inserted HDV ribozyme in linc-*klf7b* and performed PCR on pooled 48 hpf embryos. We did not observed insertions in the embryos co-injected with *rad51* mRNA. To evaluate efficiency of the integration according to different parameters (ss oligo subtype, length or concentration, NHEJ destabilising MO, *Cas9* mRNA or protein), we performed PCR on single 48hpf injected embryos (n=30). Our results show that destabilising NHEJ with *xrcc4* MO leads to an insertion rate of 60% with *cas9* RNA and 80% with CAS9 protein (Figure 21D). We also noticed that, similar to the efficiency to create deletion mutants, the insertion rate was highly affected by the gRNA-directed *Cas9* cutting efficiency.

Using this strategy, we have integrated three different short inserts (1) SV40 polyA signal (131 bp), (2) HDV ribozyme (72 bp) and (3) guanine inducible HDV ribozyme (133 bp) (Normura et al 2013) in three different lncRNA candidates (linc-*klf7b*, *megamind*, *malat1*).

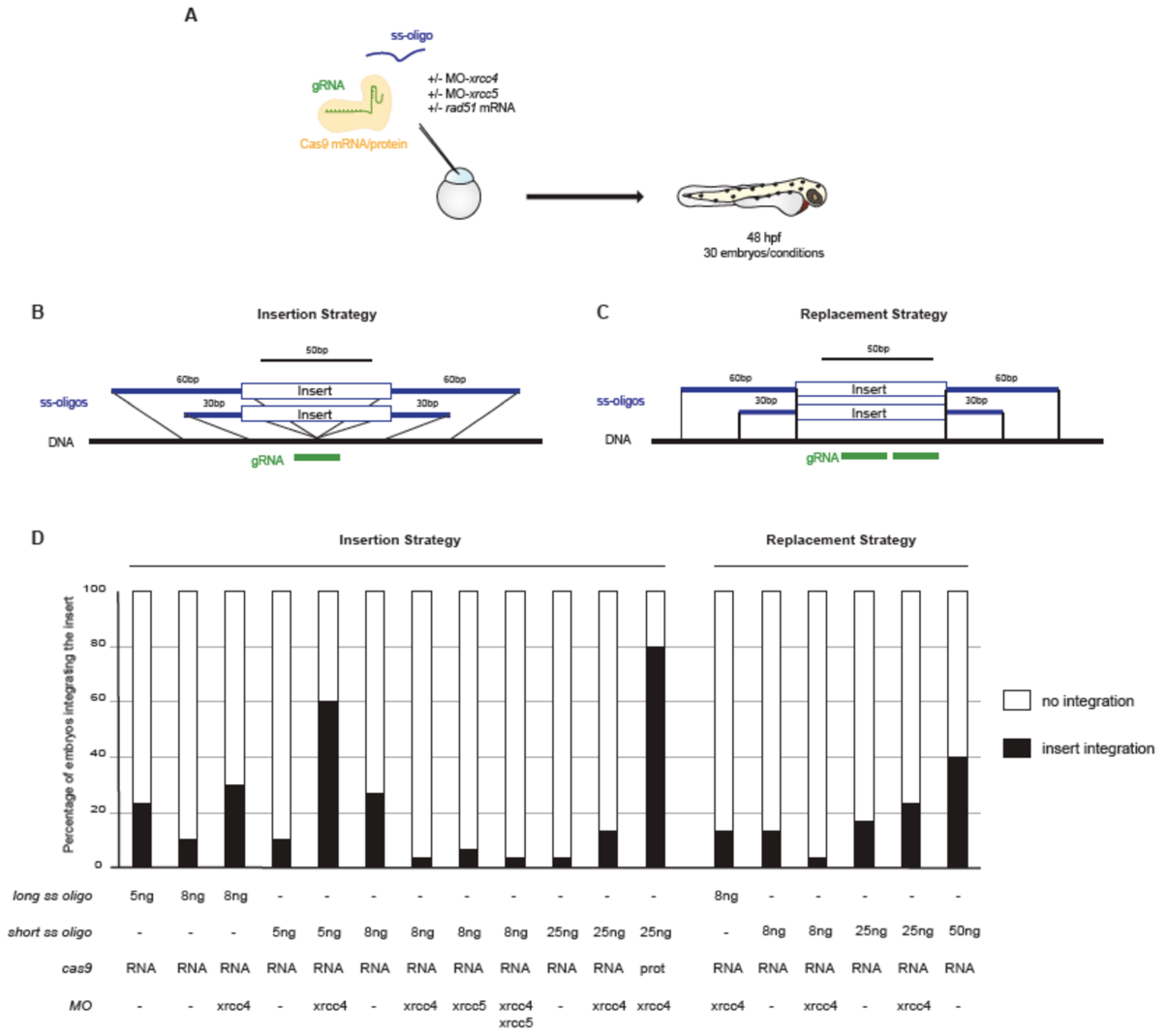


Figure 21: Promoting precise genomic integration in zebrafish through NHEJ inhibition

A Injection procedure to promote precise genomic integration in zebrafish and to screen for efficiency. **B** Single Stranded oligo (ss oligo) design for insertion strategy using single gRNA. The insert will be integrated in the genome at the gRNA double strand break locus. **C** Single stranded oligo design for replacement strategy using couple gRNAs. The insert will replace the DNA in between gRNAs. **D** Integration efficiency (in percent) according to ss oligo length and concentration, *cas9* RNA or protein, and NHEJ inhibiting morpholinos. Results were obtained by PCR on single injected embryo, using 30 embryos per conditions.

b. Technical challenges

It is important to note that the event of integration can lead to target gene fragment duplication/deletion or truncated insert. Although we observed a high rate of integration, PCR screening for the ribozyme insertion was challenging. Indeed, the ribozyme sequence high GC and hairpin secondary structure appear to negatively affect the polymerase chain reaction (PCR) efficiency, making it challenging to amplify and sequence the full-length product (Sing et al 2000, Nelms et al 2011). By optimizing the primer design and PCR conditions, we have overcome the difficulties encountered in ribozyme insertion screening and succeeded to generate linc-*klf7b*^{HDV} (Figure 22A), *malat1*^{SV40} and *malat1*^{HDV} (Figure 22D) homozygous lines.

III. Assessing RNA destabilization efficiency

Numerous *in vitro* and *ex vivo* tests have shown that the HDV ribozyme catalityc activity is present in mammalian cells (Asif-Ullah et al 2007). However the efficiency of ribozyme cleavage and the degree to which the targeted transcript will be destabilised in zebrafish is unknown. To test whether HDV ribozyme cleavage occurs in zebrafish. I have performed a cleavage assay (adapted from 5' RACE technology) in injected embryos (n=60). Targeting linc-*klf7b* (a lowly expressed syntenic lncRNA) with HDV ribozyme (Figure 22A). I detected the presence of several transcript degradation products in the injected embryos in both insertion and replacement strategies (Figure 22B), degradation products 5' are illustrated in Figure 22C. These results suggest that HDV ribozyme is active in zebrafish on lowly expressed lncRNAs.

We also generated polyA signal and Hepatitis Delta Virus (HDV) ribozyme insertion (respectively at 35bp and 5kb downstream of the TSS) in *malat1* lncRNA (Figure 22D). *malat1* is a sequence conserved 7,5kb lncRNA with high ubiquitous expression in embryo stages and adult tissues. As *malat1* genomic region presents high genomic chromatin modifications (Ulitsky et al., 2011), full locus deletion could lead to unintended effects on overlapping DNA regulatory sequences (Zhang et al., 2012). Our results show that the insertion of HDV ribozyme did not lead to *malat1* down-regulation (Figure 22D), likely either because the cleaving capacity of the HDV is too low (or inactive) in zebrafish embryos or because HDV is not functional when inserted in that particular location. Indeed, efficient ribozyme cleavage requires optimal divalent cation concentrations (preferably magnesium)

and precise pH conditions (Ke et al., 2004; Asif-Hullah et al., 2007; Nomura et al., 2013) to induce optimal folding and catalytic activities. As such, we hypothesized that HDV ribozyme inefficient activity in *malat1* could be due to low intracellular magnesium concentration in zebrafish embryos. Thus, we injected *malat1*^{HDV} embryos with increasing concentration of magnesium (10mM, 20mM, 50mM and 80mM) or incubated them in magnesium-enriched medium (50mM). Once again, we detected no *malat1* destabilisation effect upon ribozyme insertion (Figure 22D). By contrast, insertion of a SV40 polyA signal lead to a complete knock-out of *malat1* transcript in embryos and adult tissues (Figure 22E), even though zebrafish *malat1*^{SV40} KO do not present morphological, survival or fertility defects, in agreement with published mouse *malat1* KO (Eifmann et al., 2012). To conclude, our results show that SV40 polyA signal is more efficient than HDV ribozyme to destabilize lncRNAs at this particular locus.

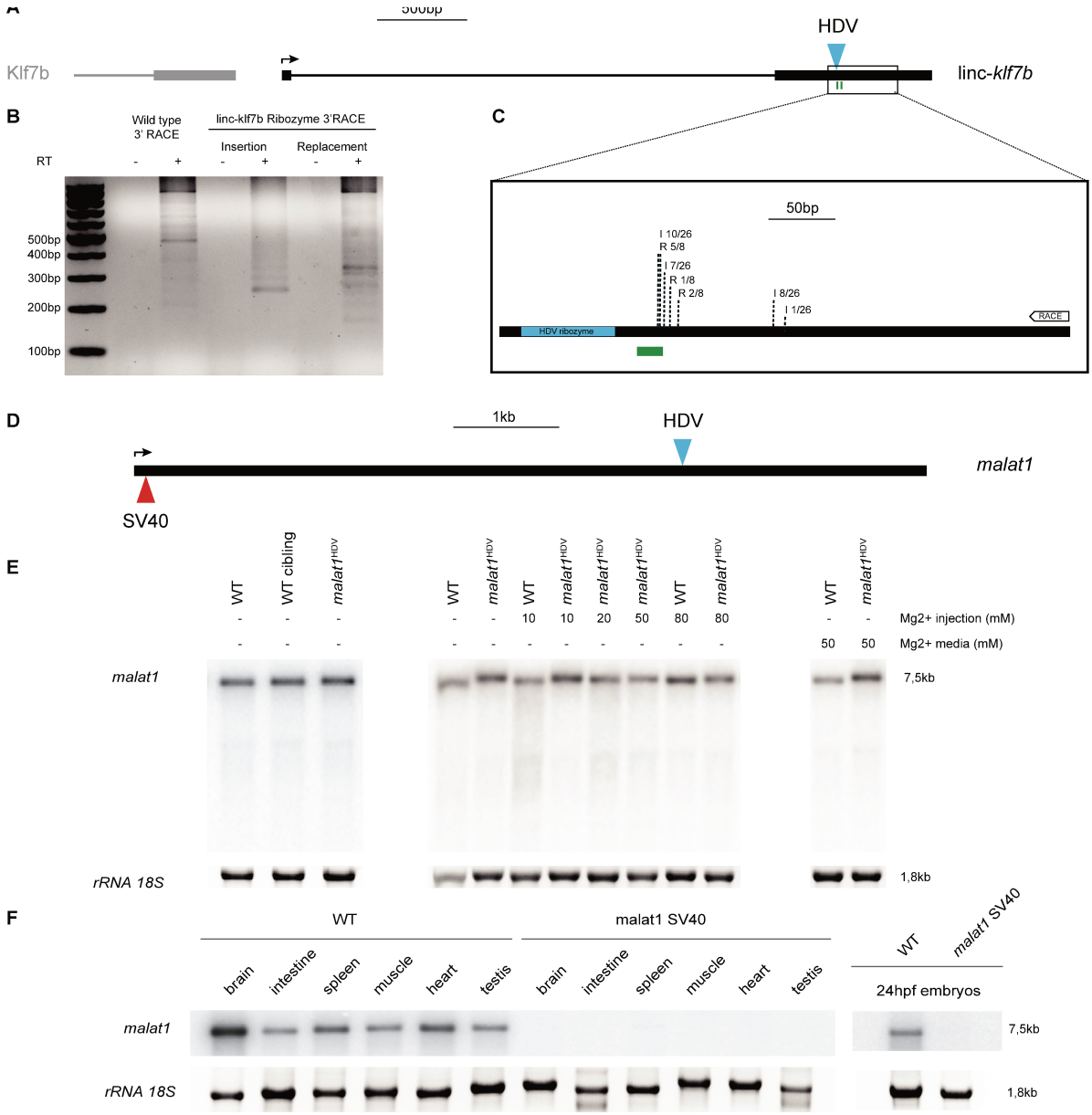


Figure 22: SV40 polyA signal inactivate lncRNA in a more efficient manner than HDV ribozyme

A Zebrafish linc-*klf7b* genomic locus with HDV ribozyme insertion position (gRNA in green). **B** Ethidium bromide gel of linc-*klf7b* HDV ribozyme 3' RACE cleavage assay performed on pools of injected embryos. **C** Zoom in on linc-*klf7b* HDV integration (blue box) position and annotation of RACE degradation products obtained with insertion (I) or replacement (R) strategies. **D** Zebrafish *malat1* genomic locus with SV40 and HDV ribozyme insertion position. **E** High molecular weight blot showing the levels of *malat1* in wild type and *malat1*^{HDV} 24hpf (hour post fertilization) embryos with/without yolk injection or incubation with magnesium cations. *malat1*^{HDV} cleavage/degradation products are expected to have a size between 2,5 and 5 kb. Each sample corresponds to a pull of 50 embryos. **F** High molecular weight RNA blot showing the levels of *malat1* in wild type and *malat1*^{SV40} adult organs and embryos (pull of 50). *Malat1*^{SV40} and *malat1*^{HDV} Northern blots were performed by

Angelo Bitetti (former PhD students in the Shkumatava lab) with a *malat1* probe of 266 nucleotides localised at the 3' end of the transcript. Gel blots and hybridization were performed in biological duplicates with 18s rRNA used as a loading reference.

DISCUSSION

I. Inactivating long noncoding RNA *in vivo*: finding the best strategy between transient and stable mutagenesis

Whereas protein-coding gene inactivation can usually be achieved through minimally-invasive methods affecting their translation, the disruption of lncRNAs requires more invasive knock-out approaches (Delás and Hannon, 2017; Marchese et al., 2017).

As transient mutagenesis strategies are by definition ephemeral, their analysis is temporally restricted to cell culture or to the first days of living embryos. Moreover, lncRNAs are often specific to a compartment of the cell, and cytoplasmic inactivating elements, such as siRNAs that have to interact with the RISC complex, may inefficiently target nuclear transcripts (Delás and Hannon, 2017).

Currently employed lncRNA stable mutagenesis approaches usually target either the DNA sequence or the transcription process (see Table 4) and present several caveats due to lncRNA locus complexity (Ziegler and Kretz, 2017). Indeed, DNA elements overlapping with lncRNA loci can themselves have important functions independent of the transcribed lncRNA molecules (Cho et al., 2018; Groff et al., 2016; Paralkar et al., 2016). Thus, large genomic deletions of lncRNA loci that inadvertantly impact the integrity of DNA structural elements confound the interpretation of any result, making it impossible to distinguish the phenotype resulting from disruption of the DNA regulatory elements from those resulting from lncRNA transcript disruption (Haemmerle and Gutschner, 2015; Marchese et al., 2017).

The advancement of CRISPR-Cas9 in zebrafish (Hwang et al., 2013) opened multiple possibilities to explore gene functionality at the scale of living organisms. We started the project at the onset of these rapid zebrafish genome editing times and targeted our candidate genes by genetic deletion. However, while creating our set of candidate mutants, we came across several challenges such as the absence of efficient gRNA for lnc-*zfp64* or lnc-*gosr2* (when co-injected with Cas9 mRNA) or the large genomic locus of lnc-*myca*, lnc-*mycb* and lnc-*sox4a/menhir*. We could not obtain homozygous genetic mutant for lnc-*klf7b*, TSS of lnc-*myca* and lnc-*mycb* (see Table 4), probably because our deletion affected functional DNA elements regulating adjacent coding genes, as reported for *PVT1* promoter region in human cell lines (Cho et al., 2018). Indeed, because lnc-*myc*/*PVT1* promoters are syntenic, it is possible that these *MYC* regulating DNA element are functionally preserved. As for lnc-

sox4a/menhir 3' deletion, its homozygous lethality could be due to DNA regulatory element off-targets, however the brain specific alternative TSS isoform observed in adults of lnc-*sox4a* TSS mutants (see Appendix) and its splicing structural similarities with human neuroblastoma *CASC15-S* (Russell et al., 2015) support RNA functionality.

Zebrafish mutants for lnc-*ppm1bb* did not show any obvious morphological defects in embryos and adult, despite the alteration of adjacent *ppm1bb*. Our results suggest that lnc-*ppm1bb* acts *in cis* on adjacent protein-coding genes, however genetic deletion did not allow us to distinguish between functionality of the lncRNA molecule and the DNA regulatory elements that could influence *ppm1bb* transcription levels. Therefore, it would be necessary to generate other stable genetic mutants, such as inserting a less-invasive premature polyA signal, to distinguish transcript functionality from that of DNA regulatory elements. The protein function of *ppm1bb* has not been characterized in zebrafish, but in standard laboratory conditions its misregulation does not appear to affect zebrafish physiology.

To further investigate lnc-*ppm1bb/ppm1bb* function *in vivo*, I would start by confirming lncRNA *cis* action in cell culture and characterize its molecular function. Next, to potentially uncover a function, I would investigate lncRNA and protein-coding gene function in a panel of stress conditions *in vivo*. As *Ppm1b* has been reported to regulate negatively necroptosis in mice by dephosphorylating RIP3(Chen et al., 2015), I would look at the impact of lnc-*ppm1bb*$^{-/-}$ on the necroptosis process using zebrafish larvae physiological retina cone cell death as a read-out (Viringipurampeer et al., 2014).

II. RNA Destabilising Element and premature polyA signal differential efficiency to inactivate lncRNA *in vivo*

We established a protocol to knock-in precisely small inserts (less than 140bp) in the zebrafish genome bypassing the Non Homologous End Joining to promote Homologous Recombination. Combining both CRISPR-Cas9 and morpholino technology (targeting xrcc4, a strategy first reported in Bertolini et al., 2006), Angelo Bitetti and I significantly improved the efficiency small sequence insertions. Up to now, our lab has successfully integrated polyA premature signals, ribozyme and also a tag sequence (not reported here) in an heritable way in the zebrafish genome. We also observed that insertion efficiency was variable according to the genomic locus, suggesting that chromatin marks or genomic structure may affect insertion rate. Properties of the desired insertion sequence may also impact its insertion efficiency or ability to be easily screened. Indeed, integration of HDV ribozyme sequences was particularly

difficult to genotype, as the high GC contents of the self-cleaving ribozyme and its hairpin secondary structures impair polymerase reading (Nelms et al., 2011; Won et al., 2017).

Recently, CRISPR-Cas9 technology has been greatly improved by the addition of an intermediate RNA molecule bridging the Cas9 protein and the guide RNA (AltR-CRISPR from IDT). Following the same strategy, bridging NHEJ disrupting or HR promoting elements to the guide RNA/Cas9 protein complex could improve knock-in efficiency.

Ribozyme are composed of different types of self-cleaving RNA sequences like the Hammerhead or the HDV ribozymes (Hamman et al., 2007) in native or inducible conformations (Nomura et al., 2013). Insertion of the HDV ribozyme minimal 72 nucleotides sequence did not lead to *malat1* down-regulationlikely either because the cleaving capacity of the HDV was too low (or inactive) in zebrafish or because the HDV ribozyme was not functional when inserted in that particular location in the zebrafish genome. There are multiple explanations for the lack of *malat1* down-regulation including HDV's requirement for high divalent cation concentrations for self cleavage and its dependence on precise tertiary structure formation, which could be disrupted by the sequences flanking the insertion site (Kapusta and Feschotte, 2014).

Hammerhead ribozyme has a reduced requirement for divalent cations (Fedoruck and Wyszomirska 2009) and a longer sequence than HDV potentially making it more stable and in zebrafish than the minimal HDV ribozyme used in this study. To screen for ribozyme cleavage capacity within different sequence contexts, it would be prudent to test ribozyme cleavage activity within its lncRNA insertion context *in vitro* or *ex vivo* prior to conducting *in vivo* investigations in order to identify efficiently cleaved sequences and to adapt the insertion site protocol accordingly. In addition, a quick screen to assess the cleavage capacity of different ribozyme subtypes or lengths in zebrafish organism could be conducted by expressing functionally reported ribozyme under the control of zebrafish endogenous promoter and inserting them in the zebrafish genome using Tol2 transposase.

Unlike the HDV ribozyme, insertion of a polyA premature signal lead to a complete knock-out of abundant mono-exonic *malat1* in zebrafish embryos and adult tissues (see Appendix). In this case, minimally invasive insertion of a single polyA signal nearby the gene TSS was sufficient to block transcription and inactivate a highly expressed long noncoding RNAs. However, it is highly probable that insertion of a premature polyA signal in lnc-*sox4a/menhir* locus would not prevent the emergence of brain specific alternative TSSs (see chapter 1). Thus, to access globally the utility of this method for lncRNA knock-out, one

should test the efficiency of a single premature polyA signal insertion on additional lncRNAs with varying expression levels and more complex exonic structures.

MATERIALS AND METHODS

I. Generation of CRISPR/Cas9 mediated zebrafish mutants

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 the Institut Curie (APAFiS#13883-201803021206135-v1) in compliance with the international guidelines. Zebrafish were staged using standard procedures (Kimmel et al., 2015). Concerning the animals induced with melanoma, 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 the Institut Curie CEEA-IC #118 (Authorization APAFiS#13883-201803021206135-v1 given by National Authority) in compliance with the international guidelines

a. Injection procedure

i. lncRNA deletion mutant

Syntenic lncRNA mutant were generated through partial locus deletion using CRISPR/Cas9-mediated genome editing technology. sgRNA couples (9ng of each, see Table 6) and Cas9 mRNA (150ng) were co-injected into one-cell stage AB zebrafish embryos (Hwang et al., 2013). sgRNAs and Cas9 mRNA were generated as described previously (Hwang et al., 2013) using the codon optimized plasmid JDS246 for the Cas9 mRNA synthesis (Addgene #43861). All of the RNAs were purified with the RNeasy Mini Kit (Qiagen).

ii. Insertion

Ribozyme and polyA insertion were generated through injection of sgRNA (50 to 100ng, alone or in couple), *Cas9* mRNA (200ng) or protein (50ng/µL kindly provided by the Concordet team- Muséum National d'Histoire Naturelle, Paris), and different concentrations of single-strand oligos (IDT, see Table 7). In order to improve Homologous Recombination (HR) efficiency, injection mix were supplemented with Non-Homologous-End-Joining (NHEJ) targeting morpholinos: MO-xrcc4 and/or MO-xrcc5 (Gene Tools LLC, see Table 8) and/or Rad51 mRNA (75 to 150 ng/µL, amplified from zebrafish cDNA and in vitro

transcribed with mMESSAGE mMACHINE T7 ultra kit according to manufacturers conditions, see Table 9).

b. Screening procedure

i. lncRNA deletion mutant

Genomic DNA was extracted through tissue (embryos or fish fins), digested with 2µg/µL of proteinase K (Roche) at 55°c for 2-3h. DNA was extracted with phenol:chloroform. Fish were genotyped by PCR (see primers list in Table 10), sequencing and mapping of genetic amplification product.

ii. Insertion

To screen for best insertion conditions, genomic DNA was extracted from 16 embryos in each injection conditions in 1xTE and 2µg/µL proteinase K (Roche) at 55°c for 3 hours then 94°c for 10 minutes. Insertion efficiency was then determined by PCR (flanking and insertion) on single embryos (see Table 10). Founders and next generations were screened through genomic DNA extraction and PCR as described for lncRNA deletion.

c. Statistics

Partial penetrant phenotype observed in double linc-*myc* mutant was analysed using Chi-square statistical test.

II. Reverse transcription and real-time PCR

a. Zebrafish WT and mutant

Total RNA was isolated from zebrafish embryos and tissues using TRIzol (Invitrogen) according to manufacturer's instructions, followed by DNAse treatment (TURBO DNA-free Ambion) and ethanol precipitated. cDNA was obtained by retro-transcription of 1µg total RNA with SuperScript IV reverse transcriptase using oligo-dT and Random Hexamer (Invitrogen), then amplified with TaqMan Universal PCR master Mix for linc-*klf7b* and PowerUp SYBR green qPCR master mix (ThermoFisher Scientific) for other lncRNA. Primers are listed in Table 11.

b. Human tissue expression

All human expression datas were obtained through EMBL-EBI Expression Atlas provided by the Genotype-Tissue Expression (GTEx) project. Primers are listed in Table 11.

c. Melanoma tissue

Total RNA was extracted from dissected tumor and brain of fish injected with melanoma inducing construct using TRIzol (Invitrogen) according to manufacturer's instruction, treated with DNAse (TURBO DNA-free Ambion) and ethanol precipitated. cDNA was obtained by retro-transcription of 300ng total RNA with SuperScript IV reverse transcriptase using oligo-dT and Random Hexamer (Invitrogen), then amplified using PowerUp SYBR green qPCR master mix (ThermoFisher Scientific). Primers for human *NRAS*, *hCASC15* and melanoma proliferation/invasion markers are listed in Table 11.

III. RNA ligase-mediated and oligo-capping Rapid Amplification of cDNA Ends

a. menhir TSS

Alternatives Transcriptions Start Sites (TSS) specific to linc-sox4a deletion brain tissue were determined by Rapid Amplification of cDNA Ends (RACE) according to manufacturer's instruction (GeneRacer^{TM} kit, life technologies). Gene Specific Primer (GSPs) were listed in Table 12.

b. Ribozyme cleavage assay

linc-*klf7b* cleavage by ribozyme was observed on pool of injected embryos by RACE. Total RNA of a pool of 50 injected or WT embryos was extracted with TRIzol (Invitrogen) according to manufacturer's instruction and purified with QIAGEN RNeasy cleaning kit. 5µg of RNA was ligated according to GeneRacer's (Invitrogen) instructions. GSPs are listed in Table 12.

IV. Melanoma induction in zebrafish

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 the Institut Curie CEEA-IC #118 (Authorization APAFiS#13883-201803021206135-v1 given by National Authority) in compliance with the international guidelines

a. Melanoma inducing Construct

PdestTol2 mitfa::NRAS^{G12D} and pdestTol2 mitfa ::NRAS^{Q61R} were kindly provided by Adam Hurlstone from Manchester University. PdestTol2 mitfa::NRAS^{G12D} construct was modified by the insertion of the mitfa::*hCASC15*, mitfa::*hCASC15* truncated, mitfa::*linc-sox4a* and mitfa::*venus* with Gibson assembly cloning using Gibson Assembly Mastermix (New England Biolab). To ensure correct subcellular localization of rescue constructs, I have inserted a zebrafish *libra* intronic sequence at cDNA exon/exon junction (Bitetti et al., 2018; Rasmussen and O'Carroll, 2011).

b. Injection procedure

Wild type and lncRNA mutant embryos were coinjected at one cell stage with 50ng/µL *Tol2* mRNA (in vitro transcribed with mMESSENGER mMACHINE T7 ultra kit Ambion) and melanoma induction construct (25ng/µL for mitfa::*linc-sox4a* rescue construct, 50 ng/µL for the others). A minimum of 2 independent injections were performed in order to generate a population of fish presenting melanoma in each conditions.

c. Melanoma monitoring procedure

i. Melanoma induced zebrafish through injection

Injected embryos were selected for cyaa::*venus* expression at 72hpf. Zebrafish larvaes were bred in standard conditions (28-29°c under a 14h light : 10h dark cycle in a filtered freshwater recirculation system). Starting from 5 to 21 weeks post-fertilisation, melanoma induced zebrafish were individually observed each week and classified according to their melanoma progression stage (No Lesion; Few melanocytic Lesion; Radial Growth Progression (RGP);

Vertical Growth Progression (VGP); VGP with tumor/nodule/hyperplasia) (Michailidou et al., 2009). All animals presenting tumor, nodule or hyperplasia covering more than 10% body surface, epidermic lesion, weight loss and/or feeding/swimming troubles were sacrified immediatly.

d. Statistics

Statistical tests were performed using PRISM software. Tumor apparition in time and survival was analysed with Mantel-Cox test for survival curves. Proportion of zebrafish at different melanoma stages and with tumors at a precise timepoint were analysed with unpaired t-test. Rare internal tumor event was analysed using hypergeometric statistical test.

V. Human melanoma cells xenotransplantation

a. Cell lines

Three different human melanoma cell lines were used in this project : SK-MEL-2 (kindly provided by Stephan Vagner, Institut Curie, Orsay), 501Mel and WN266-4 GFP (kindly provided by Adam Hurlstone, University of Manchester). Cells were cultured in DMEM (Gibco) supplemented with 10% FBS (Gibco) for 501Mel and WN266-4 and MEM (Ambion) supplemented with 10% FBS (Gibco) and 1mM L-Glutamine (Gibco) for SK-MEL-2.
501Mel and SK-MEL-2 were infected with lentiviral particles (kindly provided by Giulia Fornabaio, Institut Curie, Paris), expressing the Green Fluorescent Protein (GFP) and a Puromycin-resistance cassette. After the infection, cells were selected for 3 days in complete culture medium containing 0,75µg/mL for 501Mel and 1µg/mL for SK-MEL-2.
For graft experiments, cells were resuspended in culture medium at a concentration of 250 000 cells per µL.

b. Fish and graft

Tg(kdlr::*mcherry*) zebrafish (kindly provided by Filippo Del Bene, Institut Curie, Paris) were crossed with *menhir* mutant. Wild type and *menhir* mutant 48hpf embryos were manually dechorionated, anesthetised with 0.004% tricaïne and injected in the yolk with approximatively one nanoliter of human melanoma cell suspension (around 250 cells).

Embryos were maintained in egg water supplemented with PenStrep (25 U/mL final, Life Technology) and PTU (3µg/mL) at 34°c for 3-4 days.

c. Imaging

Starting from 72 hours post-injections, percentage of larvae presenting melanoma cells migrated oustide of injection point and average number of invading cells were estimated using Zeiss optical microscope (steREO discovery v20) combined with Zen software 4 days post-injections. For confocal microscopy, larvae were mounted in 1,2% UltraPure low-melting agarose (Invitrogen) supplemented with 0.04% tricaïne. Larvae were images by a Zeiss LSM 700 confocal microscope with a 25x oil immersion objective.

d. Image analysis and statistics

All images were analysed using Zeiss and Fiji software. Cell speed was obtained through ImageJ cell tracker software. Statistical tests Mann and Whitney were performed with PRISM software.

VI. In situ Hybridisation of the zebrafish ovary

Adult wild type zebrafish ovaries were fixed in 4% PFA (Electron Microscopy Science) for 24h at 4°c, dehydrated and embedded in paraffin blocks. Embedding and 4um sections were performed by Emerald Perlas EMBL monterotondo).
To generate probes for in situ hybridisation, *menhir* full length was amplified by PCR using Phusion DNA polymerase (Thermo Scientific) using the primers listed in Table 9. PCR product was subcloned into a pGEM-T Vector system (Promega) and confirmed by sequencing. DIG-labeled RNA probes were generated by vector linearization and T7 in vitro transcription with DIG-RNA labelling kit (Roche).
Zebrafish ovary 4um section were deparaffined and rehydrated, fixed in 4% PFA, digested with proteinase K, acetylated and hybridized with the corresponding probe in 50% formamide, 5x SSC, 5X Denhardt's solution, 500ug/ml salmon sperm DNA and 250ug/ml tRNA overnight at 56°c. Post hybridization washes were performed in 50% formamide and 2X SSC at 56°c, then 2X SSC at ambient temperature. The sections were blocked in 10%

sheep serum and incubated overnight with anti-DIG-AP antibody (Roche at 1:1000) at 37°c. Signal detection was done using NBT/BCIP substrate (Roche)

VII. Histology

Entire adult zebrafish were fixed in 4% PFA (Electron Microscopy Science) for 3 days at 4°c and decalcified in 0,25mM EDTA for 3 more days at 4°c. Zebrafish were then, sectioned and disposed in a coronal or longitudinal orientation in histology cassettes and stored in Ethanol 70%. Paraffin embedding, 4um sections and staining (Hematoxylin-Eosin-Safran) were performed by the Histim platform (Institut Cochin, Paris).

VIII. GTEx, PCAWG and TCGA investigation

Long noncoding RNA and protein-coding genes expression in healthy and tumor tissue were respectively obtained from the Genotype Tissue Expression (GTEx) and the Pan Cancer Analysis of Whole Genome (PCAWG) project available on EBI Expression Atlas.
Long noncoding RNA expression analyses according to melanoma mutation status and severity stage were performed by Nicolas Servant (Institut Curie, Paris) from data collected by The Cancer Genome Analysis project (TCGA). Statistics were analysed with Wilcoxon test.

Table 6: Single guide RNAs (sgRNA) sequences

sgRNA	**Sequence (5'-3')**
sgRNA zf lnc-*klf7b*^{del}	GGGCCGTAGAAGGCAGTGAA
sgRNA zf lnc-*klf7b*$^{del\ A31}$	GGAGTGATGAGCCAGTCATC
sgRNA zf lnc-*klf7b*$^{del\ A33}$	GTGAGACAGAGGGTTAGCCA
sgRNA zf lnc-*ppm1bb*$^{3'\ del}$	GGTGTAGTTGGGTTTAGAGA
sgRNA zf lnc-*ppm1bb*$^{3'\ del}$	GTTATTTTGGTTACCCAGGC
sgRNA zf lnc-*myca*$^{3'\ del}$	GGACCAGCTGTCAGCTTGAG
sgRNA zf lnc-*myca*$^{3'\ del}$	GCGGTATGGAAGTCTGTGTG
sgRNA zf lnc-*mycb*$^{3'\ del}$	GCTACACAGTTAAGGGGTGG
sgRNA zf lnc-*mycb*$^{3'\ del}$	GATTAGCCAGTTGTCCACTG
sgRNA zf *menhir*$^{5'del}$	GCATGATATCGGACAAGGGG
sgRNA zf *menhir*$^{5'\ del}$	GACATCCTGACGTAGGTAAA
sgRNA zf *malat1*^{HDV}	GGGAATGTGTGGGGCTTTCT
sgRNA zf *malat1*^{SV40}	GGTGAGGCGCTATGGAAGGC
sgRNA zf *megamind* ^{HDV}	GGAGCGAGAGGAGTCCATAT

zf: zebrafish – del: deletion - HDV: hepatitis delta virus

Table 7: Single strand DNA (ssDNA) oligos sequences

ssDNA	**Sequence (5'-3')**
ssDNA insertion zf lnc-*klf7b*^{HDV} short	AGCTGGGCAGTGTAAGCCCTGATGACTGGCGCCGGCATGGTCCC AGCCTCCTCGCTGGCGGCCGGTGGGCAACATGCTTCGGCATGGT GAATGGGACCTCTAGGTACCTCATCACTCCCTCCCATGGCTAACC CTCTG
ssDNA insertion zf lnc-*klf7b*^{HDV} long	CAGGTGTGAGTCTGCCAGGCTCGGTGGCACAGCTGGGCAGTGTA AGCCCTGATGACTGGCGCCGGCATGGTCCCAGCCTCCTCGCTGGC GGCCGGTGGGCAACATGCTTCGGCATGGTGAATGGGACCTCTAG GTACCTCATCACTCCCTCCCATGGCTAACCCTCTGTCTCACAGAG TCTGAAACAGCTGGTTCCCA
ssDNA replacement zf linc-*klf7b*^{HDV} short	TTAAGGGCAGGTGTGAGTCTGCCAGGCTCGGCCGGCATGGTCCC AGCCTCCTCGCTGGCGGCCGGTGGGCAACATGCTTCGGCATGGT GAATGGGACCTCTAGGTACCCTGAAACAGCTGGTTCCCACTGAG CCTAGC
ssDNA replacement zf lnc-*klf7b*^{HDV} long	GGTGTGTGCTTGGTGTTATTGCACAGCGGGTTAAGGGCAGGTGT GAGTCTGCCAGGCTCGGCCGGCATGGTCCCAGCCTCCTCGCTGGC GGCCGGTGGGCAACATGCTTCGGCATGGTGAATGGGACCTCTAG GTACCCTGAAACAGCTGGTTCCCACTGAGCCTAGCCCTGACCAG GTCAAACCTTGCACCTTCCAA
ssDNA zf *malat1*^{HDV}	GGTATGCATTTGTTTCTTGGGGGAATGTGTGCCGGCATGGTCCCA GCCTCCTCGCTGGCGGCCGGTGGGCAACATGCTTCGGCATGGTG AATGGGACCTCTAGGGGCTTTCTTGGGTGTTTCTTTATGCTTT
ssDNA zf *malat1*^{SV40}	AACATTGTGCGTCACGACGGGGTGAGGCGCACTTGTTTATTGCA GCTTAAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACA AATAAAGATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACT CATCAATGTATCTTATCATGTCTGTTATGGAAGGCAGGGAGGCTT CGTTGATCTG
ssDNA zf *megamind* ^{HDV}	CGGGCGCAAGAGGGTAAAGAGGAGCGAGAGGCCGGCATGGTCC CAGCCTCCTCGCTGGCGGCCGGTGGGCAACATGCTTCGGCATGG TGAATGGGACCTCTAGAGTCCATATTGGCCCTTCTGGGGATCATA

ssDNA: single stranded DNA – zf: zebrafish – HDV: Hepatitis Delta Virus

Table 8: Morpholinos (MOs)

Translation inhibiting MOs	**Optimal concentration**	**Sequence (5'-3')**
zf *xrcc4*	3ng	CACTACTGCTGCGACACCTCATTCC
zf *xrcc5*	3ng	CCTTATTCTCAGCAAACACCTGCAT

zf: zebrafish

Table 9: Full length RNA primer

Targeted gene	**Forward Primer Sequence (5'-3')**	**Reverse Primer Sequence (5'-3')**
zf lnc-*klf7b*	TGTGCAGAAAATAGTGTTGAATGTTGGCA	TGCAGATTGTAACGTGAGGTCTCCCTG
zf lnc-*ppm1bb*	ACCGTCACGAAGCTTGAACTGTAGTTTG	CCCCTCTCTCTGCGCTCAATAGTTTCTT
zf lnc-*myca*	CACATGGAAGAGGGCTCGAGACGAA	TGAGGCTTTTTGCCAACACTTCCCC
zf lnc-*mycb*	TGCTGTGAGCACATGGAGGAGAGAG	TTGGCAGTTTTGGTGTCACAGTTTG
zf *menhir*	CCACACGTGTTTGACAAAACTCTCCGA	TGCATAGGAATAACCTCCGAACTCAAAAC
zf *rad51*	CTCTAGCAGCAGCAGCAGTAGCTTA	TCCTGAAGTCTTGAAATGACAATAAAATTCCAGT
h *PVT1-004*	TGTGCCTGTCAGCTGCATGGAG	AGCTCATAGGTTAGGGATTTTTCAAAGGC
h *PVT1-005*	GAGGGGCGACGACGAGCTG	TTCACCAGGAAGAGTCGGGGTCTTA
h *CASC15*	GGATTACAATTTGATCGCTGGGAAT	CTCTGACTTACTCTCTGTTTCTGTCA

zf: zebrafish - m: mouse - h: human

Table 10: Mutation screening primer

Deletion screening primers	**Forward Primer sequence (5'-3')**	**Reverse Primer Sequence (5'-3')**
zf lnc-*klf7b*^{del}	GATCTCACTGGCGTCCAAGT	AGTTCCTTCTGCTCAGACACAA
zf lnc-*ppm1bb*$^{3'del}$	CCTTCTGGAGAAAGTCTGATCTG	TTCACGAATCCACCAGAGGC
zf lnc-*myca*$^{3'del}$	GTGTCATGATGATGCAGCGA	TCACGTCATTAGCAGCACGT
zf lnc-*mycb*$^{3'del}$	GCCGTTTAACCCCACAGTGA	ACCACTGTTTCCTCGTGCTC
zf *menhir*$^{5'del}$	TAAGGAGCACAAATGTCTTAATACCTCAGG	GTTTTCTCTATATGCCGACTGTTTTGATCC
Insertion screening primers	**Forward Primer Sequence (5'-3')**	**Reverse Primer Sequence (5'-3')**
zf lnc-*klf7b*^{HDV} ins	GGCATGGTGAATGGGACCTCTAGGT	CATAACAAATGGGTCCGCCGTTGAC
zf lnc-*klf7b*^{HDV} flank	ACAGCTGGGCAGTGTAAGC	CTGGTCAGGGCTAGGCTCAGT
zf *malat1*^{HDV} ins	TTCGGCATGGTGAATGGGACCTCTA	GTAGGACTGGACTGATCTGCC
zf *malat1*^{HDV} flank	GTGTACATTCTGTAACTGTTCCC	GTAGGACTGGACTGATCTGCCTCTCCACC
zf *malat1*^{SV40} ins	TCACTGCATTCTAGTTGTGGTTTGTCC	GCTTGTATTTTATCTTCGTCACGCTTGC
zf *malat1*^{SV40} flank	GTGTGGTATGTTGTGTCAAG	CCGCCATTTTGTAAATTCTTTCTAGCGTCGAG
zf *megamind*^{HDV} ins	ACAGCAAGTGAGATGGCTTAAATGGA	GTCCCATTCACCATGCCGAAG
zf *megamind*^{HDV} flank	GGAGGTTCGGGCGCAAGAGGGTAAAGAGG	CAGCCAGGCAGTGCTCGATGTAATTTCAGTTGC

del: deletion – HDV: hepatitis delta virus- ins: insertion- flank: flanking

Table 11: Real-Time PCR primer

Target gene	**Forward Primer Sequence (5'-3')**	**Reverse Primer Sequence (5'-3')**
zf lnc-*klf7b*	Taqman probe	
zf lnc-*ppm1bb* 3'	GAAACTCAGGCAGGCTTGTGA	CTGTTCTGTGTCACAAGCGATG
zf lnc-*ppm1bb* 5'	CGTCACGAAGCTTGAACTGTAG	GTCATTAAAAACACGTCGCCCC
zf lnc-*myca* 5'	CAAAGAATAACTCCACACCTGAAAG	GCAGTATTGTTGACTTTGATGGCA
zf lnc-*myca* 3'	GCTTTGCTGCAAGGAGATGAAG	CTTGGTGCTGAAATCTGTCCTCTC
zf lnc-*mycb* 5'	GAGGAGAGAGGACGTTCACTTAG	CTCTCCCTGAAGTGGTTATGAG
zf lnc-*mycb* 3'	CAGATAACCGTACCCTCTGTTGG	GGCCTTCTTTAGAGCACTGGTAC
zf *menhir* 5'	CTCTCCGAAGATGACATCCTGAC	CACAAGCCTGGTTAAAGAGAGG
zf *menhir* 3'	CATCACTCACAGTTCAGCTCTCC	GAACACGACTATCCTCCACACTC
zf *klf7b*	CGACACGGGCTACTTCTCAG	CCTTTTGGGCTCTGTCTGCA
zf *ppm1bb*	GTTTCCCCAATGCACCAAAGG	TGGCTTAGATCTGGGATGCCT
zf *myca*	CATGCTGGTCCTGGACACTC	CCTCCTCTTCTTCCTCCTCCTC
zf *mycb*	CTCACGCTGACATCTGACCAT	AGGGCTGGTAGGAGTCGTAG
zf *sox4a*	TTCAACTTTGAGTCGGGCTCC	TTCGAGCCAGTCCCCTGAT
zf *sox4b*	GTATGTTTGTATAATACGCCGCCC	CACAGTTTTCCAGTTCAGAGCG
zf *tgfb1*	GTACTTCACCAACTGCAAGCAATG	CTCTCGGAGTACATCTTAGTTGTGG
zf *βactin*	CGAGCTGTCTTCCCATCCA	TCACCAACGTAGCTGTCTTTCTG
zf *eif1α*	CTGGAGGCCAGCTCAAACAT	ATCAAGAAGAGTAGTACCGCTAGCATTAC
zf *rpil3*	TCTGGAGGACTGTAAGAGGTATGC	AGACGCACAATCTTGAGAGCAG
h *CASC15* 1	CCAAAATCAAAAGTATGGGCAGGC	CAGTTTTGTGGCAGGTAGGGG
h *CASC15* 2	CTGCGAGAGTTGTGAAAAATTGGG	CCTCAACTGCAAAGTCCTCTTAG
h *NRAS* 1	TTACGCTAGCCTCCCGAGATCC	CAAACAGGTTTCACCATCTATAACC
h *NRAS* 2	CTCCCGAGATCCATGACTGAGTAC	CAAACAGGTTTCACCATCTATAACC
h actin	GACATGGTGTATCTCTGCCTTACAG	CTTTGCGGATGTCCACGTCAC

zf: zebrafish - m: mouse - h: human

Table 12: Gene Specific Primers (GSP) for RACE

GSPs	**Sequence (5'-3')**
zf *menhir* $^{5'RACE}$ 1	ATCAGCCTTAGGTTACAGGAAGAGAGCC
zf *menhir* $^{5'RACE}$ 4	CCAGAACACGACTATCCTCCACACTCGG
zf *menhir* $^{5'RACE}$ 5	TTCCCACGCTGAAGGCTGATACTGAGAG
zf lnc-*klf7b* $^{3'cleavage}$ 1	GCATAACAAATGGGTCCGCCGTTGACGT
zf lnc-*klf7b* $^{3'cleavage}$ 1	CGATTTAGCCCCAAAGGTAAATGTGTGC
zf lnc-*klf7b* $^{3'cleavage}$ 1	TGTGTGCGCCAGAAACTGCAGGGCACCC

zf: zebrafish

BIBLIOGRAPHY

Aftab, M.N., Dinger, M.E., and Perera, R.J. (2014). The role of microRNAs and long non-coding RNAs in the pathology, diagnosis, and management of melanoma. Archives of Biochemistry and Biophysics *563*, 60–70.

Akbani, R., Akdemir, K.C., Aksoy, B.A., Albert, M., Ally, A., Amin, S.B., Arachchi, H., Arora, A., Auman, J.T., Ayala, B., et al. (2015). Genomic Classification of Cutaneous Melanoma. Cell *161*, 1681–1696.

Akhter, A., Kumagai, R.-I., Roy, S.R., Ii, S., Tokumoto, M., Hossain, B., Wang, J., Klangnurak, W., Miyazaki, T., and Tokumoto, T. (2016). Generation of Transparent Zebrafish with Fluorescent Ovaries: A Living Visible Model for Reproductive Biology. Zebrafish *13*, 155–160.

Alam, T., Medvedeva, Y.A., Jia, H., Brown, J.B., Lipovich, L., and Bajic, V.B. (2014). Promoter Analysis Reveals Globally Differential Regulation of Human Long Non-Coding RNA and Protein-Coding Genes. PLoS ONE *9*, e109443–11.

Albadri, S., Del Bene, F., and Revenu, C. (2017). Genome editing using CRISPR/Cas9-based knock-in approaches in zebrafish. Methods *121-122*, 77–85.

Albino A.P., Le Strange R., Oliff A.I., Furth M.E. and Old L.J. (1984) Transforming *ras* genes from human melanoma: a manifestation of tumour heterogeneity? Nature Mar 1-7 ; 308 (5954) :69-72

Alvarez, M.L., and DiStefano, J.K. (2011). Functional Characterization of the Plasmacytoma Variant Translocation 1 Gene (PVT1) in Diabetic Nephropathy. PLoS ONE *6*, e18671–e18678.

Amaral, P.P., and Mattick, J.S. (2008). Noncoding RNA in development. Mamm Genome *19*, 454–492.

Amaral, P.P., Leonardi, T., Han, N., Viré, E., Gascoigne, D.K., Arias-Carrasco, R., Büscher, M., Pandolfini, L., Zhang, A., Pluchino, S., et al. (2018). Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. 1–21.

Amatruda, J.F., and Patton, E.E. (2008). Genetic Models of Cancer in Zebrafish (Elsevier Inc.).

Amândio, A.R., Necsulea, A., Joye, E., Mascrez, B., and Duboule, D. (2016). Hotair Is Dispensible for Mouse Development. PLoS Genet *12*, e1006232–27.

Amsterdam A., Burgess S., Golling G., Chen., Sun., Towns K., Farrington S., Haldi M., and Hopkins N., (1999). A large-scale insertional mutagenesis screen in zebrafish. Gen Dev 13:2713-2724

Ana Rita Amândio, A.N.E.J.B.M.D.D. (2016). Hotair Is Dispensible for Mouse Development. 1–27.

Arun, G., Diermeier, S., Akerman, M., Chang, K.-C., Wilkinson, J.E., Hearn, S., Kim, Y., MacLeod, A.R., Krainer, A.R., Norton, L., et al. (2016). Differentiation of mammary tumors and reduction in metastasis upon Malat1lncRNA loss. Genes Dev. *30*, 34–51.

Asif-Ullah M., Lévesque M., Robichaud G., and Perreault J.P. (2007). Not For Distribution. Curr. Gen. Ther. 7,205-216

Astone, M., Dankert, E.N., Alam, S.K., and Hoeppner, L.H. (2017). Fishing for cures: The alLURE of using zebrafish to develop precision oncology therapies. Npj Precision Onc *1*, 1–29.

Auer, T.O., Duroure, K., De Cian, A., Concordet, J.P., and Del Bene, F. (2014). Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. Genome Research *24*, 142–153.

Barriuso, J., Nagaraju, R., and Hurlstone, A. (2015). Zebrafish: A New Companion for Translational Research in Oncology. Clinical Cancer Research *21*, 969–975.

Bassett, A.R., Akhtar, A., Barlow, D.P., Bird, A.P., Brockdorff, N., Duboule, D., Ephrussi, A., Ferguson-Smith,

A.C., Gingeras, T.R., Haerty, W., et al. (2014). Author response. eLife *3*, e1002248–14.

Beilstein, K., Wittmann, A., Grez, M., and Suess, B. (2014). Conditional Control of Mammalian Gene Expression by Tetracycline-Dependent Hammerhead Ribozymes. ACS Synth. Biol. *4*, 526–534.

BERTOLINI, L., BERTOLINI, M., ANDERSON, G., MAGA, E., MADDEN, K., and MURRAY, J. (2007). Transient depletion of Ku70 and Xrcc4 by RNAi as a means to manipulate the non-homologous end-joining pathway. Journal of Biotechnology *128*, 246–257.

Berghmans, S., Murphey, R.D., Wienholds, E., Neuberg, D., Kutok, J.L., Fletcher, C.D., Morris, J.P., Liu, T.X., Schulte-Merker, S., and Kanki, J.P. (2005) tp53 mutant zebrafish develop malignant peripehral nerve sheath tumors. Proc Natl. Acad. Sci. USA 11, 407-412

Bhan A., Soleimani M., and Mangal S.S. (2017). Long Noncoding RNA and Cancer:A New Paradigm. Cancer Research AACR.

Bhattaram, P., ndez, A.P.-M.E., Sock, E., Colmenares, C., Kaneko, K.J., Vassilev, A., DePamphilis, M.L., Wegner, M., and Lefebvre, V.E.R. (2010). Organogenesis relies on SoxC transcription factors for the survival of neural and mesenchymal progenitors. Nature Communications *1*, 1–12.

Bitetti, A., Mallory, A.C., Golini, E., Carrieri, C., Gutiérrez, H.C., 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. Nature Structural &Amp; Molecular Biology 1–14.

Branzei, D., and Foiani, M. (2008). Regulation of DNA repair throughout the cell cycle. Nat Rev Mol Cell Biol *9*, 297–308.

Brockdorff N., Ashworth A., Kay G.F., Cooper P., Smith S., McCabe V.M., Norris D.P., Penny G.D., Patel D., and Rastant S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Letters to Nature v351

Ceol, C.J., Houvras, Y., Jane-Valbuena, J., Bilodeau, S., Orlando, D.A., Battisti, V., Fritsch, L., Lin, W.M., Hollmann, T.J., Ferré, F., et al. (2011). The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. Nature *471*, 513–517.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A Long Noncoding RNA Controls Muscle Differentiation by Functioning as a Competing Endogenous RNA. Cell *147*, 358–369.

Chapman, A., del Ama, L.F., Ferguson, J., Kamarashev, J., Wellbrock, C., and Hurlstone, A. (2014). Heterogeneous Tumor Subpopulations Cooperate to Drive Invasion. CellReports *8*, 688–695.

CHEN, L., YANG, H., XIAO, Y., TANG, X., LI, Y., HAN, Q., FU, J., YANG, Y., and ZHU, Y. (2016). Lentiviral-mediated overexpression of long non-coding RNA GAS5 reduces invasion by mediating MMP2 expression and activity in human melanoma cells. International Journal of Oncology *48*, 1509–1518.

Chen, W., Wu, J., Li, L., Zhang, Z., Ren, J., Liang, Y., Chen, F., Yang, C., Zhou, Z., Sean Su, S., et al. (2015). Ppm1b negatively regulates necroptosis through dephosphorylating Rip3. Nat Cell Biol *17*, 434–444.

Chen, X., Gao, G., Liu, S., Yu, L., Yan, D., Yao, X., Sun, W., Han, D., and Dong, H. (2017). Research Article. Biomed Res Int 1–9.

Chen, Z.-Z., Huang, L., Wu, Y.-H., Zhai, W.-J., Zhu, P.-P., and Gao, Y.-F. (2016). LncSox4 promotes the self-renewal of liver tumour-initiating cells through Stat3-mediated Sox4 expression. Nature Communications *7*, 12598–13.

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., et al. (2018). Promoter of lncRNA Gene PVT1 Is a Tumor- Suppressor DNA Boundary Element. Cell 1–38.

Chodroff, R.A., Goodstadt, L., Sirey, T.M., Oliver, P.L., Davies, K.E., Green, E.D., Molnár, Z., and Ponting,

C.P. (2010). Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. Genome Biol *11*, R72–16.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kühn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol *33*, 543–548.

Chudnovsky, Y., Adams, A.E., Robbins, P.B., Lin, Q., and Khavari, P.A. (2005). Use of human tissue to assess the oncogenic activity of melanoma-associated mutations. Nat Genet *37*, 745–749.

Cizelsky, W., Hempel, A., Metzig, M., Tao, S., Hollemann, T., Kühl, M., and Kühl, S.J. (2013). sox4 And sox11 Function during Xenopus laevis Eye Development. PLoS ONE *8*, e69372–12.

Colombo, T., Farina, L., Macino, G., and Paci, P. (2015). PVT1: A Rising Star among Oncogenic Long Noncoding RNAs. Biomed Res Int *2015*, –10.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science *339*, 819–823.

Corallo, D., Candiani, S., Ori, M., Aveic, S., and Tonini, G.P. (2016). The zebrafish as a model for studying neuroblastoma. Cancer Cell International 1–9.

Cunnington, M.S., Santibanez Koref, M., Mayosi, B.M., Burn, J., and Keavney, B. (2010). Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression. PLoS Genet *6*, e1000899–17.

Dang, C.V. (2012). MYC on the Path to Cancer. Cell *149*, 22–35.

Davis, M.P., Carrieri, C., Saini, H.K., van Dongen, S., Leonardi, T., Bussotti, G., Monahan, J.M., Auchynnikava, T., Bitetti, A., Rappsilber, J., et al. (2017). Transposon-driven transcription is a conserved feature of vertebrate spermatogenesis and transcript evolution. EMBO Reports *18*, 1231–1247.

Delás, M.J., and Hannon, G.J. (2017). lncRNAs in development and disease: from functions to mechanisms. Open Biol. *7*, 170121–10.

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 Research *22*, 1775–1789.

Ding, X., Wang, X., Lin, M., Xing, Y., Ge, S., Jia, R., Zhang, H., Fan, X., and Li, J. (2016). PAUPARlncRNA suppresses tumourigenesis by H3K4 demethylation in uveal melanoma. FEBS Letters *590*, 1729–1738.

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. Molecular Cell *70*, 854–867.e859.

Dovey M., White R.M., and Zon L.I. (2009). Oncogenic NRAS Cooperates with p53 Loss to Generate Melanoma in Zebrafish. Zebrafish (4):397-404.

Du, Z., Fei, T., Verhaak, R.G.W., Su, Z., Zhang, Y., Brown, M., Chen, Y., and Liu, X.S. (2013). Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. Nat Struct Mol Biol *20*, 908–913.

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 MALAT1is compatible with life and development. RNA Biology *9*, 1076–1087.

Engreitz, J.M., Ollikainen, N., and Guttman, M. (2016). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nature Publishing Group *17*, 756–770.

Etchin, J., Kanki, J.P., and Look, A.T. (2011). Zebrafish as a Model for the Study of Human Cancer (Elsevier Inc.).

Faghihi, M.A., Modarresi, F., Khalil, A.M., Wood, D.E., Sahagan, B.G., Morgan, T.E., Finch, C.E., St Laurent, G., III, Kenny, P.J., and Wahlestedt, C. (2008). Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. Nat Med *14*, 723–730.

Fedor, M.J., and Williamson, J.R. (2005). The catalytic diversity of RNAs. Nat Rev Mol Cell Biol *6*, 399–412.

Fedorenko, I.V., Gibney, G.T., and Smalley, K.S.M. (2012). NRAS mutant melanoma: biological behavior and future strategies for therapeutic management. Oncogene *32*, 3009–3018.

Fedoruk-Wyszomirska, A., Szymań ski, M., Wyszko, E., Barciszewska, M.Z., and Barciszewski, J. (2009). Highly active low magnesium hammerhead ribozyme. J. Biochem. 145 , 451–459.

Fernando, T.R., Contreras, J.R., Zampini, M., Rodriguez-Malave, N.I., Alberti, M.O., Anguiano, J., Tran, T.M., Palanichamy, J.K., Gajeton, J., Ung, N.M., et al. (2017). The lncRNA CASC15 regulates SOX4 expression in RUNX1-rearranged acute leukemia. 1–15.

Fior, R., Póvoa, V., Mendes, R.V., Carvalho, T., Gomes, A., Figueiredo, N., and Ferreira, M.G. (2017). Single-cell functional and chemosensitive profiling of combinatorial colorectal therapy in zebrafish xenografts. Proc Natl Acad Sci USA *114*, E8234–E8243.

Fitzpatrick, G.V., Soloway, P.D., and Higgins, M.J. (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat Genet *32*, 426–431.

Flockhart, R.J., Webster, D.E., Qu, K., Mascarenhas, N., Kovalski, J., Kretz, M., and Khavari, P.A. (2012). BRAF V600Eremodels the melanocyte transcriptome and induces BANCRto regulate melanoma cell migration. Genome Research *22*, 1006–1014.

Fornabaio, G., Barnhill, R.L., Lugassy, C., Bentolila, L.A., Cassoux, N., Roman-Roman, S., Alsafadi, S., Del Bene, F. (2018) Angiotropism and extravascular migratory metastasis in cutaneous and uveal melanoma progression in a zebrafish model. Sci Rep 8(1):10448

Ganesh, S., and Svoboda, P. (2016). Retrotransposon-associated long non-coding RNAs in mice and men. Pflügers Archiv - European Journal of Physiology 1–12.

Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67–71.

Gilbert, S.F., (2000). Early Development in fish.

Glusman G., Qin S., El-Gewely M.R., Siegel A.F., Roach J.C., Hood L., and Smit A.F., (2006). A Third Approach to Gene Prediction Suggests Thousands of Additional HumanTranscribed Regions. PLoS Comput Biol

Goedert L., Pereira C.G., Roszik J., Plaça J.R., Cardoso C., Chen G., Deng W., Yennu-Nanda V.G., Silva W.A., Davies M.A., and Espreafico E.M. (2016). RMEL3, A novel BRAFV600E-associated long noncoding RNA is required for MAPK and PI3K signaling in melanoma. Oncotarget v7 n24

Gouzardi, M., Berg, K., Pieper, LM., and Schier, AF. Long non-coding RNAs are largely dispensable for zebrafish embryogenesis, viability and fertility. BioRxiv preprint posted on Jul 23 2018.

Graindorge A., Pinheiro I., Tsvetkov P., Carolis C., Buchholz F., Heard E., Taipale M., and Shkumatava A. (in revision) Systematic Identification and Quantitative Measurement of RNA-Protein Interaction by incPRINT.

Grill, C., and Larue, L. (2016). NRAS, NRAS, Which Mutation Is Fairest of Them All? Journal of Investigative Dermatology *136*, 1936–1938.

Groff, A.F., Sanchez-Gomez, D.B., Soruco, M.M.L., Gerhardinger, C., Barutcu, A.R., Li, E., Elcavage, L., Plana, O., Sanchez, L.V., Lee, J.C., et al. (2016). In Vivo Characterization of Linc-p21 Reveals Functional cis -Regulatory DNA Elements. Cell Reports *16*, 2178–2186.

Grote, P., and Herrmann, B.G. (2014). The long non-coding RNA Fendrrlinks epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. RNA Biology *10*, 1579–1585.

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. Developmental Cell *24*, 206–214.

Guil, S., and Esteller, M. (2012). Cis-acting noncoding RNAs: friends and foes. Nature Structural &Amp; Molecular Biology *19*, 1068–1075.

Gutschner, T., Baas, M., and Diederichs, S. (2011). Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. Genome Research *21*, 1944–1954.

Haemmerle, M., and Gutschner, T. (2015). Long Non-Coding RNAs in Cancer and Development: Where Do We Go from Here? Ijms *16*, 1395–1405.

Haerty, W., and Ponting, C.P. (2013). Mutations within lncRNAs are effectively selected against in fruitfly but not in human. Genome Biol *14*, R49.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. Cell *139*, 945–956.

Hammann C., and Westhof E. (2007). Searching genomes for ribozymes and riboswitches. Genome Biol. 8(4):210.

Hartig J.S. (2012). Ribozymes, Methods and Protocols. Meth in Mol Biol 848

He, M.-D., Zhang, F.-H., Wang, H.-L., Wang, H.-P., Zhu, Z.-Y., and Sun, Y.-H. (2015). Efficient ligase 3-dependent microhomology-mediated end joining repair of DNA double-strand breaks in zebrafish embryos. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis *780*, 86–96.

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 Reports *11*, 1110–1122.

Hiom, K. (2010). Coping with DNA double strand breaks. DNA Repair *9*, 1256–1263.

Hirata, M., Nakamura, K.-I., and Kondo, S. (2005). Pigment cell distributions in different tissues of the zebrafish, with special reference to the striped pigment pattern. Dev. Dyn. *234*, 293–300.

Hirata, M., Nakamura, K.-I., Kanemaru, T., Shibata, Y., and Kondo, S. (2003). Pigment cell organization in the hypodermis of zebrafish. Dev. Dyn. *227*, 497–503.

Hisano, Y., Sakuma, T., Nakade, S., Ohga, R., Ota, S., Okamoto, H., Yamamoto, T., and Kawahara, A. (2015). Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. Sci. Rep. *5*, 8841–8847.

Hombach, S., and Kretz, M. (2013). The non-coding skin: Exploring the roles of long non-coding RNAs in epidermal homeostasis and disease. BioEssays *35*, 1093–1100.

Hoo, J.Y., Kumari, Y., Shaikh, M.F., Hue, S.M., and Goh, B.H. (2016). Zebrafish: A Versatile Animal Model for Fertility Research. Biomed Res Int 1–20.

Hoshijima, K., Jurynec, M.J., and Grunwald, D.J. (2016). Precise genome editing by homologous recombination. In The Zebrafish - Genetics, Genomics, and Transcriptomics, (Elsevier), pp. 121–147.

Hosono, Y., Niknafs, Y.S., Prensner, J.R., Iyer, M.K., Dhanasekaran, S.M., Mehra, R., Pitchiaya, S., Tien, J., Escara-Wilke, J., Poliakov, A., et al. (2017a). Oncogenic Role of THOR, a Conserved Cancer/Testis Long Non-coding RNA. Cell *171*, 1559–1561.e20.

Hua-Van, A., Le Rouzic, A., Boutin, T.S., Filée, J., and Capy, P. (2011). The struggle for life of the genome's selfish architects. Biology Direct *6*, 19.

Huang, H.-Y., Cheng, Y.-Y., Liao, W.-C., Tien, Y.-W., Yang, C.-H.J., Hsu, S.-M., and Huang, P.-H. (2012). SOX4 Transcriptionally Regulates Multiple SEMA3/Plexin Family Members and Promotes Tumor Growth in Pancreatic Cancer. PLoS ONE *7*, e48637–12.

Huang, J., Arsenault, M., Kann, M., Lopez-Mendez, C., Saleh, M., Wadowska, D., Taglienti, M., Ho, J., Miao, Y., Sims, D., et al. (2013). The transcription factor sry-related HMG box-4 (SOX4) is required for normal renal development in vivo. Dev. Dyn. *242*, 790–799.

Hutchins, C.J., Rathjen, P.D., Forster, A.C., and Symons, R.H. (1986) Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid Nucleic Acids Research Volume 14 Number 9

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Kaini, P., Sander, J.D., Joung, J.K., Peterson, R.T., and Yeh, J.-R.J. (2013). Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System. PLoS ONE *8*, e68708–e68709.

Ichigozaki, Y., Fukushima, S., Jinnin, M., Miyashita, A., Nakahara, S., Tokuzumi, A., Yamashita, J., Kajihara, I., Aoi, J., Masuguchi, S., et al. (2015). Serum long non-coding RNA, snoRNA host gene 5 level as a new tumor marker of malignant melanoma. Exp Dermatol *25*, 67–69.

Ilik, I.A., Quinn, J.J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R.C., Luscombe, N., Backofen, R., et al. (2013). Tandem Stem-Loops in roX RNAs Act Together to Mediate X Chromosome Dosage Compensation in Drosophila. Molecular Cell *51*, 156–173.

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., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. Nat Genet *47*, 199–208.

Izraely, S., Sagi-Assif, O., Klein, A., Meshel, T., Tsarfaty, G., Pasmanik-Chor, M., Nahmias, C., Couraud, P.-O., Ateh, E., Bryant, J.L., et al. (2011). The metastatic microenvironment: Brain-residing melanoma metastasis and dormant micrometastasis. Int. J. Cancer *131*, 1071–1082.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., and Charpentier E. (2012). A Programmable Dual-RNA–GuidedDNA Endonuclease in AdaptiveBacterial Immunity. Science 337(6096):816-21.

John (1997). Zebrafish pigmentation mutations and the processes of neural crest development. 1–21.

Johnson, D.B., and Puzanov, I. (2015). Treatment of NRAS-Mutant Melanoma. Curr. Treat. Options in Oncol. *16*, 15–14.

Joung, J., Engreitz, J.M., Konermann, S., Abudayyeh, O.O., Verdine, V.K., Aguet, F., Gootenberg, J.S., Sanjana, N.E., Wright, J.B., Fulco, C.P., et al. (2017). Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. Nature *548*, 343–346.

Ju, B. (2015). Oncogenic KRAS promotes malignant brain tumors in zebrafish. 1–11.

Kalkat, M., De Melo, J., Hickman, K., Lourenco, C., Redel, C., Resetca, D., Tamachi, A., Tu, W., and Penn, L. (2017). MYC Deregulation in Primary Human Cancers. Genes *8*, 151–30.

Kapral, G.J., Jain, S., Noeske, J., Doudna, J.A., Richardson, D.C., and Richardson, J.S. (2014). New tools

provide a second look at HDV ribozyme structure, dynamics and cleavage. Nucleic Acids Research *42*, 12833–12846.

Kapusta, A., and Feschotte, C. (2014). Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. Trends in Genetics *30*, 439–452.

Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., and Feschotte, C. (2013). Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. PLoS Genet *9*, e1003470–20.

Kaufman, C.K., Mosimann, C., Fan, Z.P., Yang, S., Thomas, A.J., Ablain, J., Tan, J.L., Fogley, R.D., van Rooijen, E., Hagedorn, E.J., et al. (2016). A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. Science *351*, aad2197–aad2197.

Kaushik, K., Leonard, V.E., KV, S., Lalwani, M.K., Jalali, S., Patowary, A., Joshi, A., Scaria, V., and Sivasubbu, S. (2013). Dynamic Expression of Long Non-Coding RNAs (lncRNAs) in Adult Zebrafish. PLoS ONE *8*, e83616–12.

Ke, A., Zhou, K., Ding, F., Cate, J.H.D., and Doudna, J.A. (2004b). A conformational switch controls hepatitis delta virus ribozyme catalysis. 377

Khaitan, D., Dinger, M.E., Mazar, J., Crawford, J., Smith, M.A., Mattick, J.S., and Perera, R.J. (2011). The Melanoma-Upregulated Long Noncoding RNA SPRY4-IT1 Modulates Apoptosis and Invasion. Cancer Res. *71*, 3852–3862.

Kim, Y.K., Furic, L., DesGroseillers, L., and Maquat, L.E. (2005). Mammalian Staufen1 Recruits Upf1 to Specific mRNA 3′UTRs so as to Elicit mRNA Decay. Cell *120*, 195–208.

Kimmel C.B., Ballard W.W., Kimmel S.R., Ullmann B., and Schilling T.F. (2004). Stages of Embryonic Development of the Zebrafish. Dev. Dyn. 203:253-310

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (2015). Stages of Embryonic Development of the Zebrafish Stages of Embryonic Development of the Zebrafish. Dev. Dyn. 203 , 253–310.

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. Science Signaling v3-107ra8

Kleaveland, B., Shi, C.Y., Stefano, J., and Bartel, D.P. (2018). A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. Cell 1–31.

Kok, F.O., Shin, M., Ni, C.-W., Gupta, A., Grosse, A.S., van Impel, A., Kirchmaier, B.C., Peterson-Maduro, J., Kourkoulis, G., Male, I., et al. (2015). Reverse Genetic Screening Reveals Poor Correlation between Morpholino-Induced and Mutant Phenotypes in Zebrafish. Developmental Cell *32*, 97–108.

Kotkamp, K., Kur, E., Wendik, B., Polok, B.K., Ben-Dor, S., Onichtchouk, D., and Driever, W. (2014). Pou5f1/Oct4 Promotes Cell Survival via Direct Activation of mych Expression during Zebrafish Gastrulation. PLoS ONE *9*, e92356–12.

Kretz, M., Webster, D.E., Flockhart, R.J., Lee, C.S., Zehnder, A., Lopez-Pajares, V., Qu, K., Zheng, G.X.Y., Chow, J., Kim, G.E., et al. (2012). Suppression of progenitor differentiation requires the long noncoding RNA ANCR. Genes Dev. *26*, 338–343.

Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J., et al. (2012). Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature *493*, 231–235.

Kun-Peng, Z., Xiao-Long, M. and Chun-lin, Z., (2017). LncRNA FENDRR sensitizes doxorubicin-resistance of osteosarcoma cells through down-regulating ABCB1 and ABCC1. Oncotarget v8 (42) pp: 71881-71893

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013).

Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497–501.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 494 , 497–501.

Lai, K.-M.V., Gong, G., Atanasio, A., Rojas, J., Quispe, J., Posca, J., White, D., Huang, M., Fedorova, D., Grant, C., et al. (2015). Diverse Phenotypes and Specific Transcription Patterns in Twenty Mouse Lines with Ablated LincRNAs. PLoS ONE *10*, e0125522–21.

Langenau D.M., Traver D., Ferrando A.A., Kutok J.L., Aster J.C., Kanki J.P., Lin S., Prochownik E., Trede N.S., Zon L.I., and Look A.T. (2003). Myc-Induced T Cell Leukemia in Transgenic Zebrafish. Science 299(5608):887-90

Lavorgna, G., Vago, R., Sarmini, M., Montorsi, F., Salonia, A., and Bellone, M. (2016). Long non-coding RNAs as novel therapeutic targets in cancer. Pharmacological Research *110*, 131–138.

Lee, R.T.H., Ng, A.S.M., and Ingham, P.W. (2016). Ribozyme Mediated gRNA Generation for In Vitro and In Vivo CRISPR/Cas9 Mutagenesis. PLoS ONE *11*, e0166020–12.

Lee B., Sahoo A., Marchica J., Holzhauser E., Chen X., Li J.L., Seki T., Govindarajan S.S., Markey F.B., Batish M., Lokhande S.J., Zhang S., Ray A., and Perera R.J. (2017). The long noncoding RNA SPRIGHTLY acts as an intranuclear organizing hub for pre-mRNA molecules. Sci Adv. 3(5):e1602505

Lessard, L., Liu, M., Marzese, D.M., Wang, H., Chong, K., Kawas, N., Donovan, N.C., Kiyohara, E., Hsu, S., Nelson, N., et al. (2015). The CASC15 Long Intergenic Noncoding RNA Locus Is Involved in Melanoma Progression and Phenotype Switching. Journal of Investigative Dermatology *135*, 2464–2474.

Leucci, E., Vendramin, R., Spinazzi, M., Laurette, P., Fiers, M., Wouters, J., Radaelli, E., Eyckerman, S., Leonelli, C., Vanderheyden, K., et al. (2016). Melanoma addiction to the long non-coding RNA SAMMSON. Nature Publishing Group *531*, 518–522.

Li, M., Zhao, L., Page-McCaw, P.S., and Chen, W. (2016). Zebrafish Genome Engineering Using the CRISPR–Cas9 System. Trends in Genetics *32*, 815–827.

Li, R., Zhang, L., Jia, L., Duan, Y., Li, Y., Bao, L., and Sha, N. (2014). Long Non-Coding RNA BANCR Promotes Proliferation in Malignant Melanoma by Regulating MAPK Pathway Activation. PLoS ONE *9*, e100893–e100899.

Li, Y., Zhang, W., Liu, P., Xu, Y., Tang, L., Chen, W., and Guan, X. (2018). Long non-coding RNA FENDRR inhibits cell proliferation and is associated with good prognosis in breast cancer. Ott *Volume 11*, 1403–1412.

Liu, H.-T., Fang, L., Cheng, Y.-X., and Sun, Q. (2016). LncRNA PVT1 regulates prostate cancer cell growth by inducing the methylation of miR-146a. Cancer Med *5*, 3512–3519.

Liu, S.J., Horlbeck, M.A., Cho, S.W., Birk, H.S., Malatesta, M., He, D., Attenello, F.J., Villalta, J.E., Cho, M.Y., Chen, Y., et al. (2017). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science *355*, eaah7111–eaah7119.

Liu, T., Shen, S.-K., Xiong, J.-G., Xu, Y., Zhang, H.-Q., Liu, H.-J., and Lu, Z.-G. (2016). Clinical significance of long noncoding RNA SPRY4-IT1 in melanoma patients. FEBS Open Bio *6*, 147–154.

Luan W., Li L., Shi Y., Bu X., Xia Y., Wang J., Djangmah H.S., Liu X., You Y., and Xu B. (2017). Long noncoding RNA MALAT1 acts as a competing endogenous RNA to promote malignant melanoma growth and metastasis by sponging miR-22. Oncotarget 7(39):63901-63912

Maguire, L.H., Thomas, A.R., and Goldstein, A.M. (2014). Tumors of the neural crest: Common themes in development and cancer. Dev. Dyn. *244*, 311–322.

Marahens Y., Panning B., Dausman J., Strauss W., and Jaenisch R. (2007). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 11(2):156-66

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville and G. M. Church (2013). "RNA-guided human genome engineering via Cas9." Science 339(6121): 823-826.

Malina, A., Mills, J.R., Cencic, R., Yan, Y., Fraser, J., Schippers, L.M., Paquet, M., Dostie, J., and Pelletier, J. (2013). Repurposing CRISPR/Cas9 for in situ functional assays. Genes Dev. *27*, 2602–2614.

Mallory, A.C., and Shkumatava, A. (2015). LncRNAs in vertebrates: Advances and challenges. Biochimie 1–12.

Marchese, F.P., Grossi, E., Marín-Béjar, O., Bharti, S.K., Raimondi, I., González, J., Martínez-Herrera, D.J., Athie, A., Amadoz, A., Brosh, R.M., Jr., et al. (2016). A Long Noncoding RNA Regulates Sister Chromatid Cohesion. Molecular Cell *63*, 397–407.

Marchese, F.P., Raimondi, I., and Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. 1–13.

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.

Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh, H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nature Biotechnology 1–8.

Matsui, M., and Corey, D.R. (2016). Non-coding RNAs as drug targets. Nature Publishing Group *16*, 167–179.

Mazar J., Zhao W., Khalil A.M., Lee B., Shelley J., Govindarajan S.S., Yamamoto F., Ratnam M., Aftab M.N., Collins S., Finck B.N., Han X., Mattick J.S., Dinger M.E., and Perera R.J. (2014). The Functional Characterization of Long Noncoding RNA SPRY4-IT1 in Human Melanoma Cells. Oncotarget 5:19

McConnell, A.M., Mito, J.K., Ablain, J., Dang, M., Formichella, L., Fisher, D.E., and Zon, L.I. (2018). Neural crest state activation in NRAS driven melanoma, but not in NRAS-driven melanocyte expansion. Developmental Biology 1–0.

MEIERJOHANN, S., and SCHARTL, M. (2006). From Mendelian to molecular genetics: the Xiphophorus melanoma model. Trends in Genetics *22*, 654–661.

Merdrignac, A., Angenard, G., Allain, C., Petitjean, K., Bergeat, D., Bellaud, P., Fautrel, A., Turlin, B., Clément, B., Dooley, S., et al. (2018). A novel transforming growth factor beta-induced long noncoding RNA promotes an inflammatory microenvironment in human intrahepatic cholangiocarcinoma. Hepatology Communications *2*, 254–269.

Miao, Y.-R., Liu, W., Zhang, Q., and Guo, A.-Y. (2017). lncRNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. Nucleic Acids Research *46*, D276–D280.

Michailidou, C., Jones, M., Walker, P., Kamarashev, J., Kelly, A., and Hurlstone, A.F.L. (2009). Dissecting the roles of Raf- and PI3K-signalling pathways in melanoma formation and progression in a zebrafish model.

Michalik K.M., You X., Manavski Y., Doddaballapur A., Zörnig M., Braun T., John D., Ponomareva Y., Chen W., Uchida S., Boon R.A., and Dimmeler S. (2014). Long Noncoding RNA MALAT1 Regulates Endothelial Cell Function and Vessel Growth. Circ. Res. 114(9):1389-97

Mimeault, M., and Batra, S.K. (2013). Emergence of zebrafish models in oncology for validating novel anticancer drug targets and nanomaterials. Drug Discovery Today *18*, 128–140.

Mohammad, F., Mondal, T., Guseva, N., Pandey, G.K., and Kanduri, C. (2010). Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development *137*, 2493–2499.

Mohammadin, S., Edger, P.P., Pires, J.C., and Schranz, M.E. (2015). Positionally-conserved but sequence-diverged: identification of long non-coding RNAs in the Brassicaceae and Cleomaceae. BMC Plant Biology 1–12.

Mondal, T., Juvvuna, P.K., Kirkeby, A., Mitra, S., Kosalai, S.T., Traxler, L., Hertwig, F., Wernig-Zorc, S., Miranda, C., Deland, L., et al. (2018a). Sense-Antisense lncRNA Pair Encoded by Locus 6p22.3 Determines Neuroblastoma Susceptibility via the USP36-CHD7-SOX9 Regulatory Axis. Cancer Cell *33*, 417–434.e417.

Nakade, S., Tsubota, T., Sakane, Y., Kume, S., Sakamoto, N., Obara, M., Daimon, T., Sezutsu, H., Yamamoto, T., Sakuma, T., et al. (2014). Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. Nature Communications *5*, 1–8.

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.

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.

Nakagawa, S. (2016). Lessons from reverse-genetic studies of lncRNAs. BBA - Gene Regulatory Mechanisms *1859*, 177–183.

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J.C., Grützner, F., and Kaessmann, H. (2015). The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature *505*, 635–640.

Neiswender, J.V., Kortum, R.L., Bourque, C., Kasheta, M., Zon, L.I., Morrison, D.K., and Ceol, C.J. (2017). KIT Suppresses BRAF V600E-Mutant Melanoma by Attenuating Oncogenic RAS/MAPK Signaling. Cancer Res. *77*, 5820–5830.

Nelms B.L., and Labosky P.A. (2011). A predicted hairpin cluster correlate with barriers to PCR, sequencing and possibly BAC recombineering. Scient. Rep.

Nicoli, S., Ribatti, D., Cotelli, F., and Presta, M. (2007). Mammalian Tumor Xenografts Induce Neovascularization in Zebrafish Embryos. Cancer Res. *67*, 2927–2931.

Nilsson, J.A., and Cleveland, J.L. (2003). Myc pathways provoking cell suicide and cancer. Oncogene *22*, 9007–9021.

Nishikawa, F., Shirai, M., and Nishikawa, S. (2002). Site-specific modification of functional groups in genomic hepatitis delta virus (HDV) ribozyme. European Journal of Biochemistry *269*, 5792–5803.

Nissen-Meyer, L.S.H., Jemtland, R., Gautvik, V.T., Pedersen, M.E., Paro, R., Fortunati, D., Pierroz, D.D., Stadelmann, V.A., Reppe, S., Reinholt, F.P., et al. (2007). Osteopenia, decreased bone formation and impaired osteoblast development in Sox4 heterozygous mice. Journal of Cell Science *120*, 2785–2795.

Nomura, Y., Zhou, L., Miu, A., and Yokobayashi, Y. (2013). Controlling Mammalian Gene Expression by Allosteric Hepatitis Delta Virus Ribozymes. ACS Synth. Biol. *2*, 684–689.

Paralkar, V.R., Taborda, C.C., Huang, P., Yao, Y., Kossenkov, A.V., Prasad, R., Luan, J., Davies, J.O.J., Hughes, J.R., Hardison, R.C., et al. (2016). Unlinking an lncRNA from Its Associated cis Element. Molecular Cell *62*, 104–110.

Patton, E.E., Widlund, H.R., Kutok, J.L., Kopani, K.R., Amatruda, J.F., Murphey, R.D., Berghmans, S., Mayhall, E.A., Traver, D., Fletcher, C.D.M., et al. (2005). BRAF Mutations Are Sufficient to Promote Nevi Formation and Cooperate with p53 in the Genesis of Melanoma. Current Biology *15*, 249–254.

Pauli, A., Rinn, J.L., and Schier, A.F. (2011). Non-coding RNAs as regulators of embryogenesis. Nat Rev Genet *12*, 136–149.

Pei, D.-S., and Strauss, P.R. (2013). Mutation Research/Fundamental and MolecularMechanisms of Mutagenesis. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis *743-744*, 151–159.

Perry, R.B.-T., and Ulitsky, I. (2016). The functions of long noncoding RNAs in development and stem cells. Development *143*, 3882–3894.

Pollack, V.A., Alvarez, E., Tse, K.F., Torgov, M.Y., Xie, S., Shenoy, S.G., MacDougall, J.R., Arrol, S., Zhong, H., Gerwien, R.W., et al. (2007). Treatment parameters modulating regression of human melanoma xenografts by an antibody–drug conjugate (CR011-vcMMAE) targeting GPNMB. Cancer Chemother Pharmacol *60*, 423–435.

Posch, C., Sanlorenzo, M., Vujic, I., Oses-Prieto, J.A., Cholewa, B.D., Kim, S.T., Ma, J., Lai, K., Zekhtser, M., Esteve-Puig, R., et al. (2016). Phosphoproteomic Analyses of NRAS(G12) and NRAS(Q61) Mutant Melanocytes Reveal Increased CK2α Kinase Levels in NRAS(Q61) Mutant Cells. 1–8.

Prajapati, S., Verma, U., Yamamoto, Y., Kwak, Y.T., and Gaynor, R.B. (2004). Protein Phosphatase 2Cβ Association with the IκB Kinase Complex Is Involved in Regulating NF-κB Activity. J. Biol. Chem. *279*, 1739–1746.

Prensner, J.R., and Chinnaiyan, A.M. (2011). The emergence of lncRNAs in cancer biology. Cancer Discovery *1*, 391–407.

Quinn, J.J., Zhang, Q.C., Georgiev, P., Ilik, I.A., Akhtar, A., and Chang, H.Y. (2016). Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. Genes Dev. *30*, 191–207.

Ramos, A.D., Andersen, R.E., Liu, S.J., Nowakowski, T.J., Hong, S.J., Gertz, C.C., Salinas, R.D., Zarabi, H., Kriegstein, A.R., and Lim, D.A. (2015). The Long Noncoding RNA Pnky Regulates Neuronal Differentiation of Embryonic and Postnatal Neural Stem Cells. Cell Stem Cell *16*, 439–447.

Rasmussen, K.D., and O'Carroll, D. (2011). The miR-144/451eGFP allele, a novel tool for resolving the erythroid potential of hematopoietic precursors. Blood *118*, 2988–2992.

Reautschnig, P., Vogel, P., and Stafforst, T., (2017). The notorious R.N.A. in the spotlight- drug or target for the treatment. RNA Biol 14/5 651-668

Richtig, G., Ehall, B., Richtig, E., Aigelsreiter, A., Gutschner, T., and Pichler, M. (2017). Function and Clinical Implications of Long Non-Coding RNAs in Melanoma. Ijms *18*, 715–719.

Riddihough, G. (2005). In the Forests of RNA Dark Matter. Science *309*, 1507–1507.

Rudel, D., and Sommer, R.J. (2003). The evolution of developmental mechanisms. Developmental Biology *264*, 15–37.

Russell, M.R., Penikis, A., Oldridge, D.A., Alvarez-Dominguez, J.R., McDaniel, L., Diamond, M., Padovan, O., Raman, P., Li, Y., Wei, J.S., et al. (2015). CASC15-S Is a Tumor Suppressor lncRNA at the 6p22 Neuroblastoma Susceptibility Locus. Cancer Res.

Sado T., Wang Z., Sasaki H., and Li E. (2001). Regulation of imprinted X-chromosome inactivation in mice by Tsix. Development 128, 1275-1286

Sandru A., Voinea S., Panaitescu E., and Blidaru A. (2014). Survival rates of patients with metastatic malignant melanoma. Journ. of Med. and Life v7i4 pp.572-576

Santoriello, C., Gennaro, E., Anelli, V., Distel, M., Kelly, A., Köster, R.W., Hurlstone, A., and Mione, M. (2010). Kita Driven Expression of Oncogenic HRAS Leads to Early Onset and Highly Penetrant Melanoma in Zebrafish. PLoS ONE *5*, e15170–11.

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., Frendewey, D., Kellis, M., Hart, R.P., Valenzuela, D.M.,

Arlotta, P., and Rinn, J.L. (2013) Multiple knockout mouse models reveals lincRNAs are required for life and brain development. eLIFE 2;e01749 DOI:10.7554/eLife.0179

Sasaki, M., Ohnishi, M., Tashiro, F., Niwa, H., Suzuki, A., Miyazaki, J.-I., Kobayashi, T., and Tamura, S. (2007). Disruption of the mouse protein Ser/Thr phosphatase 2Cβ gene leads to early pre-implantation lethality. Mechanisms of Development *124*, 489–499.

Scahill, C.M., Digby, Z., Sealy, I.M., Wojciechowska, S., White, R.J., Collins, J.E., Stemple, D.L., Bartke, T., Mathers, M.E., Patton, E.E., et al. (2017). Loss of the chromatin modifier Kdm2aa causes BrafV600E-independent spontaneous melanoma in zebrafish. PLoS Genet *13*, e1006959–24.

Schmid, B., and Haass, C. (2013). Genomic editing opens new avenues for zebrafish as a model for neurodegeneration. J. Neurochem. *127*, 461–470.

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. CellReports *15*, 2025–2037.

Schulte-Merker, S., and Stainier, D.Y.R. (2014). Out with the old, in with the new: reassessing morpholino knockdowns in light of genome editing technology. Development *141*, 3103–3104.

Shain, A.H., and Bastian, B.C. (2016). From melanocytes to melanomas. Nature Publishing Group *16*, 345–358.

Shen, L.-J., Chen, F.-Y., Zhang, Y., Cao, L.-F., Kuang, Y., Zhong, M., Wang, T., and Zhong, H. (2013). MYCN Transgenic Zebrafish Model with the Characterization of Acute Myeloid Leukemia and Altered Hematopoiesis. PLoS ONE *8*, e59070–12.

Singh, V., Govindarajan, R., Naik, S., and Kumar A., (2000). The effect of hairpin structure on PCR amplification efficiency. Molecular Biology Today 1(3):67-69

Sleutels F., Zwart., and Barlow D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature 14;415(6873):810-3

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.

Stuelten, C.H., Parent, C.A., and Montell, D.J. (2018). Cell motility in cancer invasion and metastasis: insights from simple model organisms. Nature Publishing Group 1–17.

Sullivan, R.J., and Flaherty, K. (2012). MAP kinase signaling and inhibition in melanoma. *32*, 2373–2379.

Sun L., Sun P., Zhou Q.Y., Gao X., and Han Q. (2016). Long noncoding RNA MALAT1 promotes uveal melanoma cell growth and invasion by silencing of miR-140. Am J Transl Res 8(9):3939-3946

Tang, L., Zhang, W., Su, B., and Yu, B. (2013). Long Noncoding RNA HOTAIR Is Associated with Motility, Invasion, and Metastatic Potential of Metastatic Melanoma. Biomed Res Int *2013*, 1–7.

Taylor A.M., and Zon L.I. (2009). Zebrafish Tumor Assays: The State of Transplantation. Zebrafish (4):339-46.

Teittinen, K.J., Grönroos, T., Parikka, M., Rämet, M., and Lohi, O. (2012). The zebrafish as a tool in leukemia research. Leukemia Research *36*, 1082–1088.

Teng, Y., Xie, X., Walker, S., White, D.T., Mumm, J.S., and Cowell, J.K. (2013). Evaluating human cancer cell metastasis in zebrafish. BMC Cancer *13*, 1–1.

Thisse B., Thisse C., (2014) In Situ Hybridization on Whole-Mount Zebrafish Embryos and Young Larvae In Situ Hybridization Protocols pp53-67

Tian, Y., Zhang, X., Hao, Y., Fang, Z., and He, Y. (2014). Potential roles of abnormally expressed long

noncoding RNA UCA1 and Malat-1 in metastasis of melanoma. Melanoma Research *24*, 335–341.

Tiwari, N., Tiwari, V.K., Waldmeier, L., Balwierz, P.J., Arnold, P., Pachkov, M., Meyer-Schaller, N., Schübeler, D., van Nimwegen, E., and Christofori, G. (2013). Sox4 Is a Master Regulator of Epithelial-Mesenchymal Transition by Controlling Ezh2 Expression and Epigenetic Reprogramming. Cancer Cell *23*, 768–783.

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., et al. (2010a). The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. Molecular Cell *39*, 925–938.

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., et al. (2010). The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. Molecular Cell *39*, 925–938.

Tseng, Y.-Y., and Bagchi, A. (2015). The PVT1-MYC duet in cancer. Molecular & Cellular Oncology *2*, e974467–3.

Tseng, Y.-Y., Moriarity, B.S., Gong, W., Akiyama, R., Tiwari, A., Kawakami, H., Ronning, P., Reuland, B., Guenther, K., Beadnell, T.C., et al. (2014). PVT1 dependence in cancer with MYC copy-number increase. Nature 1–30.

Tseng, Y.-Y., Moriarity, B.S., Gong, W., Akiyama, R., Tiwari, A., Kawakami, H., Ronning, P., Reuland, B., Guenther, K., Beadnell, T.C., et al. (2015). PVT1 dependence in cancer with MYC copy-number increase. Nature *512*, 82–86.

Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. Nature Publishing Group *17*, 601–614.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: Genomics, Evolution, and Mechanisms. Cell *154*, 26–46.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011a). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell *147*, 1537–1550.

Villegas, V., and Zaphiropoulos, P. (2015). Neighboring Gene Regulation by Antisense Long Non-Coding RNAs. Ijms *16*, 3251–3266.

Vinores, S.A. (2006). Pegatinib in the treatment of wet, age-related macular degeneration. IJN 1(3) 263-268

Viringipurampeer, I.A., Shan, X., Gregory-Evans, K., Zhang, J.P., Mohammadi, Z., and Gregory-Evans, C.Y. (2014). Rip3 knockdown rescues photoreceptor cell death in blind pde6c zebrafish. *21*, 665–675.

Wang, F., Li, X., Xie, X., Zhao, L., and Chen, W. (2008). UCA1, a non-protein-coding RNA up-regulated in bladder carcinoma and embryo, influencing cell growth and promoting invasion. FEBS Letters *582*, 1919–1927.

Werner, M.S., Sullivan, M.A., Shah, R.N., Nadadur, R.D., Grzybowski, A.T., Galat, V., Moskowitz, I.P., and Ruthenburg, A.J. (2017). Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription. Nat Struct Mol Biol *24*, 596–603.

Wellbrock C., Gomez A., and Scharti M. (2002). Melanoma development and pigment cell transformation in xiphophorus. Microsc Res Tech 58(6):456-63

White R.M., Sessa A., Burke C., Bowman T., Leblanc J., Ceol C., Bourque C., Dovey M, Goessling W., Erten Burns C., and Zon L.I (2008); Transparent adult zebrafish as a tool for in vivo transplantation analysis. Cell Stem Cell 7;2 (2): 183-189

White, R.M., Cech, J., Ratanasirintrawoot, S., Lin, C.Y., Rahl, P.B., Burke, C.J., Langdon, E., Tomlinson, M.L., Mosher, J., Kaufman, C., et al. (2011). DHODH modulates transcriptional elongation in the neural crest and melanoma. Nature *471*, 518–522.

Won M., and Dawid I.B. (2017). PCR artifact in testing for homologous recombination in genomic editing in

zebrafish. PLOS ONE

Wu, C.-F., Tan, G.-H., Ma, C.-C., and Li, L. (2018). The Non-Coding RNA Llme23 Drives the Malignant Property of Human Melanoma Cells. 1–10.

Wu, Q., Xiang, S., Ma, J., Hui, P., Wang, T., Meng, W., Shi, M., and Wang, Y. (2018). Long non-coding RNA CASC15 regulates gastric cancer cell proliferation, migration and epithelial mesenchymal transition by targeting CDKN1A and ZEB1. Mol Oncol 1–44.

Xing, Y., Wen, X., Ding, X., Fan, J., Chai, P., Jia, R., Ge, S., Qian, G., Zhang, H., and Fan, X. (2017). CANT1 lncRNA Triggers Efficient Therapeutic Efficacy by Correcting Aberrant lncing Cascade in Malignant Uveal Melanoma. Molecular Therapy *25*, 1209–1221.

Xu, T.-P., Huang, M.-D., Xia, R., Liu, X.-X., Sun, M., Yin, L., Chen, W.-M., Han, L., Zhang, E.-B., Kong, R., et al. (2014). Decreased expression of the long non-coding RNA FENDRR is associated with poor prognosis in gastric cancer and FENDRR regulates gastric cancer cell metastasis by affecting fibronectin1 expression. *7*, 1–15.

Yan, X., Hu, Z., Feng, Y., Hu, X., Yuan, J., Zhao, S.D., Zhang, Y., Yang, L., Shan, W., He, Q., et al. (2015). Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. Cancer Cell *28*, 529–540.

Yanez R.J., and Porter A.C.G (1999). Gene targeting is enhanced in human cells overexpressing hRAD51. Gene Ther 6, 1282-1290.

Yap, K.L., Li, S., Muñoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.-M. (2010). Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. Molecular Cell *38*, 662–674.

Yen, J., White, R.M., and Stemple, D.L. (2014). ScienceDirectZebrafish models of cancer: progress and future challenges. Current Opinion in Genetics & Development *24*, 38–45.

Yoon, J.-H., Abdelmohsen, K., and Gorospe, M. (2013). Posttranscriptional Gene Regulation by Long Noncoding RNA. J. Mol. Biol. *425*, 3723–3730.

Yu, Y.-H., Hu, Z.-Y., Li, M.-H., Li, B., Wang, Z.-M., and Chen, S.-L. (2015). Cardiac hypertrophy is positively regulated by long non-coding RNA PVT1. Int J Clin Exp Pathol *8*, 2582–2589.

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 Reports *2*, 111–123.

Zhang, H., Alberich-Jorda, M., Amabile, G., Yang, H., Staber, P.B., Di Ruscio, A., Welner, R.S., Ebralidze, A., Zhang, J., Levantini, E., et al. (2013). Sox4 Is a Key Oncogenic Target in C/EBPα Mutant Acute Myeloid Leukemia. Cancer Cell *24*, 575–588.

Zhang, Y., Huang, H., Zhang, B., and Lin, S. (2016). TALEN- and CRISPR-enhanced DNA homologous recombination for gene editing in zebrafish (Elsevier Ltd).

Zhao, L., Arsenault, M., Ng, E.T., Longmuss, E., Chau, T.C.-Y., Hartwig, S., and Koopman, P. (2017). SOX4 regulates gonad morphogenesis and promotes male germ cell differentiation in mice. Developmental Biology *423*, 46–56.

Zhao, W., Mazar, J., Lee, B., Sawada, J., Li, J.-L., Shelley, J., Govindarajan, S., Towler, D., Mattick, J.S., Komatsu, M., et al. (2016). The Long Noncoding RNA SPRIGHTLY Regulates Cell Proliferation in Primary Human Melanocytes. Journal of Investigative Dermatology *136*, 819–828.

Zheng, X., Hu, H., and Li, S. (2016). High expression of lncRNA PVT1 promotes invasion by inducing epithelial‐to-mesenchymal transition in esophageal cancer. Oncol Lett 1–6.

Zhigui, L., Shuang, H., Hongqiang, Y., Jing, G., and Zhuo, Y. (2016). Autophagy ameliorates cognitive impairment through activation of PVT1 and apoptosis in diabetes mice. Behavioural Brain Research *305*, 265–277.

Ziegler, C., and Kretz, M. (2017). The More the Merrier—Complexity in Long Non-Coding RNA Loci. Front. Endocrinol. *8*, 101–106.

ABBREVIATIONS

Airn	*Antisense Of IGFR2 Non Protein-Coding RNA*
AKT	Protein Kinase B
ANCR	*Angelman Syndrome Chromosome Region*
ANRIL	*Antisense Noncoding RNA In The INK4 Locus*
ASO	Allele Specific Oligonucleotides
ATA	Alternative cleavage and polyA signal
BAC	Bacteria Artificial Chromosome
BANCR	*BRAF Activated Non Protein-Coding RNA*
Bp	base pair
CANT1 CASC15 New Transcript 1	
CASC15	*Cancer Susceptibility 15*
CASC15-S	*Cancer Susceptibility 15 Short*
CDK	Cyclin Dependent Kinase
CHD7	Chromodomain Helicase DNA binding protein 7
Chr	Chromosome
ChIRP-MS	Chromatin isolation by RNA precipitation followed by Mass Spectrometry
CK2α	Casein Kinase 2
CM	Cutaneous Melanoma
CNS	Central Nervous System
CONCR	*Cohesion regulator noncoding RNA*
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTCF	CCCTC Binding Factor
Cyaa	Calmodulin sensitive adenylated cyclase
Dbh	Dopamine beta hydroxylase
DBP	DNA binding proteins
DDXII	Dead H box Helicase 11
DNA	Desoxyribo Nucleic Acid
dpf	day post fertilization
DSB	Double strand break
EBI	European Bioinformatics Institute
EMICERI	*EQTN MOB3B IFNK C9orf72 enhancer RNA I*
ENU	N-ethyl-N-nitrosourea
ERK	Extracellular signal Regulated Kinase
Fendrr	*FOXF1 Adjacent NonCoding Developmental Regulatory RNA*
GAS5	*Growth arrest specific 5*
GDP	Guanine DiPhosphate
GFP	Green Fluorescent Protein
Gosr2	Golgi SNAP Receptor Complex Member 2
gRNA	guide RNA
GSP	Gene Specific Primer
GTEx	Gene Tissue Expression
GTP	Guanine TriPhosphate
GTPase	Guanine TriPhosphatase
Gy	Gray (irradiation unit)
HDV	Hepatitis Delta Virus

HOTAIR	*HOX Transcript Antisense RNA*
hpf	hour post fertilization
HR	Homologous Recombination
IGF2BP1	Insulin Like Growth Factor 2 mRNA Binding Protein 1
incPRINT	in cell protein-RNA interaction
INK4	cyclin independent kinase inhibitors
kb	kilobase
Kcnq1ot1	*Potassium voltage-gated channel subfamily member Q1 opposite transcript 1*
KD	Knock-Down
Kdrl	Kinase insert domain receptor like
Kdm2aa	Lysine demethylase 2aa
KI	Knock-in
Kita	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog a
Klf7b	Kruppel Like Factor 7b
KO	Knock-Out
libra	*lncRNA involved in behavioural alterations*
Linc-MD1	*long intervening noncoding muscle differentiation 1*
LNA	Locked Nucleic Acid
lncRNA	long noncoding RNA
Malat1	*Metastasis Associated Lung Adenocarcinoma Transcript 1*
MAPK	Mitogen Activated Protein Kinase
MEK	Mitogen-activated protein kinase
menhir	*MElaNoma Hindrance long noncoding RNA*
Mg2+	Magnesium
miRNA/miR	microRNA
mitfa	Microphtalmia Transcription Factor
mM	millimollar
MMEJ	Micro-homology Mediated End Joining
MMP	Matrix MetalloProteinase
Mn2+	Manganese
MO	Morpholino
mTOR	mammalian Target Of Rapamycin
mRNA	messenger RNA
my	million years
ncRNA	noncoding RNA
Neat1	*Nuclear Paraspeckle Assembly Transcript 1*
NF1	Neurofibromin
NHEJ	Non Homologous End Joining
NOP14	Nucleolar protein 14
Nrep	Neuronal Regeneration Related Protein
nts	nucleotides
ORF	Open reading frame
PAUPAR	*Pax6 upstream antisense RNA*
PCAWG	PanCancer Analysis of Whole Genomes
PCR	Polymerase Chain Reaction
PIM2	Proviral Integrations of Moloney virus 2
PI3K	Phosphoinositide 3-kinase
Pnky	*Pinky*

POP1	ribonuclease P/MRP subunit
Ppm1bb	Protein phosphatae Mg2+/Mn2+ dependent 1bb
PTEN	Phosphatase and TENsin homolog
PVT1	*Plasmacytoma Variant Transcript 1*
qRT-PCR	quantitative Retro Transcriptase Polymerase Chain Reaction
RBP	RNA binding Proteins
RDE	RNA Destabilizing Element
RFP	Red Fluorescent Protein
RGP	Radial Growth Progression
RISC	RNA-induced silencing complex
RIP3	Receptor interacting protein 3
RNA	Ribo Nucleic Acid
RNAi	RNA interference
RNAseq	RNA sequencing
rRNA	ribosomal RNA
SAMMSON	*Survival Associated Mitochondrial Melanoma Specific Oncogenic Non-Coding RNA*
SETDB1	SET Domain Bifurcated 1
SHAPE	Selective 2'-Hydroxy Acylation analysed by Primer Extension
shRNA	small hairpin RNA
siRNA	small interfering RNA
SLNCR1	*SRA Like non-coding RNA1*
SNGH5	*SnoRNA host gene 5*
SNP	Single Nucleotide Polymorphism
SOX4	SRY-Box 4
SPRIGHTLY	*Sprouty 4 intron 1*
SS	Splice Site
Ss oligos	single stranded oligos
TCGA	The Cancer Genome Atlas
TE	Transposable Elements
TF	Transcription factor
TFBS	Transcription Factor Binding Site
THOR	*Testis-associated Highly conserved Oncogenic long noncoding RNA*
TINCR	*Tissue differenciation-inducing non-protein-coding RNA*
TLINC	*TGFβ-induced long noncoding RNA*
TPM	Transcript Per Million
tRNA	transfert RNA
Tsix	*Transcript Xist Antisense RNA*
TSS	Transcription Start Site
UCA1	*Urothelial Cancer Associated 1*
UM	Uveal Melanoma
USP36	Ubiquitin specific peptidase 36
VGP	Vertical Growth Progression
wpf	weeks post fertilization
WT	Wild Type
Xist	*X inactive specific transcript*
Xrcc4	X-ray repair cross complementing 4
Zfp64	Zinc Finger protein 64
Znf507	Zinc Finger protein 507

RESUME (Version longue)

a) Introduction

En biologie moléculaire, « l'ADN produit l'ARN qui produit la protéine » a été le dogme central en vigueur depuis les années 60, suggérant que la molécule d'ARN n'avait d'autres fonctions que celles de transmettre l'information entre l'ADN et la machinerie de traduction protéique (Crick, 1966). Ce dogme a été réfuté dans les années 2000 par les résultats obtenus avec l'avancée des techniques de séquençages du génome, annonçant l'avènement du monde des ARN qui ne codent pas pour des protéines, ou ARN non codants (Glusman et al., 2006 ; Ulitsky et al., 2011 ; Derrien et al., 2012). Les ARN non codants (ou ncARN) représentent une grande famille de gène qui ont été classés selon leur taille : les petits ncARN (<50 nucléotides), les ncARN intermédiaires (entre 50 et 200 nucléotides) et les longs ncARN (>200 nucléotides). Les longs ARN non codants (ou lncARN) sont structurellement similaires aux ARN messagers (transcription par l'ARN polymérase II, épissage, coiffe en 5' et queue polyA en 3') si ce n'est qu'ils ne présentent pas de potentiel codant en protéines (Mallory et al., 2015). De nombreux lncARN ont été identifiés dans divers organismes, des procaryotes aux organismes multicellulaires les plus complexes où ils ont été associés à de multiples processus physiologiques et pathologiques (Ulitsky et al., 2011 ; Ilik et al., 2013 ; Nesculea et al., 2015 ; Mohammadin et al., 2015). Pour autant, la fonction et le rôle de la plupart des longs ARN non codants restent encore méconnues.

La conservation d'un gène dans l'évolution est très souvent un indicateur de fonctionnalité et les lncARN peuvent être conservés à différents niveaux. Tout comme les gènes codants pour des protéines, les lncARN peuvent présenter une conservation de séquence (1), ce phénomène est toutefois peu fréquent à l'échelle des vertébrés (seulement 2% des lncARN présentent une conservation partielle de séquence entre l'homme et le poisson zèbre) (Ulitski et al., 2011 ; Hezroni et al., 2015). Il a également été observé que les lncARN sont exprimés de façon très spécifique au niveau tissulaire, et que cette expression est conservée dans l'évolution (2) (Chodroff et al., 2010 ; Necsulea et al., 2015). Tout comme les protéines, les ARN sont des molécules qui forment des structures tridimensionnelles hautement dynamiques, et ces structures peuvent être également conservées dans l'évolution (3) (Quinn et al., 2016). Les lncARN ont aussi été observés comme conservés à l'échelle de

leur schéma d'épissage (4) (Hezroni et al., 2015) et de leur position génomique (5) (Brockdorff et al., 1991 ; Ulitsky et al., 2011 ; Hezroni et al., 2015). En 2011, Igor Ulitsky et al., ont découvert que près de 35% des lncARN du poisson zèbre ont leur position conservée chez l'homme (Ulitsky et al., 2011), et ce malgré une distance évolutive de plus de 450 millions d'années et une duplication génomique intégrale observée chez les téléostéens (Hoegg et al., 2004). Ce phénomène de conservation de position est appelé synténie.

Parmi les nombreux lncARN identifiés par le séquençage des génomes de nombreuses espèces, seul un très faible pourcentage ont vu leur fonction caractérisées de façon précise. D'une façon générale, les lncARN ont des fonctions de régulateurs de l'expression génique (Marchese et al., 2017). Pour cela, ils peuvent interagir avec l'ADN (Yap et al., 2010 ; Grote et al., 2013 ; Joung et al., 2017), d'autres ARN (Tripathi et al., 2010 ; Yoon et al., 2013 ; Bitetti et al., 2017) ou encore des protéines (Leucci et al., 2016 ; Hosono et al., 2017 ; Mondal et al., 2018) via la formation de structure secondaire. Ils peuvent donc moduler de nombreux processus cellulaires, allant de l'organisation 3D du génome à la régulation de l'épissage des ARN messagers ou l'assemblage des complexes protéiques.

On peut analyser la fonction d'un lncARN selon deux approches : soit en identifiant les partenaires (le plus souvent protéiques) du transcrit, soit en analysant les conséquences moléculaires et phénotypiques de son inactivation. Les lncARN sont pourtant des molécules complexes à inactiver (Ziegler and Kretz, 2017). Pour les gènes codants pour des protéines, la perturbation du cadre de lecture lors de l'étape de traduction suffit à inactiver complètement leur fonction. Mais de telles approches ne sont pas envisageables pour les gènes non codants, qui présentent des loci génomiques complexes (longs de plusieurs dizaines voir centaines de kilobases, présentant de multiple isoformes ou chevauchant d'autres gènes) (Ziegler and Kretz, 2017). De plus, le faible taux de conservation de séquence des lncARN ne permet pas d'identifier et de cibler les domaines fonctionnels de façon efficaces (Ulitski et al., 2011 ; Hezroni et al., 2015). Il existe toutefois plusieurs méthodologies pour inactiver les lncARN de façon transitoires ou stables. Les approches transitoires consistent en la déstabilisation de la molécule d'ARN, et ne peuvent être efficaces pour évaluer la fonction d'un gène à l'échelle de l'organisme. Les approches d'inactivation stables consistent en la suppression ou le remplacement du gène, la suppression ou la répression de leur promoteur, ou encore l'insertion de signaux prématurés de terminaison de transcription (queues polyA) ou d'éléments dits « déstabilisants ». Cette dernière approche apparaît comme la moins invasive, elle nécessite toutefois l'intégration précise et efficace de séquences exogènes dans le

génome. S'il existe de nombreuses méthodes permettant d'effectuer de telles intégrations, elles peuvent toutefois être assez inefficaces dans certains organismes modèles comme le poisson zèbre (Hwang et al., 2013 ; Hisano et al., 2015 ; Albadri et al., 2017). En collaboration avec Angelo Bitetti, étudiant en thèse au sein du même laboratoire, j'ai consacré une partie de ma thèse à l'amélioration de ces techniques d'intégration génomiques comme chez le poisson zèbre, ainsi qu'à l'étude de l'efficacité de certains éléments déstabilisant sur les lncARN (signaux polyA prématurés ou ribozymes).

b) Projet de thèse

Lors de ma thèse, j'ai analysé la relation entre le phénomène synténique et la conservation de fonction des lncARN en utilisant le modèle poisson zèbre. Pour cela, j'ai caractérisé le profil d'expression de plusieurs lncARN synténique dans l'embryons et les tissus adultes de poisson zèbre, et généré plusieurs lignées de poissons zèbres mutants pour ces gènes candidats. Dans la littérature, peu de défauts morphologiques, physiologiques ou comportementaux ont été rapportés comme spécifiques à l'inactivation d'un lncARN, et ce probablement à cause de mécanismes redondants qui permettent à l'organisme de compenser la perte de fonction du gène. La caractérisation phénotypique de mutants lncARN à l'échelle d'un organisme entier est donc un exercice difficile. J'ai toutefois réussi à décrire des anomalies physiologiques dues à l'inactivation de gènes lncARN chez le poisson zèbre. J'ai en effet observé d'importantes malformations lors du développement embryonnaires des doubles mutants lnc-*myca* et lnc-*mycb*. J'ai également mis en place un système permettant d'évaluer l'impact des lncARN sur la progression du cancer de la peau et décrit le rôle d'un lncARN suppresseur de tumeur, le gène *menhir*. Les résultats obtenus pendant ma thèse attestent donc de la fonctionnalité et de l'importance des lncARN *in vivo*.

c) Fonctionnalité des lncARN dans le développement embryonnaire des vertébrés

Certains lncARN, tels que *XIST* ou *TSIX* (Maharens et al., 1997 ; Sado et al., 2001), ont été reconnus comme indispensables à la survie et au développement normal des organismes. Ils présentent en effet des fonctions cruciales dans l'inactivation du chromosome X et la compensation de dose entre les sexes. Toutefois, certains lncARN semblent ne pas être nécessaires au bon fonctionnement d'un organisme, malgré leur niveau d'expression parfois très élevé. *MALAT1* et *HOTAIR* sont des lncARN qui ont été décrits comme fonctionnels dans

de nombreuses lignées cellulaires, toutefois leur fonction *in vivo* restent encore méconnues (Amândio et al., 2016). Les modèles murins mutants pour ces gènes ne présentent pas ou peu de défauts phénotypiques (Zhang et al., 2012 ; Nagakawa et al., 2012 ; Eissmann et al., 2012), la seule anomalie rapportée chez les mutants *MALAT1* étant une malformation du réseau vasculaire rétinien (Michalik et al., 2014). Il est donc possible qu'à l'échelle de l'organisme, des mécanismes redondants compensent l'inactivation du lncARN et suppléent sa perte de fonction. D'après mes connaissances, de tels dispositifs moléculaires n'ont toutefois pas encore été identifiés.

Le lncARN *PVT1*, ou *Plasmacytoma Variant Transcript 1*, a été dépeint comme un oncogène dans de nombreux cancers, mais il n'a pas été décrit comme essentiel aux processus physiologiques normaux tels que le développement embryonnaire (Tseng et al., 2015 ; Chen et al., 2017). *PVT1* représente un locus comprenant 25 isoformes (de différentes tailles, avec de multiples points d'initiation et de terminaison) en aval du gène codant *MYC*. Si *PVT1* a sa séquence conservée au sein des mammifères, il a été décrit comme strictement synténique parmi les vertébrés (Hezroni et al., 2015). Chez le poisson zèbre, les deux isoformes protéiques *myca* et *mycb* sont associés à des lncARN présentant une conservation de position, d'orientation transcriptionnelle et de profils d'épissages avec certains isoformes *PVT1*.

Afin d'inactiver ces gènes chez le poisson zèbre, j'ai supprimé la majeur partie de leur séquence génomique en utilisant la technique d'édition du génome CRISPR-Cas9 (Hwang et al., 2013). Les individus mutants pour un seul lnc-*myc* ne présentent aucun défaut morphologiques ou problèmes de viabilité et de fertilité dans les conditions de laboratoire. En combinant les deux mutations, j'ai observé l'émergence de sévères anormalités morphologiques chez les embryons lnc-*myca*$^{-/-}$ lnc-*mycb*$^{-/-}$. Le phénotype observé est partiellement pénétrant (65%), toutefois seuls 22% des individus double mutants survivent jusqu'à l'âge adulte.

Afin de caractériser plus précisément l'aspect moléculaire de ce phénotype embryonnaire, j'ai analysé les niveaux d'expression des lncARN et de leur gène codants adjacents par qRT-PCR. Alors que les niveaux transcriptionnels des protéines *myca* et *mycb* paraissent constants entre les embryons sauvages, les simples et les doubles mutants, j'ai observé une forte surexpression de l'isoforme lnc-*mycb* dans le mutant lnc-*myca*$^{-/-}$ et inversement. Le phénotype embryonnaire observé chez les doubles mutants serait alors spécifique à la perte de fonction des deux isoformes lnc-*myc*. Ces résultats m'ont donc permis

de dévoiler la présence d'un potentiel mécanisme de compensation entre les homologues de *PVT1*, et ceux, malgré l'absence de séquences communes aux deux isoformes.

Une caractérisation plus complète des phénotypes morphologiques et surtout moléculaires est encore nécessaire. En effet, Il est probable qu'un second mécanisme de compensation génétique soit mis en place dans les doubles mutants lnc-*myc*, expliquant la pénétrance partielle du phénotype et la survie de 22% des individus. Des analyses transcriptionnelles nous aiderons à dévoiler ce mécanisme compensatoire. Chez l'homme, *PVT1* a été décrit comme stabilisant la protéine *MYC* de part son interaction. Nous analyserons donc les niveaux d'expression des protéines *myca* et *mycb* par Western Blot. Enfin, nous évaluerons la conservation de fonction des isoformes lnc-*myc* avec leurs homologues humain *PVT1*. Pour cela, nous générerons des lignées transgéniques exprimant un isoforme du lncARN humain et tenterons une expérience de sauvetage phénotypique du mutant ou « rescue ».

Les mutants *Fendrr* sont les seuls mutants lncARN à avoir illustrés un phénotype embryonnaire aussi sévère (Grote et al., 2013 ; Sauvageau et al., 2013), et à notre connaissance, aucun autre mécanisme de compensation n'a encore été rapporté dans le domaine des lncARN. Notre étude, une fois approfondie, dévoilera donc une nouvelle facette de la conservation des lncARN.

d) *Fonctionnalité des lncARN dans la genèse et la progression du mélanome cutané*

En sus de leurs fonctions reconnues dans les processus développementaux et physiologiques, les lncARN sont fréquemment étudiés en oncologie. Ils sont en effet souvent dérégulés chez les patients souffrant de cancers (Aftab et al., 2014 ; Yan et al., 2015 ; etc.). Certains lncARN présentent même de fort potentiels comme cibles thérapeutiques ou bio-marqueurs de la progression tumorale (Prensner and Chinnaiyan, 2011 ; Bhan et al., 2017).

Le mélanome cutané est le cancer des mélanocytes, les cellules responsables de la pigmentation de la peau. Les lésions mélanocytaires évoluent du stade de naevus mélanocytaire, en mélanome de stade I naevus dysplasique, en mélanome précancéreux ou « *in situ* » de stade II, puis au stade de mélanome invasif présentant des métastases locales au stade III ou distantes au stade IV (Shain and Bastian, 2016). Si les mélanomes de stades I et II peuvent être facilement traités via des approches chirurgicales, les mélanomes métastatiques

sont extrêmement mortels, et leur incidence augmente de 3,7% chaque année (Sandru et al., 2014).

Jusqu'à aujourd'hui, 17 lncARN ont été rapportés comme dérégulés chez les patients souffrants de mélanome (Aftab et al., 2014 ; Richtig et al., 2017). Dans la plupart des cas, le rôle de ces lncARN dans la mélanogenèse a été étudié via des mutations transitoires dans des lignées cellulaires de mélanome humain. *THOR* est le seul exemple de lncARN analysé dans un organisme modèle par inactivation stable du transcrit (Hosono et al., 2017). En 2017, Hosono et al., ont utilisé le modèle poisson zèbre et observé que la progression du mélanome était moins agressive chez les mutants *thor*$^{-/-}$. En exprimant de façon exogène l'isoforme *THOR* humain, ils ont également réussi à effectuer un sauvetage phénotypique et à restituer son agressivité au cancer (Hosono et al., 2017). Le lncARN *THOR* présente une conservation de séquence et de profil d'expression, mais est également le premier lncARN reconnu comme présentant une fonction oncogénique conservée dans l'évolution.

Le poisson zèbre, ou *Danio rerio* est un modèle courant en oncologie depuis les années 2000 (Amsterdam et al., 1999 ; Langenau et al., 2003), particulièrement dans l'étude du mélanome cutané. Les mélanocytes humains présentent des caractéristiques communes avec les mélanophores du poisson zèbre, ces deux types cellulaires ayant la même origine embryologique, la même fonction productrice de mélanine, ainsi que des propriétés génétiques, histologiques et physiologiques similaires (John et al., 1997 ; Shain and Bastian, 2016). Les lésions et tumeurs mélanocytaires du poisson zèbre présentent également des caractéristiques histologiques et moléculaires hautement analogues aux lésions cutanées humaines (Patton et al., 2005).

Contrairement à d'autres types de poissons comme les xiphophores (Wellbrock et al., 2002 ; Meierjhann et Schartl., 2006), les poissons zèbres ne développent pas de mélanome cutané de façon naturelle. Le développement mélanocytaire doit en effet être induit. A ce jour, quatre types de méthodologies ont été utilisée afin d'étudier le mélanome sur le modèle poisson zèbre. (1) L'utilisation d'agents mutagènes ou d'insertion de séquences rétrovirales ou transposons sont des techniques classiques de mutagenèses aléatoires qui ont déjà été utilisées sur le poisson zèbre pour induire le développement tumoral (Amsterdam et al., 1999). Plus récemment, l'étude des acteurs et de la progression tumorale des mélanomes cutanés et uvéaux ont été étudiés par des approches de mutagenèse dirigée (2) ou de xénogreffes (3) sur le modèle poisson zèbre. Enfin, les embryons de poisson zèbre sont de plus en plus utilisés pour tester l'efficacité de nouvelles drogues ou traitements (Ju et al.,

2015 ; Fior et al., 2017). Les embryons de poisson zèbre sont de petites tailles et se développent ex-utero. De plus, les embryons et larves de poisson zèbre sont perméables aux molécules solubles, et présentent une sensibilité aux drogues similaires à celle des modèles murins (Fior et al., 2017).

Afin d'étudier l'impact des lncARN sur la progression du mélanome, nous avons utilisé une technique de mutagenèse dirigée. Nous avons inséré dans le génome du poisson zèbre une construction exprimant un oncogène humain et induisant la progression mélanocytaire. De nombreuses constructions ont été établies comme efficaces pour induire la cancérogenèse chez le poisson zèbre (Patton et al., 2005 ; Michailidou et al., 2009 ; Santoriello et al., 2010). Nous avons choisi de déclencher le développement mélanocytaire en insérant l'oncogène $NRAS^{G12}$ sous le contrôle du promoteur mitfa, un gène spécifique des mélanocytes. Si cette mutation n'est retrouvée que dans 3-4% des patients atteints de mélanome (TCGA, Akbani et al., 2015), les mélanomes $NRAS^{G12}$ sont parmi les tumeurs les plus agressives et il existe très peu de solutions thérapeutiques pour traiter ces cancers (Sullivan et Flaherty., 2012). Afin de caractériser de nouvelles cibles thérapeutiques, nous avons évalué la progression du mélanome dans des lignées de poissons zèbres mutants pour six lncARN. Quatre d'entre eux ont un homologue humain qui a été rapporté comme dérégulé chez les patients souffrants de mélanome (*malat1*, lnc-*myca* et lnc-*mycb* sont les homologues de *PVT1*, lnc-*sox4a* est l'homologue de *CASC15*), le cinquième est associé à la neurogenèse (*megamind,* Kok et al., 2015) et le dernier n'a pas de fonctionnalité connue à ce jour (lnc-*ppm1bb*). Afin d'induire le mélanome, j'ai injecté la construction mitfa :$NRAS^{G12}$ en combinaison avec l'ARNm de la transposase Tol2 dans des embryons sauvages ou mutants pour les différents lncARN. L'insertion de la construction oncogénique avec le système Tol2 étant aléatoire, nous avons généré un minimum de 50 individus porteur de la construction par lignées, afin d'obtenir des données qui seront statistiquement significatives. J'ai suivi la progression du mélanome de la semaine 5 à la semaine 21 post-injection, rapportant chaque semaine le stade de sévérité du cancer et l'apparition de nodules ou de tumeurs pour chaque individu. Pour cela, j'ai utilisé un système de classification de la sévérité des lésions mélanocytaires : absence de lésion, naevus dysplasique, progression radiale, progression verticale externe, et enfin formation de nodules, tumeurs ou hyperplasies caudales (Michailidou et al., 2009).

Je n'ai observé aucune différence dans la progression mélanocytaire et la tumorigenèse des mutants *megamind*, lnc-*ppm1bb*, *malat1* ou lnc-*myc*, et ceux malgré le rôle présumé des homologues humain *MALAT1* et *PVT1* dans le mélanome (Tian et al., 2014 ; Chen et al.,

2017). Toutefois, j'ai observé une altération de la progression mélanocytaire des mutants lnc-*sox4a*, caractérisée par une augmentation de la tumorigenèse et de l'agressivité du mélanome. Nous avons donc renommé ce lncARN *menhir* (pour *Melanoma Hindrance long noncoding RNA*).

Le lncARN *menhir* est l'homologue du lncARN humain *CASC15* (ou *Cancer Susceptibility 15)* chez le poisson zèbre. *CASC15* et *menhir* ne présentent pas de conservation de séquence, mais sont des lncARN synténiques (Ulitsky et al., 2011 : Hezroni et al., 2015). En analysant l'expression de *menhir* dans les tissus et les tumeurs mélanocytaires du poisson zèbre, j'ai observé que *menhir* et *CASC15* sont également conservé au niveau du profil d'expression.

Les mutants *menhir*$^{-/-}$, obtenus via la délétion du site d'initiation de la transcription du lncARN, présentent également une diminution de la survie et une augmentation de la sévérité des lésions mélanocytaires par rapport aux autres lignées de poissons zèbres. En décrivant la progression du mélanome pendant 15 semaines, j'ai également observé un phénomène de tumorigenèse interne d'origine mélanocytaire uniquement chez les mutant *menhir*$^{-/-}$. Les sévères lésions mélanocytaires externes et internes observées sur les mutants *menhir*$^{-/-}$ suggèrent donc que ce lncARN jouerait un rôle de gène suppresseur de tumeur.

Chez l'homme, le déclenchement du processus métastatique est mortel dans 80% des patients atteints de mélanome. Il est donc essentiel de caractériser le rôle de *menhir* dans le processus métastatique. Celui-ci peut être divisé en plusieurs étapes : (1) perte de la polarité des cellules tumorales, (2) altérations de l'architecture tissulaire, (3) percée de la membrane basale par les cellules tumorales, (4) intravasation des cellules tumorales dans les vaisseaux sanguins ou lymphatiques, (5) extravasation, (6) colonisation du nouvel organe, (7) expansion de la colonie métastatique selon la permissivité de l'environnement parfois associée à (8) un processus d'angiogenèse (Stuelten et al., 2018). Si de nombreux tests *in vitro* ont été développés pour analyser le processus métastatique, ils ne peuvent toutefois pas prendre en compte les propriétés biologiques inhérentes à un organisme vivant. Les propriétés de transparence des embryons et des adultes font du poisson zèbre un excellent modèle pour étudier l'initiation et la progression métastatique des xénogreffes en microscopie *in vivo*. Les propriétés invasives et profils moléculaires de nombreuses lignées cancéreuses humaines ont déjà été caractérisées lors de xénogreffes chez le poisson zèbre (Teng et al., 2013). Afin d'étudier si le potentiel métastatique des cellules cancéreuse peut être modulé par le lncARN *menhir*, nous avons transplanté des cellules issues de lignées de mélanome humain dans des

embryons de poissons zèbres sauvage et *menhir*$^{-/-}$. Les cellules humaines exprimant la GFP, nous avons pu suivre leur dissémination au sein de l'embryon. A quatre jours post-greffes, nous avons observé une augmentation du nombre de cellules 501mel ayant quitté le site de transplantation et colonisé l'organisme des embryons *menhir*$^{-/-}$. Ces résultats suggèrent donc que le rôle suppresseur de tumeur de *menhir* est également nécessaire au maintien de l'intégrité des processus de défenses à l'invasion métastatique.

Chez l'homme, *CASC15* consiste en un locus comprenant de multiples isoformes, exprimés de façon différentielle selon les types de cancers. La fonctionnalité de ce transcrit semble également être différente selon les cancer : Il est décrit comme un gène suppresseur de tumeur dans le mélanome uvéal et les neuroblastomes (Russels et al., 2015 ; Xing et al., 2017 ; Mondal et al., 2018) et comme un oncogène dans les leucémies ou les cancers du système digestif (Chen et al., 2016 ; Fernando et al., 2017 ; Merdrignac et al., 2018 ; Wu et al., 2018). Dans le mélanome cutané *CASC15* a été décrit comme impliqué dans la transition entre les stades prolifératifs et invasifs des cellules tumorales (Lessard et al., 2015).

Afin d'évaluer la possible conservation de fonction des lncARN synténiques *CASC15* et *menhir*, nous avons modifié la construction mitfa::NRAS^{G12} en y insérant une cassette contenant un isoforme humain *CASC15* exprimé sous le contrôle du promoteur mitfa. Cette nouvelle construction a été injectée chez les mutants *menhir*$^{-/-}$ et nous avons évalué la progression mélanocytaire de la même façon décrite précédemment. Les résultats que nous avons obtenus pour l'instant montrent une diminution de l'agressivité du mélanome, caractérisée par une décroissance de la tumorigenèse, une augmentation du taux de survie et un réduction de la sévérité de la progression mélanocytaire, suggérant que l'expression exogène de *CASC15* est suffisante au sauvetage phénotypique des mutants *menhir*$^{-/-}$ et que les deux homologues ont leur fonction conservée dans l'évolution.

La description du lncARN *CASC15* et de sa fonction oncogénique ou suppresseur de tumeur est de plus en plus détaillé dans les modèles de culture cellulaire. Notre étude est la première à reconsidérer son rôle dans l'évolution et *in vivo* via l'inactivation de son homologue *menhir* chez le poisson zèbre. Nous avons pour l'instant caractérisé l'impact de l'inactivation de *menhir* et son sauvetage phénotypique par *CASC15* en observant les lésions cutanées externes. Afin d'approfondir notre caractérisation de la fonctionnalité de ces lncARN, nous étudierons le potentiel invasif des lésions cutanées observées via l'analyse de coupes histologiques. Nous évaluerons également l'impact de *CASC15* sur le potentiel

métastatique intrinsèque aux cellules tumorales en répétant les expériences de transplantations avec des lignées cellulaires mutantes pour ce lncARN. En cas d'altération du potentiel invasif, nous tenterons un sauvetage phénotypique en induisant l'expression du lncARN *menhir* complet ou partiellement tronqué, afin de déterminer quels sont les domaines fonctionnels du transcrit. Si les résultats obtenus confirment nos précédentes observations, nous aurons donc identifié le premier lncARN suppresseur de tumeur étudié *in vivo* ayant sa fonction conservée dans l'évolution. Notre étude décrira également que deux lncARN peuvent avoir leur fonction conservée dans l'évolution en absence de conservation de séquence.

e) Conclusion

Le phénomène synténique a déjà été observé dans de nombreux organismes, allant des plantes (Mohammadin et al., 2015) aux invertébrés (Quinn et al., 2016) et aux vertébrés (Ulitsky et al., 2011 ; Hezroni et al., 2015 ; Amaral et al., 2018). Nos résultats ont permis de mettre en lumière que même sans conservation de séquence, le phénomène synténique peut prédire la fonctionnalité d'un lncARN.

D'après nos connaissances, ces résultats sont également le premier exemple de conservation de fonction d'un gène indépendamment de la conservation de sa séquence chez les vertébrés. Il est fort probable que la fonctionnalité d'un lncARN vienne de ses structures secondaires et de ses capacités à interagir avec différents partenaires protéiques, et ce indépendamment de sa séquence (comme suggéré dans Dominguez et al., 2018). Une prochaine étape à la caractérisation des lncARN décrits ci-dessus sera donc l'identification de leurs partenaires protéiques afin de décrire précisément leur mécanisme moléculaire.

APPENDIX

« Strategies for Genetic Inactivation of Long Noncoding RNAs in Zebrafish »

Manuscript in submission in RNA

Strategies for Genetic Inactivation of Long Noncoding RNAs in Zebrafish

Perrine Lavalou[1], Helene Eckert[1], Sara Majello[1], Angelo Bitetti[1], Antoine Graindorge[1], Alena Shkumatava[1,*]

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

[*]Corresponding author
E-mail: alena.shkumatava@curie.fr (AS)

ABSTRACT

The number of annotated long noncoding RNAs (lncRNAs) continues to grow, however their functional characterization in model organisms has been hampered by the lack of reliable genetic inactivation strategies. While partial or full deletions of lncRNA loci disrupt lncRNA expression, they do not permit the formal association of a phenotype with the encoded transcript. Here, we examined several alternative strategies for generating lncRNA null alleles in zebrafish and found that they often resulted in unpredicted changes to lncRNA expression. Removal of the transcriptional start sites (TSSs) of lncRNA genes resulted in hypomorphic mutants due to the usage of either constitutive or tissue-specific alternative TSSs. Deletions of short, deeply conserved lncRNA regions can also lead to overexpression of truncated transcripts. By contrast, a knock-in of a polyadenylation signal enabled complete inactivation of *malat1*, the most abundant vertebrate lncRNA. In summary, lncRNA null alleles require extensive *in vivo* validation and we propose insertion of transcription termination sequences as the most reliable approach to generate lncRNA-deficient zebrafish.

INTRODUCTION

Thousands of lncRNAs have been identified in multiple vertebrate species (Hezroni et al., 2015; Necsulea et al., 2014), but their biological functions remain mostly unknown. To study lncRNAs *in vivo*, genetic mutants have been generated in model animals, primarily using a mouse model, but have also more recently been reported in zebrafish (Amandio et al., 2016; Anderson et al., 2016; Bitetti et al., 2018; Bond et al., 2009; Eissmann et al., 2012; Goff and Rinn, 2015; Grote et al., 2013; Han et al., 2014; Han et al., 2018; Hosono et al., 2017; Ip et al., 2016; Isoda et al., 2017; Kleaveland et al., 2018; Kok et al., 2015; Kotzin et al., 2016; Lai et al., 2015; Leighton et al., 1995; Li et al., 2013; Marahrens et al., 1997; Nakagawa et al., 2012; Nakagawa et al., 2014; Ripoche et al., 1997; Sado et al., 2001; Sauvageau et al., 2013; Sleutels et al., 2002; Zhang et al., 2012).

Genetic inactivation of lncRNAs is less straight-forward than of coding genes, where deletion of an exon or a point mutation in the open reading frame (ORF) often leads to stop codons or frame-shift mutations and a subsequent loss of function. Several complementary strategies have been implemented to achieve genetic loss of lncRNA function including full or partial deletion of the lncRNA locus, deletion and replacement of the lncRNA locus by a reporter gene (Nakagawa et al., 2012; Sauvageau et al., 2013), deletion of lncRNA transcriptional start site (TSS) and upstream regulatory regions (Fitzpatrick et al., 2002; Zhang et al., 2012) and sequence inversions (Bitetti et al., 2018) (Fig.1). Although commonly used, these lncRNA inactivation strategies have several caveats and limitations. Full deletions of lncRNA loci, which often span several kilo bases, or lncRNA replacement by a reporter gene are invasive and might lead to phenotypes that are caused by removal of regulatory DNA motifs. Deletions of lncRNA TSS and upstream promoter regions may result in usage of alternative TSSs or cryptic promoters and/or impact the expression of neighboring genes. A less invasive and more accurate approach is to inactivate lncRNAs by integrating of a premature polyadenylation (polyA) cassette. This strategy has been successfully implemented in several recent mouse lncRNA studies (Anderson et al., 2016; Ballarino et al., 2018; Grote et al., 2013) (Fig. 1). Whereas lncRNA locus

deletion and/or partial lncRNA gene inversion strategies have been applied in zebrafish to genetically inactivate lncRNAs (Bitetti et al., 2018; Hosono et al., 2017; Kok et al., 2015), analyses of complementary lncRNA silencing approaches including the minimally invasive insertion of the polyA sequences have not yet been carried out.

Here, we examined the efficiency of several strategies for CRISPR-Cas9-mediated inactivation of lncRNAs in zebrafish. Careful evaluation of lncRNA zebrafish mutants demonstrated that caution is required when analyzing each individual mutant allele. When genetically manipulating lncRNA loci, we found that usage of constitutive or tissue-specific alternative TSSs, overexpression or destabilization of truncated lncRNA transcripts commonly take place *in vivo,* minimizing or confounding the effect of the intended genetic intervention. By contrast, using our improved zebrafish knock-in protocol, a minimally invasive insertion of a premature polyadenylation signal into the *malat1* locus diminished *malat1* transcripts to undetectable levels, effectively establishing a *malat1* null allele in zebrafish.

RESULTS

Deletion of the conserved region of the lncRNA *cyrano* leads to overexpression of the truncated transcript

A small fraction of zebrafish lncRNAs are conserved to mammals representing a promising set of candidates for functional interrogations (Hezroni et al., 2015; Ulitsky et al., 2011). The conserved regions of lncRNAs are usually relatively short, ranging between 50-300 nucleotides (nt) (Hezroni et al., 2015; Ulitsky et al., 2011) and can be efficiently targeted for CRISPR-Cas9-mediated deletions in zebrafish, offering a relatively minimally invasive strategy for functional inactivation (Fig. 1). To examine the effect of this strategy on lncRNA expression, the deeply conserved lncRNA *cyrano* (Ulitsky et al., 2011) was chosen for genetic interrogations in zebrafish. We generated a ~280 base pair (bp) deletion of the most conserved region of the 5.5 kb long *cyrano*, hereafter referred as *cyrano*$^{\Delta CR}$ (Ulitsky et al., 2011) (Fig. 2A, B).

Interestingly, we detected elevated levels of the residual truncated transcript in homozygous *cyrano*$^{\Delta CR}$ zebrafish embryos across *cyrano*$^{\Delta CR}$ adult tissues apart from the brain (Fig. 2C, D; Supplementary Fig. 1A). These results suggest that removal of a relatively small region of a lncRNA may have an unexpected effect on the transcript levels, potentially leading to its unintended overexpression.

TSS deletion of the *cyrano* locus results in hypomorphic zebrafish mutants

Next, we tested if deleting the sequences surrounding and containing lncRNA TSS elements is a reliable alternative strategy for zebrafish lncRNA genetic inactivation. To this end, we generated a minimally invasive *cyrano*$^{\Delta TSS}$ mutant allele by removing sequences containing the *cyrano* TSS (0 to +84) (Fig. 2E). Although the levels of the *cyrano* transcript were reduced in *cyrano*$^{\Delta TSS}$ fish, the transcript was still robustly detectable by qRT-PCR and RNA blot analysis, resulting in a hypomorphic *cyrano*$^{\Delta TSS}$ mutant (Fig. 2F, G). The 5' RACE analysis demonstrated that in the absence of the two main TSSs usually used in wild type animals, an upstream alternative TSS maintains *cyrano* expression in *cyrano*$^{\Delta TSS}$ mutant zebrafish (Supplementary Fig. 1B-D).

Notably, neither the *cyrano*$^{\Delta CR}$ mutant, which removed the deeply conserved miR-7 site (Ulitsky et al., 2011) nor the *cyrano*$^{\Delta TSS}$ mutant fish exhibited obvious morphological defects. This observation is consistent with the recent mouse study (Kleaveland et al., 2018) and is in contrast to previous studies that used a morpholino-based knockdown approach to inactivate *cyrano* (Sarangdhar et al., 2018; Ulitsky et al., 2011).

lncRNA TSS removal leads to tissue-specific alternative TSS usage maintaining lncRNA expression

To test if the usage of alternative TSSs is a prevalent cellular mechanism to maintain lncRNA gene expression, we examined the effect of TSS deletions

on additional lncRNAs in zebrafish. We have generated a lnc-*sox4*$^{\Delta TSS}$ mutant allele by removing ~200bp surrounding the lnc-*sox4* TSS (-43 to +157) (Fig. 3A, B). lnc-*sox4* (Ulitsky et al., 2011) that is highly expressed in the zebrafish ovary was successfully abolished in embryos, in the ovary and across other tested adult tissues in lnc-*sox4*$^{\Delta TSS}$ zebrafish (Fig. 3C, D). However, robust expression of lnc-*sox4* was detected in the lnc-*sox4*$^{\Delta TSS}$ brain (Fig. 3D). The 5'RACE analysis confirmed that a tissue-specific alternative TSS, located in an intron 70kb downstream of the main TSS (Fig. 3B; Supplementary Fig. 2A, B), was used in the lnc-*sox4*$^{\Delta TSS}$ animals and maintained lncRNA expression specifically in the adult brain (Fig. 3D). Interestingly, while homozygous lnc-*sox4*$^{\Delta TSS}$ fish were viable and fertile, our alternative strategy to eliminate lnc-*sox4* expression by deleting the last exon failed to generate homozygous embryos (Supplementary Fig. 2 C, D).

Moreover, we generated an additional lncRNA TSS mutant by removing ~389 bp surrounding the lnc-*pou2af1* TSS (-74 to +315) (Fig. 4A, B). Similar to the lnc-*sox4*$^{\Delta TSS}$ allele, the level of lnc-*pou2af1* was efficiently abolished in lnc-*pou2af1*$^{\Delta TSS}$ embryos and in a set of tested lnc-*pou2af1*$^{\Delta TSS}$ adult tissues (Fig. 4C, D). However, in the skin, in which lnc-*pou2af1* is highly expressed, kidney, intestine and testis, the expression of lnc-*pou2af1* was robustly detected by qRT-PCR (Fig. 4E). The 5'RACE analysis showed that several alternative TSSs, located ~1kb upstream of the main TSS, were used in the lnc-*pou2af1*$^{\Delta TSS}$ animals in a tissue-specific manner (Fig. 4B; Supplementary Fig. 3A, B). Interestingly, expression of lnc-*pou2af1* from alternative TSSs generated new tissue-specific lncRNA exons at the 5' of the transcript (Fig. 4B; Supplementary Fig. 3B, C).

Together, our data showed that in the absence of the main TSS, alternative TSSs can be used in a tissue-specific manner, generating hypomorphic mutants and minimizing the effect of the intended gene inactivation.

Insertion of the polyadenylation signal resulted in a *malat1* null allele in zebrafish

Having noted that usage of alternative TSSs may be a common cellular mechanism to confer lncRNA expression, we tested if a knock-in of a polyA signal into a lncRNA locus can be applied in zebrafish as a minimally invasive alternative to generate lncRNA null alleles. This approach has been successfully used to inactivate lncRNAs in mice (Anderson et al., 2016; Ballarino et al., 2018; Grote et al., 2013; Isoda et al., 2017). The *malat1* locus produces one of the most abundant lncRNAs in vertebrate genomes (Hezroni et al., 2015; Ulitsky et al., 2011). Because *malat1* is a mono-exonic lncRNA of ~7,5kb and its locus contains multiple TSSs and clustered enhancers forming a so-called super-enhancer (Perez-Rico et al., 2017), any deletion strategy of the locus including TSS removal has a strong potential to affect *cis* regulatory elements (Fig. 5A). Therefore, we used our improved protocol for the efficient targeted knock-in to insert a 131bp SV40 polyA signal into the *malat1* locus of zebrafish (Fig. 5B; see Material and Methods for details). The targeted knock-in of the polyA sequence completely abolished *malat1* expression in zebrafish embryos and in all examined adult tissues (Fig. 5C, D). Despite efficient inactivation of *malat1*, *malat1*^{polyA} zebrafish were viable and fertile and displayed no obvious morphological defects. The lack of overall morphological abnormalities is consistent with previously reported *Malat1*$^{-/-}$ mice (Eissmann et al., 2012; Nakagawa et al., 2012; Zhang et al., 2012) and is in contrast to the morpholino-based *malat1* inactivation in zebrafish (Wu et al., 2018). Taken together, compared to the lncRNA deletion strategies, polyA signal insertion was the most efficient and least invasive approach in zebrafish.

DISCUSSION

The identification of lncRNAs in model vertebrates, their comparative genomics analyses and recent progress in the genome editing technologies lead to generation of multiple mutant lncRNA alleles in vertebrate animals. Because common strategies for genetic inactivation of lncRNAs often do not allow to distinguish between the functions mediated by the lncRNA transcript

and those mediated by overlapping DNA regulatory motifs, the generation and interpretation of lncRNA null alleles is not always trivial. Here, we compared zebrafish lncRNA mutant alleles generated using several alternative and commonly applied CRISPR-Cas9 strategies for lncRNA inactivation.

We demonstrated that relatively small deletions of lncRNA conserved regions, representing attractive target sequences to eliminate or diminish lncRNA functions (Bitetti et al., 2018; Kleaveland et al., 2018), can result in unexpected changes in lncRNA expression such as the overaccumulation of the remaining transcript. Given our poor understanding of lncRNA functional domains and their cellular mechanisms of action, ectopic expression of partial lncRNA transcripts confounds any mechanistic or functional interpretation.

We demonstrated that relatively small deletions of conserved regions of lncRNAs, which represent attractive target sequences to eliminate or diminish lncRNA functions (Bitetti et al., 2018; Kleaveland et al., 2018), might result in unexpected changes in lncRNA levels, such as overexpression of the remaining transcript. One possibility is that deletion of the conserved region of *cyrano*, which removed a highly conserved and extensively paired site to miR-7 (Ulitsky et al., 2011), stabilized the *cyrano* RNA in zebrafish. Alternatively, deletion of this region of zebrafish *cyrano* might have caused transcriptional upregulation. For example, if deletion of this region abrogated *cyrano* function, cells might have boosted transcription of the locus in a futile attempt to restore *cyrano* activity. Interestingly, deletion of the conserved region of mouse *cyrano* does not lead to increased lncRNA levels (Kleaveland et al., 2018), which suggests that *cyrano* regulation has diverged between fish and mammals. A better understanding of *cyrano* regulation and function will help identify the source of this ectopic effect on the remaining lncRNA transcript observed in fish and how this effect might complicate interpretation of the deletion results.

Moreover, we showed that the removal of TSS and upstream regulatory regions, a commonly used approach considered to be straightforward to interpret, can result in the presence of either constitutive or tissue-specific

alternative TSSs that preclude efficient inactivation of lncRNAs and result in hypomorph mutant animals. Although not shown in this study, usage of temporal-specific alternative TSSs might also contribute to the maintenance of lncRNA expression at specific developmental stages, complicating the analysis and interpretation of TSS mutant alleles in animal models.

Importantly, our improved protocol for efficient targeted knock-in in zebrafish enabled examination of the effect of a polyA signal insertion into the most abundant and enhancer-dense lncRNA locus. We demonstrate that this minimally invasive genome editing strategy, previously shown to be successful for lncRNA inactivation in mice (Anderson et al., 2016; Ballarino et al., 2018; Grote et al., 2013; Isoda et al., 2017), is a highly effective strategy in zebrafish. Given the ease of our improved knock-in protocol, which combines a combination of a single-strand oligo as a template for homologous recombination and the inhibition of the non-homologous end joining, we anticipate that the insertion of a polyA sequence will become a widespread strategy for generating lncRNA mutant alleles in zebrafish.

Taken together, evaluation of several independent lncRNA mutant alleles in zebrafish indicates that a combination of complementary lncRNA inactivation approaches and their careful analyses are required for robust and accurate lncRNA functional analyses.

FIGURES

Figure 1: Strategies for genetic inactivation of lncRNAs in animals

Figure 2: Genetic perturbations of the lncRNA *cyrano* in zebrafish result in overexpression and hypomorphic alleles

(A) Gene architecture of the lncRNA *cyrano*. Shown are the corresponding CAGE (Haberle et al., 2014; Nepal et al., 2013), H3K4me3 ChIP-Seq (Ulitsky et al., 2011) and RNA-seq tracks from wild type zebrafish. Vertebrate conservation plots based on the 8-genome alignment indicate the location of conserved sequences. (B) The *cyrano*$^{\Delta CR}$ mutant allele showing the deletion of the most conserved region of the transcript (dotted, blue line) in zebrafish. Position of the qRT-PCR product is indicated. (C) *cyrano* expression in wild type and homozygous *cyrano*$^{\Delta CR}$ embryos detected by qRT-PCR at 2 hours post fertilization (hpf), 24 hpf and 72 hpf. *eef1α1* was used as a reference gene. (D) *cyrano* expression across wild type and homozygous *cyrano*$^{\Delta CR}$ adult tissues detected by qRT-PCR. *eef1α1* was used as a reference gene. (E) The *cyrano*$^{\Delta TSS}$ zebrafish allele showing deletion of the sequence around the TSS (dotted, blue line). Indicated are positions of the 5' RACE primer, qPCR product, RNA blot probe and alternative TSS. (F) *cyrano* expression in 72 hpf wild type and homozygous *cyrano*$^{\Delta TSS}$ embryos detected by an RNA blot. *18S* rRNA was used as a reference gene. (G) *cyrano* expression in 72 hpf wild type and homozygous *cyrano*$^{\Delta TSS}$ embryos detected by qRT-PCR. *eef1α1* was used as a reference gene.

Figure 3: Presence of a tissue-specific alternative TSS leads to a brain-specific rescue of lnc-*sox4* expression

(A) The lnc-*sox4* locus in zebrafish. Shown are the corresponding CAGE (Haberle et al., 2014; Nepal et al., 2013), H3K4me3 ChIP-Seq (Ulitsky et al., 2011) and RNA-seq tracks from wild type zebrafish. (B) The lnc-*sox4*$^{\Delta TSS}$ mutant allele showing deletion of the sequence around the TSS (dotted, blue line). Indicated are positions of the 5' RACE primers, qPCR primers and alternative TSS. (C) lnc-*sox4* expression in 24 hpf and 72 hpf wild type and homozygous lnc-*sox4*$^{\Delta TSS}$ embryos detected by qRT-PCR. *eef1α1* was used

as a reference gene. (D) lnc-*sox4* expression across adult wild type and homozygous lnc-*sox4*$^{\Delta TSS}$ zebrafish tissues detected by qRT-PCR. *eef1α1* was used as a reference gene.

Figure 4: Usage of tissue-specific alternative TSSs maintains lnc-*pou2af1* expression in a set of adult tissues

(A) The lnc-*pou2af1* locus in zebrafish. Shown are the corresponding CAGE (Haberle et al., 2014; Nepal et al., 2013), H3K4me3 ChIP-Seq (Ulitsky et al., 2011) and RNA-seq tracks from wild type zebrafish. (B) The lnc-*pou2af1*$^{\Delta TSS}$ mutant allele showing deletion of the sequence around the TSS (dotted, blue line). Indicated are positions of the 5' RACE and qPCR primers and alternative TSS. (C) lnc-*pou2af1* expression in 24 hpf wild type and homozygous lnc-*pou2af1*$^{\Delta TSS}$ embryos detected by qRT-PCR. *eef1α1* was used as a reference gene. (D), (E) lnc-*pou2af1* expression across a set of adult wild type and homozygous lnc-*pou2af1*$^{\Delta TSS}$ zebrafish tissues detected by qRT-PCR. *eef1α1* was used as a reference gene.

Fig. 5: Efficient inactivation of the lncRNA *malat1* in zebrafish by insertion of the premature polyadenylation signal

(A) The *malat1* locus in zebrafish. Shown are the corresponding CAGE (Haberle et al., 2014; Nepal et al., 2013), H3K4me3 ChIP-Seq (Ulitsky et al., 2011), RNA-seq and H3K27ac ChIP-Seq (Perez-Rico et al., 2017) tracks from wild type zebrafish. (B) Generation of the *malat1*^{polyA} allele by a targeted knock-in of the polyA signal. The hybridization site of the RNA-blot probe is indicated as a grey box. ha, homologous arms. (C) *malat1* expression in wild type (WT) and homozygous *malat1*^{polyA} embryos detected by an RNA blot. 18S rRNA was used as a loading reference. hpf, hours post fertilization. (D) *malat1* expression across wild type and homozygous *malat1*^{polyA} adult zebrafish tissues detected by an RNA blot. 18S rRNA was used as a loading reference.

Supplementary Fig. 1

(A) *cyrano* expression in the wild type and homozygous *cyrano*$^{\Delta CR}$ brain detected by qRT-PCR. *eef1α1* was used as a reference gene. (B) DNA

fragments amplified by 5' RACE in 72 hpf wild type and homozygous *cyrano*$^{\Delta TSS}$ embryos. (C) The sequence of the *cyrano* wild type locus in zebrafish. The nucleotides defining the borders of the first exon of *cyrano* are shown in blue and underlined. Positions of two main TSS (WT TSS1 and WT TSS2) are indicated with arrows. The numbers in brackets show the frequency of the TSS usage in wild type animals. (D) The sequence of the zebrafish *cyrano*$^{\Delta TSS}$ allele. The nucleotides defining the borders of the first exon of *cyrano* are shown in red and underlined. The deletion is shown as a dotted line. The position of the alternative TSS is indicated with a red arrow. The numbers in the brackets show the frequency of the alternative TSS usage in *cyrano*$^{\Delta TSS}$ embryos.

Supplementary Fig. 2

(A) DNA fragments amplified by 5' RACE in the wild type and homozygous lnc-*sox4*$^{\Delta TSS}$ brain. (B) The sequence of the zebrafish lnc-*sox4*$^{\Delta TSS}$ allele. The nucleotides defining the borders of the tissue-specific exon expressed in the lnc-*sox4*$^{\Delta TSS}$ brain are shown in blue and underlined. The position of the alternative TSS is indicated with a red arrow. The intron sequences are indicated in black. (C) The lnc-*sox4*$^{\Delta 3'exon}$ mutant allele showing deletion of the sequence at the 3'end of the gene (dotted, blue line). (D) Number of identified lnc-*sox4*$^{\Delta 3'exon}$ fish.

Supplementary Fig. 3

(A) DNA fragments amplified by 5' RACE in the wild type and homozygous lnc-*pou2af1*$^{\Delta TSS}$ skin. (B) The lnc-*pou2af1*$^{\Delta TSS}$ mutant allele is shown on the top. Deletion of the sequence is shown as dotted, blue line. lnc-*pou2af1* skin-specific expression from alternative TSSs (red arrows) results in multiple alternative isoforms. (C) The sequence of the zebrafish lnc-*pou2af1*$^{\Delta TSS}$ allele. The nucleotides defining the borders of the tissue-specific exon expressed in the lnc-*pou2af1*$^{\Delta TSS}$ skin are shown in red and underlined. The positions of the alternative TSSs are indicated with red arrows. The intron sequences are indicated in black. The deletion is shown as a dotted line.

MATERIALS AND METHODS

Generation of lncRNA mutant alleles in zebrafish

All lncRNA mutant alleles were generated using CRISPR-Cas9 mediated genome editing. To generate lnc-*sox4*$^{\Delta TSS}$, lnc-*pou2af1*$^{\Delta TSS}$, *cyrano*$^{\Delta TSS}$ and *cyrano*$^{\Delta CR}$ alleles, two sgRNAs (9ng each, Supplementary Table 1) and 150ng *in vitro* transcribed *Cas9* mRNA were co-injected into the one-cell stage AB zebrafish embryos (Hwang et al., 2013). To generate lnc-*sox4a*$^{\Delta 3'exon}$ allele, two sgRNAs (100 ng each, Supplementary Table 1) and Cas9 protein (50ng/µL, a gift of the Concordet lab, Muséum d'Histoire Naturelle, Paris) were co-injected into the one-cell stage AB zebrafish embryos (Hwang et al., 2013). sgRNAs and *Cas9* mRNA were generated as described previously (Hwang et al., 2013), using the codon-optimized plasmid JDS246 for the *Cas9* mRNA synthesis (Addgene #43861), purified with RNeasyMini Kit (Qiagen). Genomic DNA was extracted as described previously (Bitetti et al., 2018) and used for genotyping by PCR, DNA sequencing and mapping of genetic amplification product. The genotyping primers are listed in Supplementary Table 2.

All zebrafish were bred and maintained at the 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 the 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).

Generation of the *malat1*^{polyA} allele by CRISPR/Cas9–mediated homologous recombination in zebrafish

Zebrafish *malat1*^{polyA} mutant was generated by insertion of a single SV40 polyA signal (131bp) into the *malat1* locus. Briefly, one cell stage embryos were injected with a single guide RNA (100ng, Supplementary Table 1), Cas9 protein (50ng/µL, a gift of the Concordet lab, Muséum d'Histoire Naturelle,

Paris), a morpholino against *xrcc4* to suppress NHEJ (Non Homologous End Joining) (3ng/µL, Gene Tools LLC, Supplementary Table 1) and a 191nt single strand DNA oligo with 30bp homology arms flanking both sides of the SV40 polyA sequence (200ng, designed and manufactured by Ultramer IDT, Supplementary Table 1). Genomic DNA was extracted as described previously (Bitetti et al., 2018) and polyA insertion was detected by PCR using primers listed in Supplementary Table 2, DNA sequencing and mapping of genetic amplification product.

qRT-PCR

Total RNA was isolated from zebrafish embryos and adult tissues by TRIzol extraction (Invitrogen) according to manufacturer's instruction, followed by DNAse treatment (TURBO DNA-free Ambion) and ethanol precipitated. cDNA was obtained by retro-transcription of 500ng total RNA (heart and spleen) or 1µg total RNA (other tissues) with SuperScript IV reverse transcriptase (Invitrogen) using a mix of oligo-dT and Random Hexamer (lnc-*sox4a*) or oligo-dT only (*cyrano*) and amplified using PowerUp SYBR green (ThermoFisher Scientific) according to manufacturer's instruction (primers are reported in Supplementary Table 3).

RNA blots

Total RNA was isolated using TRIzol (Invitrogen) according to manufacturer's instructions, separated on 1% agarose gels containing 0.8% formaldehyde and transferred to nylon membrane (Nytran SPC, GE Healthcare) by capillary action. Blots were hybridized with α-UTP ^{32}P-labeled RNA probes at 68°c in ULTRAhyb buffer (Ambion) as recommended by the manufacturer. RNA probe template was amplified from zebrafish brain cDNA by PCR using the primers listed in Supplementary Table 3 (the sequence of the T7 promoter is underlined) and *in vitro* transcribed (RNA Maxiscript, Ambion) in the presence of α-UTP^{32}P.

RNA ligase-mediated and oligo-capping Rapid Amplification of cDNA Ends (5' RACE)

TSS usage was determined by Rapid Amplification of cDNA Ends (RACE) according to manufacturer's instruction (GeneRacer^{TM} kit, life technology). Gene specific primers listed in Supplementary Table 3 were used to amplify lncRNA 5' RACE products through PCR and nested PCR, subcloned into the PCR BLUNT II TOPO vector (Invitrogen) and transformed in the NEB TOP-10 cells. Minimum 12 colonies were sequenced and the sequences were aligned to the corresponding lncRNA genomic locus.

Supplementary tables:

Table 1: guide RNAs, ssDNA oligos, morpholinos

sgRNA	**Sequence 5'-3'**
sgRNA *cyrano*$^{\Delta CR}$ 1	GTATAGTAGTTCCTATCATA
sgRNA *cyrano*$^{\Delta CR}$ 2	CATCATTTAATGGAAGACAT
sgRNA *cyrano*$^{\Delta TSS}$ 1	GGTCCCGTGTGCTGCTACTG
sgRNA *cyrano*$^{\Delta TSS}$ 2	GTTGTTCGGGCCAGGCTCTG
sgRNA lnc-*sox4*$^{\Delta TSS}$ 1	GCATGATATCGGACAAGGGG
sgRNA lnc-*sox4*$^{\Delta TSS}$ 2	GACATCCTGACGTAGGTAAA
sgRNA lnc-*sox4*$^{\Delta 3'exon}$ 1	GTAAGATCATGGGTCCATAC
sgRNA lnc-*sox4*$^{\Delta 3'exon}$ 2	GTGAAGCGACTGACGCTGGT
sgRNA lnc-*pou2af1*$^{\Delta TSS}$ 1	GTGTGCGACCCGGCAGTGAA
sgRNA lnc-*pou2af1*$^{\Delta TSS}$ 2	GAGCGGAATGCGCAGAAAGT
sgRNA *malat1*^{polyA}	GGTGAGGCGCTATGGAAGGC
ssDNA oligos	**Sequence 5'-3'**
ssDNA *malat1*^{polyA}	AACATTGTGCGTCACGACGGGGTGAGGCGCA CTTGTTTATTGCAGCTTAAATGGTTACAAATAA AGCAATAGCATCACAAATTTCACAAATAAAGA TTTTTTTCACTGCATTCTAGTTGTGGTTTGTCC AAACTCATCAATGTATCTTATCATGTCTGTTAT GGAAGGCAGGGAGGCTTCGTTGATCTG
Morpholinos	**Sequence 5'-3'**
xrcc4 MO	CACTACTGCTGCGACACCTCATTCC

Table 2: Genotyping primers

Primer names	**Sequence 5'-3'**
cyrano$^{\Delta CR}$ Fwd	ACCACAGCTCAGAGCCACACTTGGA
cyrano$^{\Delta CR}$ Rev	CTGGCACTACAAATCCCGCCACCCT

cyrano$^{\Delta TSS}$ Fwd	CAATTTGTTCTCTTCAATTTTACCCTCGTCC
cyrano$^{\Delta TSS}$ Rev	TGGAGTGAAGAGTAATTTTCAACAAATTTG
lnc-*sox4*$^{\Delta TSS}$ Fwd	TAAGGAGCACAAATGTCTTAATACCTCAGG
lnc-*sox4*$^{\Delta TSS}$ Rev	GTTTTCTCTATATGCCGACTGTTTTGATCC
lnc-*sox4*$^{\Delta 3'exon}$ Fwd	TAAGGAGCACAAATGTCTTAATACCTCAGG
lnc-*sox4*$^{\Delta 3'exon}$ Rev	GTTTTCTCTATATGCCGACTGTTTTGATCC
lnc-*pou2af1*$^{\Delta TSS}$ Fwd	GGGCTACAAATATCAGTGAAACTG
lnc-*pou2af1*$^{\Delta TSS}$ Rev	CATTTACCAACGCTCTAGCTG
malat1^{polyA} flanking Fwd	GTGTGGTATGTTGTGTCAAG
malat1^{polyA} flanking Rev	CCGCCATTTTGTAAATTCTTTCTAGCGTCGAG
malat1^{polyA} insert Fwd	TCACTGCATTCTAGTTGTGGTTTGTCC
malat1^{polyA} insert Rev	GCTTGTATTTTATCTTCGTCACGCTTGC

Table 3: qPCR primers and primers to amplify RNA blot probes

qPCR primers	**Sequence 5'-3'**
qPCR *cyrano* Fwd	AAACCTTTCTAGCGGGGTGC
qPCR *cyrano* Rev	TGATCCAAGTGTGGCTCTGAG
qPCR lnc-*sox4* Fwd	CATCACTCACAGTTCAGCTCTCC
qPCR lnc-*sox4* Rev	GAACACGACTATCCTCCACACTC
qPCR lnc-*pou2af1* Fwd	CCTAAATCTCTAGGTATCGTTCAACTGGG
qPCR lnc-*pou2af1* Rev	GCAAATGATATGAGTAAGCATTGCGTGAC
qPCR *eef1α1* Fwd	CAGCATTATCCAGTCCTTAAGTAGAGTGC
qPCR *eef1α1* Rev	GCGTCATCAAGAGCGTTGAGAAG
RNA blot primers	**Sequence 5'-3'**
NB *cyrano* Fwd	GGCTCAGTAGCTTAGAATACGCAGG
NB *cyrano* Rev	TAATACGACTCACTATAGGGCTTCATGAGGATA ATGAGTCATCAG
NB *malat1* Fwd	GGGTGTAAAGCGCCGCTACC
NB *malat1* Rev	GCGGTAATACGACTCACTATAGGGCTTGCATTC AGGTGACGTGATCC

5' RACE primers	**Sequence 5'-3'**
5' RACE *cyrano*	AACAATATGACCAGTCGATGGCACC
5' RACE *cyrano* nested	ACACAAGAAGAGTTTGTGGGGGAGT
5' RACE lnc-*sox4*	ATCAGCCTTAGGTTACAGGAAGAGAGCC
5' RACE lnc-*sox4* nested 1	CCAGAACACGACTATCCTCCACACTCGG
5' RACE lnc-*sox4* nested 2	TTCCCACGCTGAAGGCTGATACTGAGAG
5' RACE lnc-*pou2af1*	CATGGCTGCTATCATATGCCACGCCCACTAC
5' RACE lnc-*pou2af1* nested	GTGTTTCTAAGGGAGACTCCAACCACCAAGTCC

AUTHOR CONTRIBUTIONS

P.L. developed the protocol for the targeted knock-in in zebrafish and contributed to the design, generation and analysis of the *lnc-sox4a*$^{\Delta TSS}$, *lnc-sox4a*$^{\Delta 3'exon}$ and *malat1*^{polyA} alleles. H.E. contributed to the design, generation and analysis of the *lnc-pou2af1*$^{\Delta TSS}$ and the maintenance and analyses of *cyrano* and *malat1*^{polyA} alleles. S.M. contributed to the design, generation and analyses of the *cyrano* alleles. A.B. contributed to the design, generation and analyses of the *malat1*^{polyA} allele. A.G. contributed to the design and generation of the *cyrano* alleles. P.L. and A.S. wrote the final version of the manuscript. A.S. conceived and supervised the study.

ACKNOWLEDGEMENTS

We thank all members of the Shkumatava lab for useful discussions. This work was supported by grants from ERC (FLAME-337440), ATIP-Avenir, and La Fondation Bettencourt Schueller to A.S. and PSL and La Ligue Nationale Contre Le Cancer doctoral fellowships to P.L.

REFERENCES

Amandio, A.R., Necsulea, A., Joye, E., Mascrez, B., and Duboule, D. (2016). Hotair Is Dispensible for Mouse Development. PLoS Genet *12*, e1006232.

Anderson, K.M., Anderson, D.M., McAnally, J.R., Shelton, J.M., Bassel-Duby, R., and Olson, E.N. (2016). Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. Nature *539*, 433-436.

Ballarino, M., Cipriano, A., Tita, R., Santini, T., Desideri, F., Morlando, M., Colantoni, A., Carrieri, C., Nicoletti, C., Musaro, A.*, et al.* (2018). Deficiency in the nuclear long noncoding RNA Charme causes myogenic defects and heart remodeling in mice. EMBO J *37*.

Bitetti, A., Mallory, A.C., Golini, E., Carrieri, C., Carreno Gutierrez, H., Perlas, E., Perez-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.

Bond, A.M., Vangompel, M.J., Sametsky, E.A., Clark, M.F., Savage, J.C., Disterhoft, J.F., and Kohtz, J.D. (2009). Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. Nat Neurosci *12*, 1020-1027.

Eissmann, M., Gutschner, T., Hammerle, M., Gunther, S., Caudron-Herger, M., Gross, M., Schirmacher, P., Rippe, K., Braun, T., Zornig, M.*, et al.* (2012). Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. RNA Biol *9*, 1076-1087.

Fitzpatrick, G.V., Soloway, P.D., and Higgins, M.J. (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat Genet *32*, 426-431.

Goff, L.A., and Rinn, J.L. (2015). Linking RNA biology to lncRNAs. Genome Res *25*, 1456-1465.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Wahrisch, S., Beisaw, A., Macura, K., Blass, 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.

Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M.*, et al.* (2014). Two independent transcription initiation codes overlap on vertebrate core promoters. Nature *507*, 381-385.

Han, P., Li, W., Lin, C.H., Yang, J., Shang, C., Nuernberg, S.T., Jin, K.K., Xu, W., Lin, C.Y., Lin, C.J.*, et al.* (2014). A long noncoding RNA protects the heart from pathological hypertrophy. Nature *514*, 102-106.

Han, X., Luo, S., Peng, G., Lu, J.Y., Cui, G., Liu, L., Yan, P., Yin, Y., Liu, W., Wang, R.*, et al.* (2018). Mouse knockout models reveal largely dispensable but context-dependent functions of lncRNAs during development. J Mol Cell Biol *10*, 175-178.

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.

Hosono, Y., Niknafs, Y.S., Prensner, J.R., Iyer, M.K., Dhanasekaran, S.M., Mehra, R., Pitchiaya, S., Tien, J., Escara-Wilke, J., Poliakov, A.*, et al.* (2017). Oncogenic Role of THOR, a Conserved Cancer/Testis Long Non-coding RNA. Cell *171*, 1559-1572 e1520.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Kaini, P., Sander, J.D., Joung, J.K., Peterson, R.T., and Yeh, J.R. (2013). Heritable and precise zebrafish genome editing using a CRISPR-Cas system. PLoS One *8*, e68708.

Ip, J.Y., Sone, M., Nashiki, C., Pan, Q., Kitaichi, K., Yanaka, K., Abe, T., Takao, K., Miyakawa, T., Blencowe, B.J.*, et al.* (2016). Gomafu lncRNA knockout mice exhibit mild hyperactivity with enhanced responsiveness to the psychostimulant methamphetamine. Sci Rep *6*, 27204.

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.*, et al.* (2017). Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. Cell *171*, 103-119 e118.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev Dyn *203*, 253-310.

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.

Kok, F.O., Shin, M., Ni, C.W., Gupta, A., Grosse, A.S., van Impel, A., Kirchmaier, B.C., Peterson-Maduro, J., Kourkoulis, G., Male, I.*, et al.* (2015). Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. Dev Cell *32*, 97-108.

Kotzin, J.J., Spencer, S.P., McCright, S.J., Kumar, D.B.U., Collet, M.A., Mowel, W.K., Elliott, E.N., Uyar, A., Makiya, M.A., Dunagin, M.C.*, et al.* (2016). The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan. Nature *537*, 239-243.

Lai, K.M., Gong, G., Atanasio, A., Rojas, J., Quispe, J., Posca, J., White, D., Huang, M., Fedorova, D., Grant, C.*, et al.* (2015). Diverse Phenotypes and Specific Transcription Patterns in Twenty Mouse Lines with Ablated LincRNAs. PLoS One *10*, e0125522.

Leighton, P.A., Ingram, R.S., Eggenschwiler, J., Efstratiadis, A., and Tilghman, S.M. (1995). Disruption of imprinting caused by deletion of the H19 gene region in mice. Nature *375*, 34-39.

Li, L., Liu, B., Wapinski, O.L., Tsai, M.C., Qu, K., Zhang, J., Carlson, J.C., Lin, M., Fang, F., Gupta, R.A.*, et al.* (2013). Targeted disruption of Hotair leads to homeotic transformation and gene derepression. Cell Rep *5*, 3-12.

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.

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.

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.

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J.C., Grutzner, F., and Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature *505*, 635-640.

Nepal, C., Hadzhiev, Y., Previti, C., Haberle, V., Li, N., Takahashi, H., Suzuki, A.M., Sheng, Y., Abdelhamid, R.F., Anand, S.*, et al.* (2013). Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. Genome Res *23*, 1938-1950.

Perez-Rico, Y.A., Boeva, V., Mallory, A.C., Bitetti, A., Majello, S., Barillot, E., and Shkumatava, A. (2017). Comparative analyses of super-enhancers reveal conserved elements in vertebrate genomes. Genome Res *27*, 259-268.

Ripoche, M.A., Kress, C., Poirier, F., and Dandolo, L. (1997). Deletion of the H19 transcription unit reveals the existence of a putative imprinting control element. Genes Dev *11*, 1596-1604.

Sado, T., Wang, Z., Sasaki, H., and Li, E. (2001). Regulation of imprinted X-chromosome inactivation in mice by Tsix. Development *128*, 1275-1286.

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.

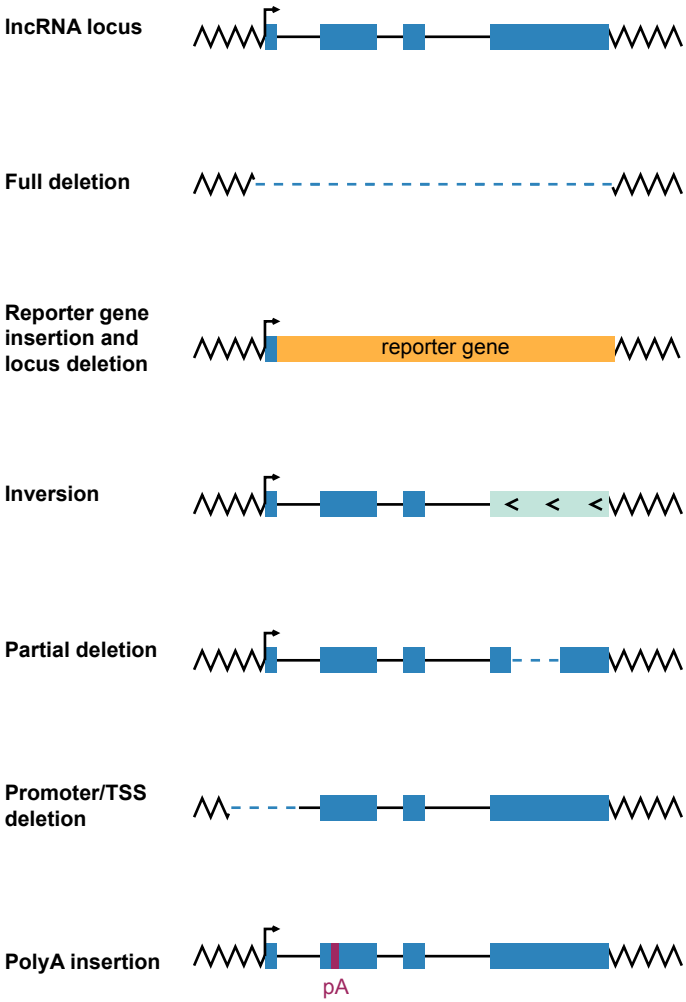
Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Hacisuleyman, E., Li, E., Spence, M.*, et al.* (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife *2*, e01749.

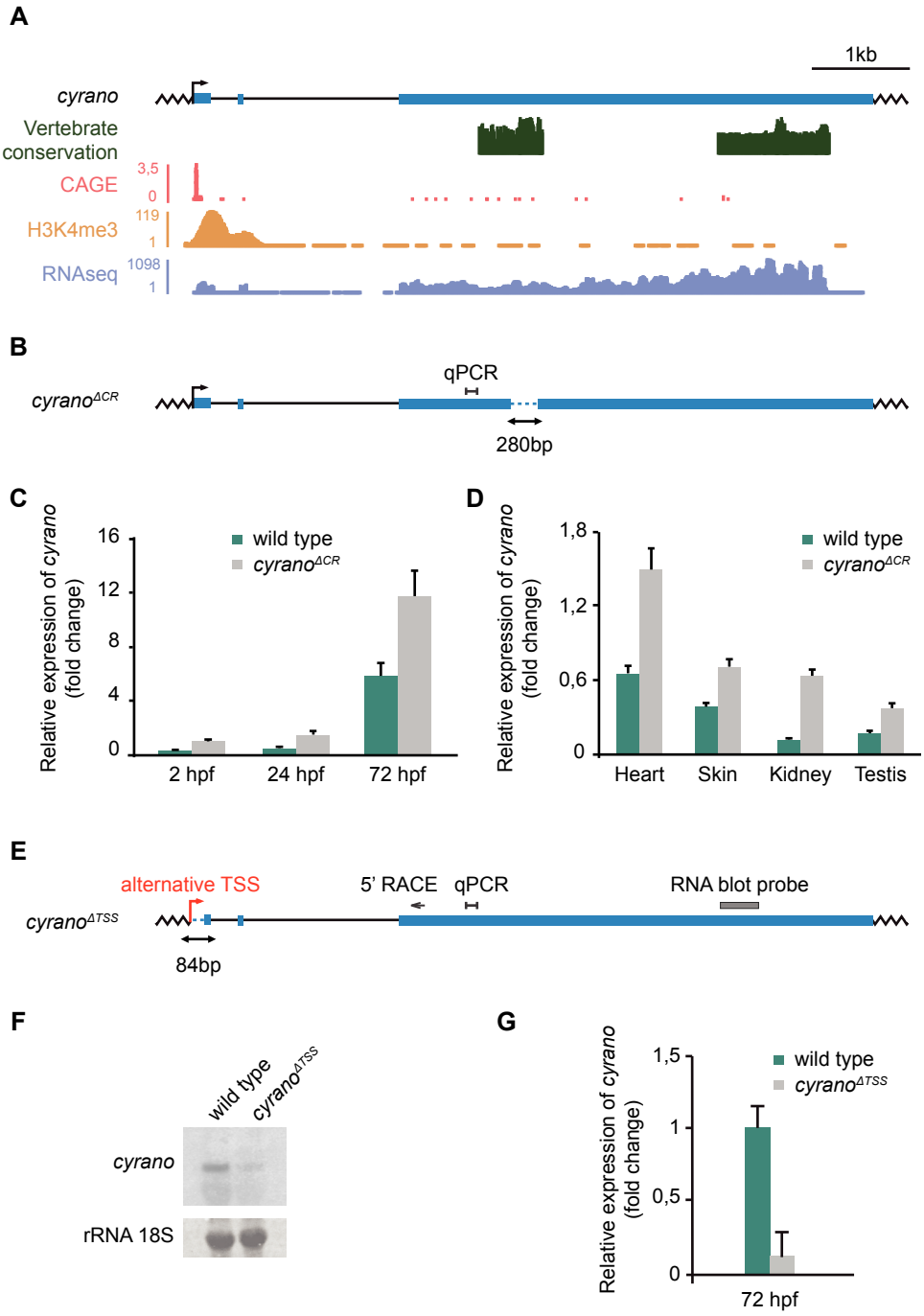
Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810-813.

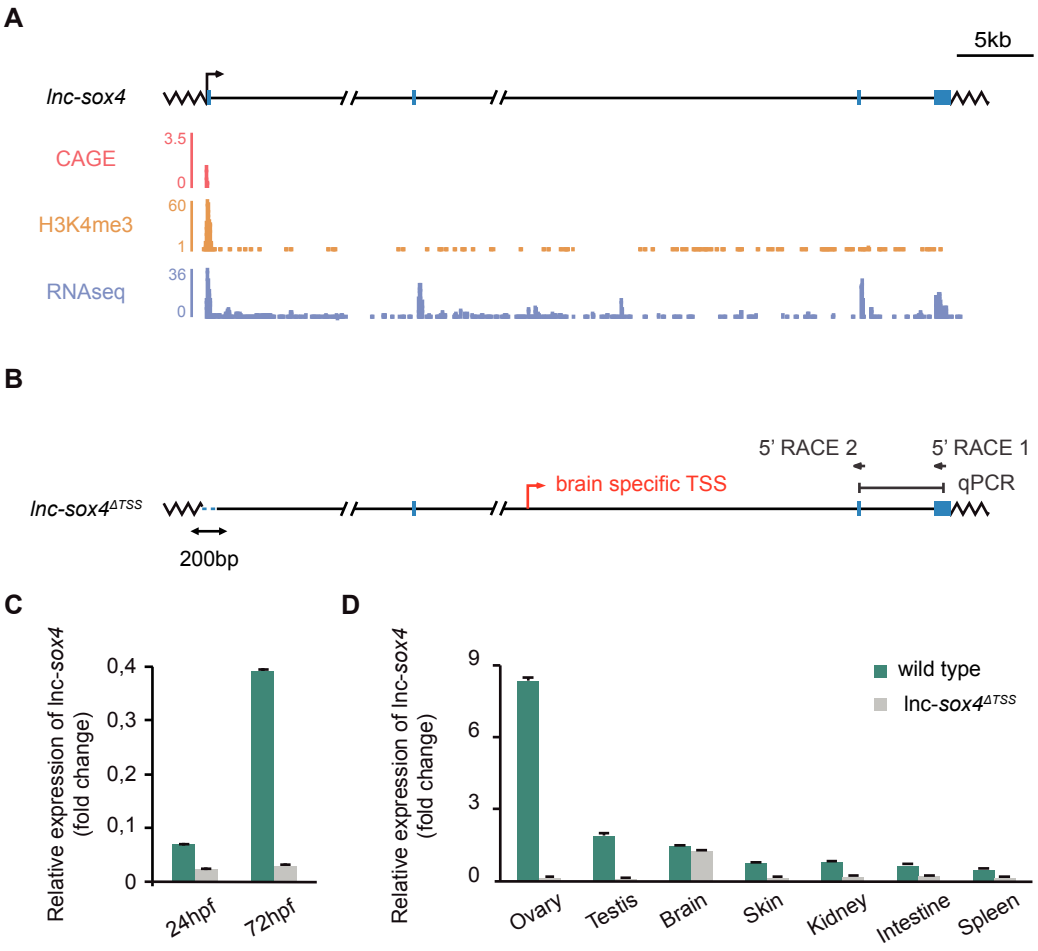
Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell *147*, 1537-1550.

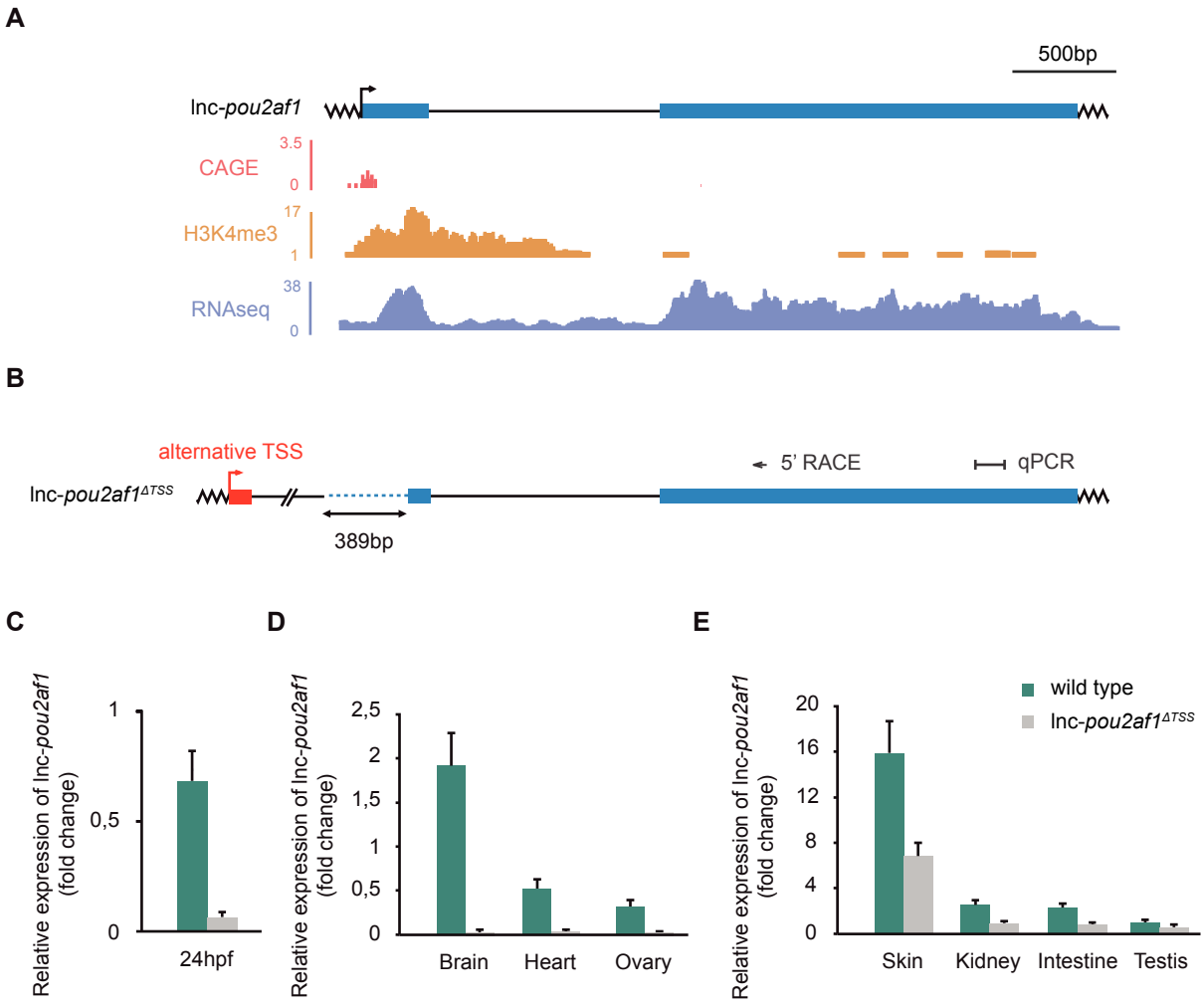
Wu, M., Zhang, S., Chen, X., Xu, H., and Li, X. (2018). Expression and function of lncRNA MALAT-1 in the embryonic development of zebrafish. Gene.

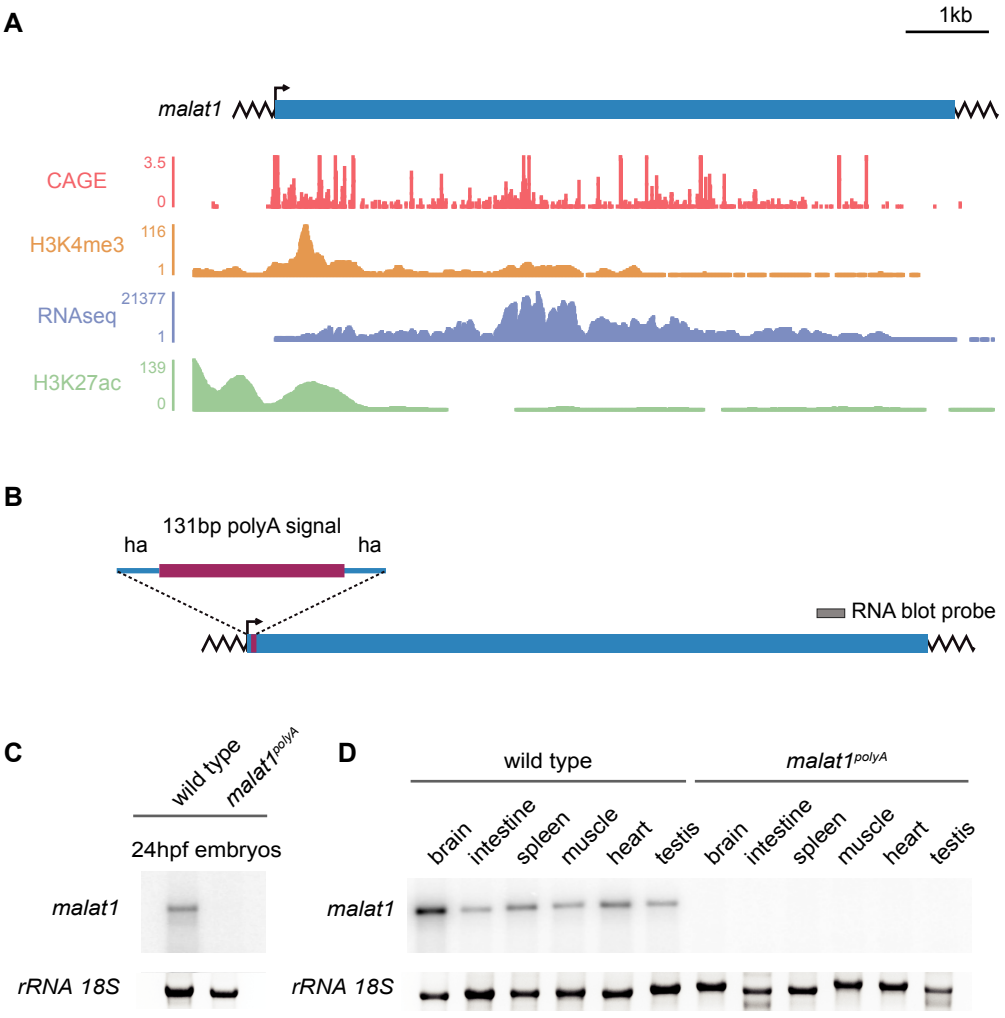
Zhang, B., Arun, G., Mao, Y.S., Lazar, Z., Hung, G., Bhattacharjee, G., Xiao, X., Booth, C.J., Wu, J., Zhang, C.*, et al.* (2012). The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. Cell Rep *2*, 111-123.


lncRNA locus
Full deletion
Reporter gene insertion and locus deletion
reporter gene
Inversion
Partial deletion
Promoter/TSS deletion
PolyA insertion
pA

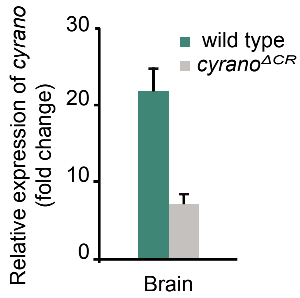

A
1kb
cyrano
Vertebrate conservation
CAGE
3,5
0
H3K4me3
119
1
RNAseq
1098
1
B
cyrano^ΔCR
qPCR
280bp
C
Relative expression of cyrano (fold change)
16
12
8
4
0
wild type
cyrano^ΔCR
2 hpf
24 hpf
72 hpf
D
Relative expression of cyrano (fold change)
1,8
1,2
0,6
0
wild type
cyrano^ΔCR
Heart
Skin
Kidney
Testis
E
alternative TSS
5' RACE
qPCR
RNA blot probe
cyrano^ΔTSS
84bp
F
wild type
cyrano^ΔTSS
cyrano
rRNA 18S
G
Relative expression of cyrano (fold change)
1,5
1
0,5
0
wild type
cyrano^ΔTSS
72 hpf


A
5kb
lnc-sox4
CAGE
3.5
0
H3K4me3
60
1
RNAseq
36
0
B
5' RACE 2
5' RACE 1
brain specific TSS
qPCR
lnc-sox4ΔTSS
200bp
C
Relative expression of lnc-sox4 (fold change)
0,4
0,3
0,2
0,1
0
24hpf
72hpf
D
Relative expression of lnc-sox4 (fold change)
9
6
3
0
wild type
lnc-sox4ΔTSS
Ovary
Testis
Brain
Skin
Kidney
Intestine
Spleen

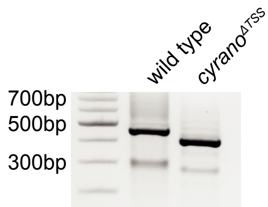

A
500bp
lnc-pou2af1
CAGE
3.5
0
H3K4me3
17
1
RNAseq
38
0
B
alternative TSS
lnc-pou2af1ΔTSS
5' RACE
qPCR
389bp
C
Relative expression of lnc-pou2af1 (fold change)
1
0,5
0
24hpf
D
Relative expression of lnc-pou2af1 (fold change)
2,5
2
1,5
1
0,5
0
Brain
Heart
Ovary
E
Relative expression of lnc-pou2af1 (fold change)
wild type
lnc-pou2af1ΔTSS
20
16
12
8
4
0
Skin
Kidney
Intestine
Testis


A
1kb
malat1
CAGE
3.5
0
H3K4me3
116
1
RNAseq
21377
1
H3K27ac
139
0
B
131bp polyA signal
ha
ha
RNA blot probe
C
wild type
malat1polyA
24hpf embryos
malat1
rRNA 18S
D
wild type
malat1polyA
brain
intestine
spleen
muscle
heart
testis
brain
intestine
spleen
muscle
heart
testis
malat1
rRNA 18S

A



B



C

WT TSS 1 (16/20)

GTCCAATAAGAGCAAAGATATCGATGTGCACGGGTGTGCGCCGGAAGAGCAGCGCCGCAGTAGCAG CACA

WT TSS 2 (4/20)

CGGGACCGAAATGGCGTAACGCGCAGTCGAGCACCGCAGCAGCGCAGAGGGGCCAGGCTCAGCGAACACCA

CAGAGCCTGGCCCGAACAACTTCAACCAGCAAATTTGTTGAAAATTACTCTTCACTCCAGTCCACATGTTTCCAA

cyrano exon 1

ACCAGCACGAGATCACCGCTGACG GTAAGTGCGGGATGAGTGTTGACGTGCACT

D

alternative TSS (12/12)

GTCCAATAAGAGCAAAGATATCGATGTGCACGGGTGTGCGCCGGAAGAGCAGCGCCGCAGTAGCAG C- - - - - - -

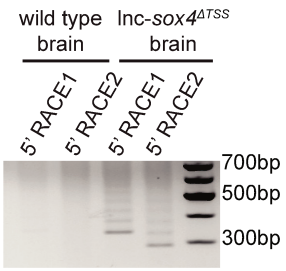
- -

- - - - - - - - - - - GCCCGAACAACTTCAACCAGCAAATTTGTTGAAAATTACTCTTCACTCCAGTCCACATGTTTCCAA

cyrano$^{\Delta TSS}$ exon 1

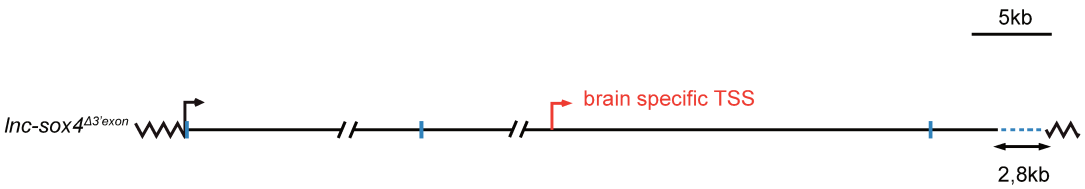
ACCAGCACGAGATCACCGCTGACG GTAAGTGCGGGATGAGTGTTGACGTGCACT

A



B

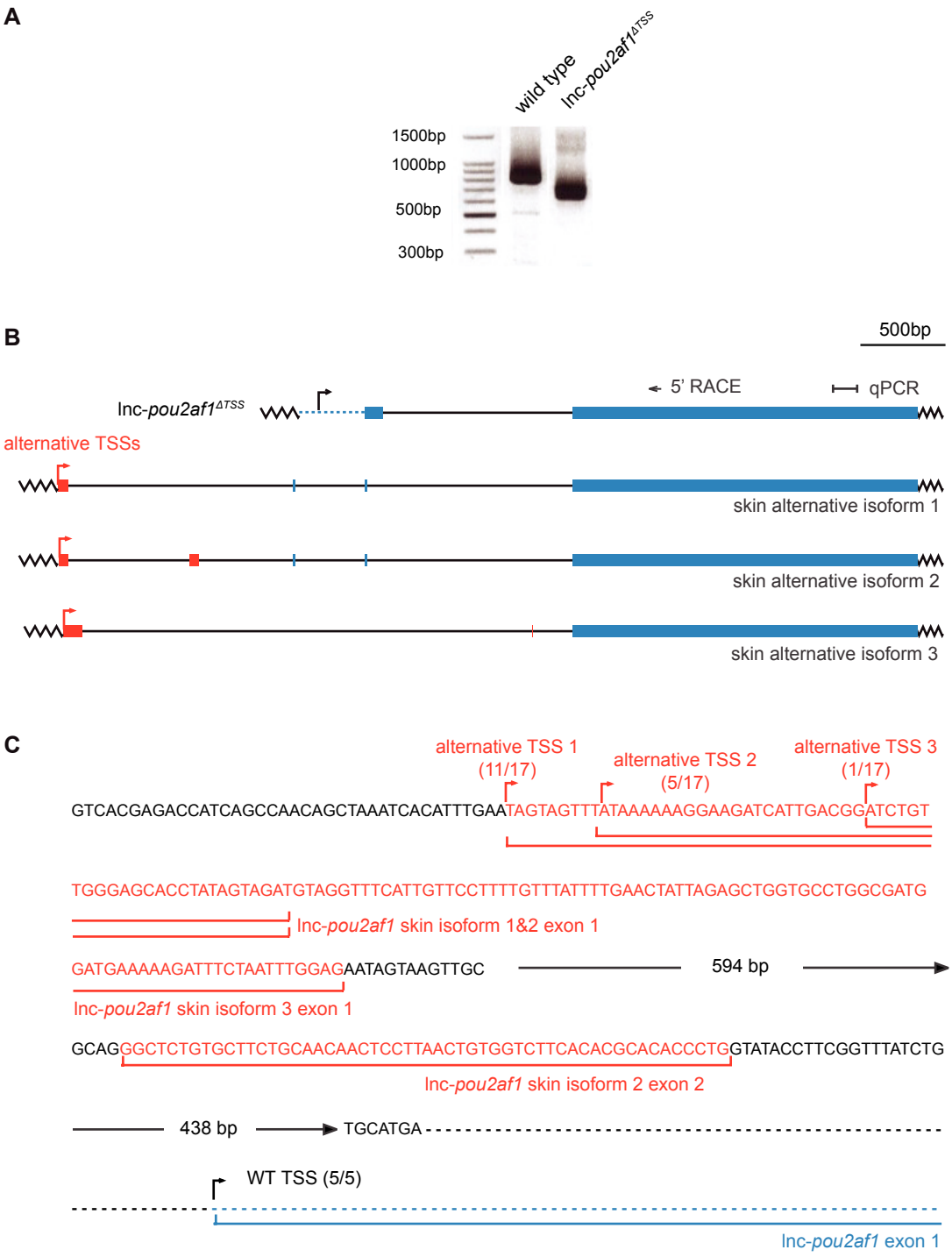
C



D

	Founder (F0)	Heterozygous (F1)	Homozygous (F2)
Number of lnc-*sox4*$^{\Delta 3'exon}$ fish/ total number of fish genotyped	7*/15	42/116	0/149

* Germ line transmission rate ranging from 2/24 to 14/24


A
wild type
lnc-pou2af1ΔTSS
1500bp
1000bp
500bp
300bp
B
500bp
lnc-pou2af1ΔTSS
5' RACE
qPCR
alternative TSSs
skin alternative isoform 1
skin alternative isoform 2
skin alternative isoform 3
C
alternative TSS 1
(11/17)
alternative TSS 2
(5/17)
alternative TSS 3
(1/17)
GTCACGAGACCATCAGCCAACAGCTAAATCACATTTGAATAGTAGTTTATAAAAAAGGAAGATCATTGACGGATCTGT
TGGGAGCACCTATAGTAGATGTAGGTTTCATTGTTCCTTTTGTTTATTTTGAACTATTAGAGCTGGTGCCTGGCGATG
lnc-pou2af1 skin isoform 1&2 exon 1
GATGAAAAAGATTTCTAATTTGGAGAATAGTAAGTTGC
594 bp
lnc-pou2af1 skin isoform 3 exon 1
GCAGGGCTCTGTGCTTCTGCAACAACTCCTTAACTGTGGTCTTCACACGCACACCCTGGTATACCTTCGGTTTATCTG
lnc-pou2af1 skin isoform 2 exon 2
438 bp
TGCATGA
WT TSS (5/5)
lnc-pou2af1 exon 1

Résumé

L'identification de divers gènes cibles impliqués dans la progression cancéreuse est cruciale afin de décrypter les mécanismes sous-jacents au cancer et de développer des stratégies thérapeutiques efficaces. Les lncARN (longs ARN non codants) sont similaires aux ARN messagers d'un point de vue moléculaire, mais ne présentent pas de potentiel codant pour des protéines. Ils sont fréquemment dérégulés et mutés dans de nombreux types de cancers. Tout comme les gènes codants pour des protéines, les lncARN des vertébrés peuvent être conservés à plusieurs niveaux: séquence, profil d'expression ou position génomique (synténie). Seuls 2% des lncARN du poisson zèbre présentent une préservation de séquence avec l'homme, tandis que plus de 35% sont conservés au niveau synténique, indiquant la présence d'une pression évolutive préservant la position génomique des lncARN. Afin d'évaluer si ce phénomène synténique peut prédire la conservation fonctionnelle des lncARN, j'ai établi un criblage génétique inverse évaluant le rôle des lncARN dans le développement du mélanome. Ces études ont été effectuées chez le poisson zèbre, un modèle animal présentant de multiples similarités génétiques, histologiques et physiologiques avec la peau humaine.

En utilisant la technologie d'édition du génome CRISPR-Cas9 pour générer les lignées de poissons zèbres mutants pour une sélection de 6 lncARN candidats, j'ai mesuré l'impact de la perte de fonction de ces lncARN sur la progression du mélanome, induit chez le poisson zèbre via l'expression de l'oncogène humain $NRAS^{G12}$ et la xénogreffes de cellules de mélanome humain. Lors de cette étude, j'ai identifié *menhir* (MElaNoma HIndrance long noncoding RNA) comme un gène suppresseur de tumeur dans le mélanome. En effet, les poisson zèbres mutants pour *menhir* présentent une altération de l'agressivité du mélanome caractérisée par (1) une augmentation de la tumorigenèse, (2) une baisse de la survie, (3) une augmentation de la sévérité du mélanome et (4) une augmentation du potentiel métastatique due à une plus grande permissivité à l'invasion des cellules du mélanome. Afin d'analyser si la fonction anti-oncogène de *menhir* est conservée dans l'évolution, nous avons exprimé l'homologue humain *CASC15* (Cancer Susceptibility 15) dans les mélanocytes des poissons zèbres mutants pour *menhir* affectés par le mélanome. Malgré l'absence de conservation de séquence, l'expression de *CASC15* atténue le phénotype d'agressivité du mélanome des poissons mutants pour *menhir*, entrainant une diminution de la progression du cancer, de la tumorigenèse et une amélioration de la survie des individus mutants affectés par le mélanome.

Par conséquent, mes résultats identifient un nouveau lncARN suppresseur de tumeur dans le mélanome et montrent que la conservation de position génomique peut être corrélée avec une conservation de fonction.

Mots Clés

Longs ARN non codants, Synténie, Zebrafish, Mélanome

Abstract

The identification of diverse target genes involved in cancer progression is crucial to decipher the mechanisms underlying cancer and to develop effective targeted treatment therapies. LncRNAs (long noncoding RNAs) are molecularly similar to messenger RNAs but lack protein coding potential. Their deregulation and misexpression as well as the presence of mutations in lncRNA loci have been linked to diseases including cancers. Like protein-coding genes, vertebrate lncRNAs can be conserved at multiple levels (sequence, expression pattern, genomic position or synteny). In contrast to only 2% sequence conservation, more than 35% of zebrafish lncRNAs are conserved at the syntenic level to human, indicating an evolutionary pressure to preserve lncRNA position in the genome. To assess if positional conservation is a predictor of functional conservation, I implemented a reverse genetic screen assaying the role of lncRNAs in melanoma development using zebrafish, an optimal oncology model with multiple skin genetics, histological and physiological similarities with human.

Using CRISPR-Cas9 genome editing technology to generate syntenic lncRNA zebrafish loss of function mutants, I profiled the impact of lncRNA loss on melanoma induced by the human $NRAS^{G12}$ oncogene and in human melanoma cell xenografts. Among six candidate lncRNAs, we identified *menhir* (MElaNoma HIndrance long noncoding RNA) as a melanoma tumor suppressor. *menhir* zebrafish mutants display impaired melanoma aggressiveness characterized by (1) accelerated tumorigenesis, (2) decreased mutant survival, (3) increased melanoma severity and (4) increased metastatic potential due to a higher permissiveness to melanoma cell invasion. To assess if the tumor suppressor function of zebrafish *menhir* is conserved throughout evolution, the human putative ortholog of *menhir* called *CASC15* (Cancer Susceptibility 15) was expressed in melanocytes of the zebrafish melanoma *menhir* mutant. Despite lack of sequence conservation, *CASC15* expression mitigated the mutant *menhir* melanoma phenotype as evidenced by reduced melanoma progression, decreased tumorigenesis and enhanced survival of zebrafish affected with melanoma.

Thus, our results identify a novel melanoma tumor suppressor lncRNA and show that conserved genomic location of lncRNAs can be used to posit functional conservation in vertebrates.

Keywords

Longs noncoding RNAs, Synteny, Zebrafish, Melanoma