Identification and targeting of functional non-coding RNA–protein interactions

Xavier Sabaté Cadenas

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Institut Curie

Identification and targeting of non-coding RNA – protein interactions

Identification et ciblage thérapeutique d'ARNs non-codants et leurs interacteurs protéiques

Soutenue par
Xavier SABATÉ CADENAS

Le 19 septembre 2022

École doctorale n° 577
Structure et Dynamique des Systèmes Vivants

Spécialité
Génétique

institut Curie

Composition du jury :

Lionel LARUE Directeur de Recherche Institut Curie	*Président*
Ramesh PILLAI Professeur University of Geneva	*Rapporteur*
Clément CARRÉ Associate Professor IBPS, Sorbonne Université	*Rapporteur*
Reini LUCO Chargée de Recherche Institut Curie	*Examinatrice*
Shahad ALBADRI Chargée de Recherche Institut de la Vision, Sorbonne Université	*Examinatrice*
Alena SHKUMATAVA Directrice de Recherche Institut Curie	*Directrice de thèse*

'He vingut pel camí dels desigs i me'n tornaré pel camí dels records'

– Mercè RODOREDA

Acknowledgments

I would like to express my deepest gratitude to each and every person who made this manuscript possible.

To Alena, for that summer transatlantic call that brought me to Paris, for your advice, teaching, and support.

To all the members of the Shkumatava lab. To Anna, for our endless conversations, climbing and psychology sessions, a moment alone next the tomb of Christ, and everything that made PhDing with you unexpectedly unique. À Louise, parce que se lever à 7h du mat, supporter mes questions de grammaire et partager les heures d'injection ont été les dommages collatéraux de ce qui est devenu bien plus qu'une camaraderie de laboratoire. T'avoir à mes côtés a été un cadeau. To Vedrana, because the bad jokes and high-risk high-gain memes have been the best price to pay for your collegiality and backing. Thank you, thank you. À Matthieu, parce que ça a bel et bien été un plaisir de partager ce projet avec toi et de voir comme tu es devenu un chercheur indépendant. À Carine, parce que l'incPRINT ne nous a pas fait périr. But also to Flavio, Allison, Florian, Antoine and Perrine. This manuscript materialized thanks to all of you.

To Shahad Albadri, Clément Carré, Lionel Larue, Reini Luco and Ramesh Pillai for accepting my invitation to be part of my thesis jury.

To Germano Cecere and Igor Ulitsky, for their guidance and counseling during the thesis advisory committee.

To my past teachers and mentors. Als professors de l'Hospital de Bellvitge i del Clínic. Anna Angulo, Dragan Hrnčić, Ann Ehrenhofer-Murray, Josep-Maria Canals Coll i Andrés Míguez. I am here because of your support as well. To past labmates and international science friends. Tim and Sara, I felt you by my side in each coffee I took. Andrew, for your healing words. Carla and Kat, because I could not miss you more. Ania, Felicitas, Helena and Sandra, also thanks to you.

To the numerous collaborators, platforms and Curie services that made the projects a reality. Azzurra de Pace, Dónal O'Carroll and Kamil Kranc. Eitan Shaulian and Vikash Kumar. Nicolas Servant and Helene Fradin. Karine Laud-Duval and Valerie Petit. FACS, NGS and Olivier Renaud. The zebrafish

facility, Armelle and Tarek. The lavarie, David and Josette. The BDD administration and acceuil. The Training Unit and Ana Rita, for your encouragement, counseling, and diplomacy. The LabEx team, Ines and Shauna and the COFUND team.

To the 3rd floor of BDD, my home. À Aurélien, pour cet appel téléphonique dans neuf ans où on discutera de tout ce qui s'est passé depuis cette conversation en espagnol pendant la fête d'anniversaire du département. À l'inimitable Soraya, hawaïenne, bagué, narratrice, heteroclite et meilleure copine de festival, confinement et voyage. To Yazan, because my critical questions never killed how enriching was to listen to you. E anche a te, Manuela, perché vincere il miliardario non ha niente da invidiare al nostro incontro. You all have been an unimageable dream team. À Samuel, pour ces bières. To Michel, Daniel and Katia, the senior voices that brought advocacy, guidance, and discussion to the cell culture room and alongside a coffee. And also to Seynabou, Emma, Carlos, Ignacio, Aya, Ramy, Camila, Gwen, Tiphaine, Audrey, Danny, Laia and Camille.

To many Curie scientists. To the Non-Coding Genome superteam, Julien, Dominika, Sonia, but specially Sara, my partner in crime, and Marina for her patience, enthusiasm, and advice. To Raphaël, Pedro, Allison, and Deborah, for their accompaniment. Raquel, prometo que alguien de la unidad heredará la bicicleta. To all Curie friends. Fairouz, Megan, Mathilde, Lorraine, and Francesca. To Rocco. And to Anouk and Tina. To the IC3i family. To Javiera, Pallavi, Marcel, Aafrin, Anny, Marci, Sandra, Silvia, Sam, Tomaso, Anne-Celine and Özge. To Darinne, for that bookshop hopping we will do with no rain. A la Raquel i la Júlia, per fer-me sentir una mica més a prop de casa. And to Deep and Jaime, for your veteran voice.

À tous ces Parisiens dont j'ai croisé le chemin, vous avez été les rayons de soleil qui illuminent les journées grises et nuageuses. Étienne, Anca, Alina, Kevin, Gerard, Capucine, Yoann et Marcos, Sylvain, Manon, Aloïs, Logan, Sabrina et Dylan. To Roxana, Halide, Negin and Andrea, for the gift of sharing such precious moments with you. And to my beloved more-than-french-class-colleagues Josien, Vivek, Reyer, Ana Isa, Eleonora, Guglielmo and Adrià.

Als amics que, tot i estar a centenars de quilòmetres, s'han fet sentir més a prop que mai. Aina, Núria, Jana, Joan i Georgina, gràcies per ser-hi. Y por aquellos a quien no les ha bastado con estar cerca que también han querido compartir viaje, gracias por esa voz de la razón, Claudia y Gonçalo.

I a la família. Montse i Mercè. Àvia Tere i àvia Mercè. Jordi, Papa i Mama. Per la vostra estima, escolta i suport incondicional. Tinc un raconet del cor guardat per cadascun de vosaltres.

Thank you. Merci. Gràcies

Abstract

The importance of RNA-protein interactions in cancer biology has been increasingly acknowledged in the past years. RNA-binding proteins (RBPs) regulate the maturation, stability, and subcellular localization of RNA transcripts. RBPs are also the drivers of the function of long non-coding RNAs implicated in multiple biological processes. In addition, RNA-RBP interactions have arisen as regulators of tumorigenesis and therefore as potential therapeutic targets.

The human long non-coding RNA (lncRNA) *CASC15* has been reported to be dysregulated in human metastatic melanoma. Our laboratory recently identified its zebrafish ortholog, whose primary sequence is poorly conserved throughout evolution. Using an inducible zebrafish skin cancer model that closely parallels the human process, we demonstrated that the zebrafish *casc15* attenuates melanoma initiation and progression. Importantly, the expression of human *CASC15* in the zebrafish lncRNA mutant rescued the melanoma progression phenotype. We hypothesized that the same RBP interactions in both orthologous transcripts underlie their functional conservation. In my first project, I investigated the molecular mechanism of action of *CASC15* in a human cell line. Similar to zebrafish, the depletion of CASC15 in human melanoma cells mutant increases their migration behavior. In addition, by combining different molecular and biochemical approaches, I identified a common set of functional protein interactors that drive the functional conservation of *CASC15* orthologs. These findings shed light on the relationship between mechanism of action and functional compatibility between lncRNA from different animal species divergent in sequence.

In parallel, I also studied YTHDF2, an important RBP implicated in multiple human cancers. This protein recognizes an abundant methylation mark present in specific RNA transcripts and regulates their stability. YTHDF2 knockout compromises leukemic stem cells while preserving normal hematopoiesis in vivo. Because of this cancer-specificity property, YTHDF2 has emerged as a unique therapeutic target. However, the identification of small molecules that target RNA-protein interactions in cell is currently limited by the absence of unbiased scalable screening methods. In my second project, I have adapted our RNA-protein interaction technology to a drug screening platform. I identified several FDA-approved compounds that specifically inhibit YTHDF2 binding to its target RNAs and established the pipeline to control for the specificity of the identified hits. In order to find novel small molecule inhibitors, I expanded the collection of tested drugs by following a ligand-based computational screen as well as including fragments of the hit compounds. In conclusion, this work demonstrates that YTHDF2 protein is druggable and corroborates the reliability of our method to be used for the identification of inhibitors targeting specific RNA-protein interactions.

Together, I have established in vivo models, cellular systems, and robust molecular tools, leading to the identification and targeting of functional RNA-RBP interactions with crucial roles in human cancers.

Résumé

L'importance des interactions ARN-protéines dans la biologie du cancer a été de plus en plus reconnue au cours des dernières années. Les protéines de liaison à l'ARN (RBP) régulent la maturation, la stabilité et la localisation subcellulaire des transcrits d'ARN, et sont également les moteurs de la fonction des ARN non codants impliqués dans de multiples processus biologiques. En outre, les interactions ARN-RBP sont apparues comme des régulateurs de l'oncogenèse et donc comme des cibles thérapeutiques potentielles.

Le long ARN non codant (lncRNA) humain CASC15 a été signalé comme étant dérégulé dans le mélanome métastatique humain. Notre laboratoire a récemment identifié son orthologue chez le poisson zèbre, dont la séquence primaire est peu conservée au cours de l'évolution. À l'aide d'un modèle inductible de cancer de la peau chez le poisson zèbre, qui correspond étroitement au processus humain, nous avons démontré que *CASC15* chez le poisson zèbre atténue l'initiation et la progression du mélanome. Ainsi, l'expression de *CASC15* humain chez le poisson zèbre mutant pour ce lncRNA permet le sauvetage du phénotype de progression du mélanome. Nous supposons que chez les deux orthologues, les mêmes interactions ARN-RBPs expliquent leur conservation fonctionnelle. Dans mon premier projet, j'ai étudié le mécanisme d'action moléculaire de *CASC15* dans une lignée cellulaire humaine. Comme pour le poisson zèbre, la délétion de *CASC15* dans les cellules de mélanome humain accélère leur migration. De plus, en combinant différentes approches moléculaires et biochimiques, j'ai identifié un ensemble d'interacteurs protéiques commun chez les orthologues de *CASC15* qui conduisent à leur conservation fonctionnelle. Ces résultats mettent en lumière la relation entre le mécanisme d'action et la compatibilité fonctionnelle entre les lncRNA de différentes espèces animales divergentes au niveau de leur séquence.

En parallèle, j'ai également étudié YTHDF2, une RBP impliquée dans de multiples cancers humains. Cette protéine reconnaît une marque de méthylation présente dans des transcrits d'ARN spécifiques qui régule leur stabilité. La délétion de YTHDF2 empêche l'apparition et le développement de la leucémie aiguë myéloïde tout en préservant l'hématopoïèse normale in vivo. En raison de cette spécificité cancéreuse, YTHDF2 est apparu comme une cible thérapeutique unique. Cependant, l'identification de petites molécules qui ciblent les interactions ARN-RBPs dans les cellules est actuellement limitée par l'absence de méthodes de criblage non biaisées. Dans mon deuxième projet, j'ai adapté notre technologie d'interaction ARN-RBPs à une plateforme de criblage de médicaments. J'ai identifié plusieurs composés approuvés par la FDA qui inhibent spécifiquement la liaison de YTHDF2 à ses ARN cibles et j'ai établi la procédure pour contrôler la spécificité des résultats identifiés. Afin de trouver de nouvelles molécules, j'ai appliqué un crible computationnel basé sur les ligands identifiés ainsi que des fragments de ces derniers. L'ensemble de ce travail démontre que la protéine YTHDF2 peut être traitée par des

médicaments et corrobore la fiabilité de notre méthode pour l'identification d'inhibiteurs ciblant des interactions ARN-RBPs spécifiques.

En conclusion, j'ai établi des modèles in vivo, des systèmes cellulaires et des outils moléculaires robustes, menant à l'identification et au ciblage d'interactions fonctionnelles ARN-RBPs ayant des rôles cruciaux dans les cancers humains.

Lay abstract

RNA molecules participate in all functions in our body and in many of its diseases. The RNA *CASC15* regulates skin cancer progression in humans, and zebrafish contain an equivalent molecule. We discovered that both have the same role in slowing down skin cancer, despite having a different sequence. We hypothesized they do so thanks to binding the same proteins. During my thesis, I discovered that human cells migrate faster when *CASC15* is missing and identified proteins that bind *CASC15* of both species. In parallel, the RNA-binding protein YTHDF2 has been shown essential for the initiation of blood cancer. Blood stem cells deficient for YTHDF2 will not develop cancer while continuing with their normal function. Given that YTHDF2 is a unique therapeutic target, I developed a technique and identified drugs that could stop this protein from binding to its target RNAs. This work shows the importance of studying RNA-protein binding for the understanding of cancer.

Résumé grand public

Les molécules d'ARN participent aux fonctions de notre organisme et à l'apparition de maladies. L'ARN *CASC15* régule la progression du cancer de la peau chez l'homme, et le poisson zèbre a une molécule équivalente. Nous avons découvert que ceux-ci ont le même rôle dans le ralentissement de ce cancer, malgré une différence dans leur séquence. Nous supposons qu'elles le font grâce à la liaison aux mêmes protéines. Pendant ma thèse, j'ai découvert que les cellules humaines migrent plus vite lorsque *CASC15* est absent et j'ai identifié les protéines qui se lient à *CASC15* chez ces deux espèces. En parallèle, la protéine de liaison à l'ARN YTHDF2 s'est avérée essentielle pour l'initiation du cancer du sang. Les cellules souches du sang dépourvues de YTHDF2 ne développent pas de cancer et maintiennent une fonction normale. J'ai mis au point une technique et identifié des médicaments inhibant YTHDF2. Ce travail montre l'importance de l'étude de la liaison ARN-protéine pour comprendre le cancer.

Abbreviations

AML – acute myeloid leukemia

ASO – antisense oligonucleotide

aTSS – alternative transcription start site

bp – base pair

CASC15 – CAncer Susceptibility Candidate 15 lncRNA

Chr – chromosome

CLIP – cross-linking immunoprecipitation

cyaa – crystallin alpha a

DDX50 – DEAD-box 50

DMSO – dimethyl sulfoxide

DUSP12 – dual specificity phosphatase 12

esiRNA – endoribonuclease-prepares small interfering RNA

gRNA – guide RNA

HIV – human immunodeficiency virus

HTS – high-throughput screen

h.CASC15 – human *CASC15*

incPRINT – in cell protein-RNA interaction

kb – kilobase

lncLOOM – lncRNA linear order conserved motifs

lncRNA – long non-coding RNA

LOF – loss of function

NMR –Nuclear Magnetic Resonance

ns – non significant

nt – nucleotide

MEME – multiple EM for motif elicitation

miRNA – micro RNA

mRNA – messenger RNA

m^6A – N^6-methyladenosine

ORF – open reading frame

polyA – polyadenylation

PRC – polycomb repressive complex

RBP – RNA binding protein

RIP – RNA immunoprecipitation

RLUC – Renilla luciferase

RNAi – RNA interference

RNA-seq – RNA sequencing

RNP – ribonucleoprotein

RRE – Rev responsive element

RT-qPCR – quantitative reverse transcription polymerase chain reaction

SEEKR – sequence evaluation through k-mer representation

SPR – Surface Plasmon Resonance

TE – transposable element

TF – transcription factor

Tnfrsf2 – tumor necrosis factor receptor superfamily 2

TRE – transactivation response element

TSS – transcription start site

UTR – untranslated region

WT – wild type

YTHD – YT521-B homology domain

zf.casc15 – zebrafish *casc15*

501mel – 501 melanoma

Table of Contents

List of Figures

CHAPTER 2

PREAMBLE

Traditionally, scientists have attributed to RNA a central role in the flow of genetic information and have considered proteins the effector tools of biological systems. Today, we know that this model that opposes informational versus catalytic functions is incomplete. RNA does not merely act as an intermediary between DNA and protein synthesis but carries out key functions in the cell. Beyond the coding potential of ribonucleic transcripts, non-coding RNAs control transcriptional and post-transcriptional processes and modulate protein activity (Geisler & Coller, 2013; Statello et al., 2021; R. W. Yao et al., 2019).

RNA rarely acts alone. As we will see throughout this dissertation, the formation of ribonucleoprotein (RNP) complexes is ubiquitous for different RNA species. In fact, protein binding defines the function of non-coding RNAs including non-coding parts of messenger RNAs (mRNAs), called 5' and 3' untranslated regions (UTRs), small and long non-coding RNAs (lncRNAs), with rare exceptions proving the rule (Guerrier-Takada et al., 1963; Kruger et al., 1982). One may argue whether the protein recognizes and interacts with a target RNA to regulate its metabolism and function or, contrarily, if a transcript recruits an RNA-binding protein (RBP) affecting its fate and role (Hentze et al., 2018). Regardless of which component is the driving force in mediating the interaction, RNAs are functionally engaged with RBPs throughout their entire life cycle (Choder, 2011; Kastelic & Landthaler, 2017).

Strangely enough, most of the graphical representations of RNA and proteins are disproportionate. In the bibliography, as well as herein, cartoons are rarely drawn to scale, and owing to displaying reasons RNAs are illustrated relatively smaller than they are. This misleading depiction can make the reader believe a transcript might bind only a handful of proteins. Yet, a transcript of 100 nt in length is able to recruit between 5 to 20 proteins (Cid-Samper et al., 2018; Ribeiro et al., 2018). An accurate sketch can help envision this conformation (Fig.1). RNA transcripts can, thus, constitute a platform or scaffold of proteins, integrating diverse regulatory RBPs to further perform complex functions (Engreitz, Ollikainen, et al., 2016).

In addition, RNA-protein interactions have miscellaneous dynamics: transient or stable interactions can occur, and the combinatorial assembling can take place in a cooperative, sequential, or competitive fashion (Achsel & Bagni, 2016). This extensive crosstalk between RNA and RBPs adds a layer of complexity and diversity to their functionalities. Non-coding RNAs are known to guide, recruit and decoy their protein partners to regulate chromosome architecture and transcription, interfere with translation and signaling pathways, modulate the activity and location of RBPs, and sequester them in order to control their abundance and availability for alternative functions (Aillaud & Schulte, 2020; Beckmann et al., 2016; Gil & Ulitsky, 2020; Statello et al., 2021; R. W. Yao et al., 2019). The next pages present how non-coding RNAs and proteins function together and serve as therapeutic targets.

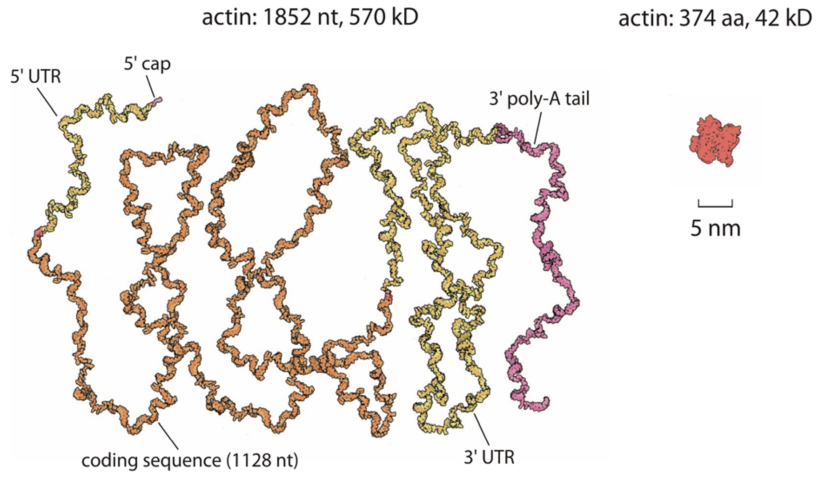


Figure 1. RNA and protein size and mass relativized
Scaled drawings of actin protein and its mRNA transcript. mRNAs are bigger than the protein they code for since the mass of a codon is an order of magnitude greater than an average amino acid, and contain untranslated regions (UTRs) and polyadenylation (polyA) tail. Extracted from (Flamholz et al., 2014).

CHAPTER 1

Mechanistic conservation of the human and zebrafish *CASC15* lncRNA orthologs in melanoma inhibition

1. INTRODUCTION

In a typical eukaryotic cell, non-coding RNA species are more abundant than protein-coding transcripts. Ribosomal and transfer RNAs account for the great majority of the total RNA mass, which is increased in diversity with long and small non-coding RNAs. LncRNAs are mRNA-like transcripts longer than 200 nt that do not have coding potential. This heterogeneous type of RNA is highly prevalent not only in vertebrates but in all eukaryotic genomes (Cabili et al., 2011; J. Liu et al., 2012; Nagalakshmi et al., 2008; Nam & Bartel, 2012; Pauli et al., 2012; Tupy et al., 2005).

The evolutionary origin, biological function, and mechanism of action of lncRNAs have been the core question marks in the spotlight of lncRNA research. Although classical long-known transcripts such as *Xist* helped pave the way, the nature of most lncRNAs remains still puzzling. In the last decade, the intense exploration of this group of genes has shed light on these problematics and uncovered key properties, numerous functions, and essential biological processes in which they are involved in.

1.1. Origin and evolution of lncRNAs

The phylogenetic study of lncRNA is constrained by, as we will see in the coming sections, the rapid divergence of lncRNA sequences across evolution and their fast turnover (Hezroni et al., 2015; Kutter et al., 2012). Protein-coding genes are highly conserved and benefit from open reading frame (ORF)-based searches, which are unhelpful for our transcripts of interest. Although much has been theorized about the origin of lncRNAs (Ganesh & Svoboda, 2016; Kapusta & Feschotte, 2014; Ponting et al., 2009; Ulitsky, 2016; Ulitsky & Bartel, 2013), examples of non-coding transcripts illustrating evolutionary scenarios are, to date, limited.

Novel lncRNAs can arise from previously existing genes. Duplication of a lncRNA gene followed by erosion of sequence similarity is a plausible mechanism, exemplified by *NEAT1*/*MALAT1* lncRNAs (Stadler, 2010) or the zebrafish (*Danio rerio*) *megamind* paralogues (Ulitsky et al., 2011). Likewise, the birth of a lncRNA gene can occur through the loss of the coding potential of an existing ORF-containing gene. *XIST*, *JPX*, *FTX*, (Duret et al., 2006; Romito & Rougeulle, 2011), and another several dozen transcripts (Hezroni et al., 2017) are known to have protein-coding ancestry. Of note, lncRNA-to-protein transition events have also been described, that is, non-coding genes that gain coding potential throughout evolution. *Libra*/*Nrep* transcripts (Bitetti et al., 2018) are a prime example of this phenomenon, termed OVERPRINTING, proposed for some hominid-specific protein-coding loci as well (J. Y. Chen et al., 2015; C. Y. Li et al., 2010; Toll-Riera et al., 2009; C. Xie et al., 2012).

An alternative explanation for the formation of non-coding genes is the de novo emergence out of non-transcribed genomic regions. The arising of transcripts from non-genic DNA, also known as EXAPTATION, has been intensively studied across multiple species for protein-coding loci (van Oss & Carvunis, 2019) and is considered a major mechanism for the formation of new genes (Neme & Tautz, 2013; Vakirlis et al., 2020). Yet, experimental evidence for this kind of non-coding transcript emergence remains anecdotal (Heinen et al., 2009; K. D. Wilson et al., 2020).

De novo non-coding transcriptional units can originate from a series of mutations or genomic rearrangements that ultimately create the appropriate combination of a promoter, splice site, and polyadenylation (polyA) sequence acquisition in non-transcribed locations of the genome. As an alternative to mutational processes, the integration of transposable elements (TE) has been proposed to be a considerable driver of de novo non-coding gene formation (Fort et al., 2014; Kapusta et al., 2013; Kelley & Rinn, 2012; Mariño-Ramírez et al., 2005). TEs can carry a functional promoter or stabilize a cryptic exon, subsequently providing favorable conditions for giving rise to a novel lncRNA (Davis et al., 2017; Fueyo et al., 2022).

These evolutionary innovations can go beyond the generation of new lncRNA genes. TEs have also been described to play a significant role in the diversification, growing complexity, and sophistication

of lncRNA loci. It has been suggested that TEs can provide cell localization signals (Carlevaro-Fita et al., 2018; Lubelsky & Ulitsky, 2018) and functional domains mediating interaction with other biomolecules, like the binding of RBPs (Johnson & Guigó, 2014). Despite their heterogeneity, scientists have systematically compared lncRNAs aiming to extract and generalize some of their characteristics.

1.2. General lncRNA commonalities

The synthesis of mature lncRNA transcripts resembles that of mRNAs. Their transcription is performed by RNA polymerase II, their introns are spliced-out and they undergo capping and polyadenylation. Nonetheless, the nature and life cycle of lncRNAs contrast with those of other RNA species regarding their cellular fate and other characteristics.

In the first place, the architectural features of non-coding loci differ from those of mRNAs. LncRNA genes tend to be shorter in size and contain shorter and fewer exons (Cabili et al., 2011; Derrien et al., 2012; Hezroni et al., 2015). Their transcription pattern, too, presents differences compared to protein-coding transcripts: lncRNA promoters are weaker and contain fewer transcription factor (TF) binding motifs (Mattioli et al., 2019), and they often undergo premature transcription termination (Schlackow et al., 2017; Tilgner et al., 2012).

Once transcribed, they are less efficiently spliced (Mukherjee et al., 2017; Tilgner et al., 2012) and tend to be less stable and have a shorter half-life when compared to mRNAs (Clark et al., 2012), resulting in an overall lower level of expression (Cabili et al., 2011; S. J. Liu et al., 2016; Ravasi et al., 2006). LncRNAs are predominantly enriched in the nucleus (Derrien et al., 2012; Mukherjee et al., 2017), although the subcellular localization can be context-specific and might even differ for the conserved lncRNAs in different species (Guo et al., 2020; Much, Smallegan, et al., 2022).

Their abundance is specific at both the tissue and cell type level (Cabili et al., 2011; S. J. Liu et al., 2016). Whereas a small proportion of lncRNAs are ubiquitously expressed, often the highly abundant ones, most non-coding transcripts are expressed at lower levels and are present only in specific organs or cell types (Jiang et al., 2016). Brain and testis show the highest enrichment of uniquely-expressed lncRNAs (Ransohoff et al., 2018).

Whole-genome alignment studies comparing genomic sequences between species allowed the identification of exon-intron structure conservation (Nitsche et al., 2015; Washietl et al., 2014) as well as similar tissue expression patterns of lncRNAs among different organisms (Necsulea et al., 2014). However, early on it became clear that lncRNAs are poorly conserved sequence-wise (J. Wang et al., 2004). The extrapolation of the functional annotation system from protein-coding genes, based almost entirely on sequence homology, appeared not to be useful for non-coding transcripts. In fact, lncRNAs show particular types of conservation which will be discussed in detail hereunder.

1.3. What does conservation mean in the lncRNA world?

In order to infer whether a locus is biologically relevant without the benefit of experimental evidence, comparative genomics studies traditionally relied on nucleotide sequence conservation. A crucial role of a gene was implied when its sequence is conserved throughout evolution. Conversely, lack of primary sequence identity was used to argue against functional significance.

From a sequence-centric point of view, lncRNAs appear to be poorly conserved. Roughly 12% of human and mouse lncRNA have orthologous transcripts with alignable sequences in other species (Cabili et al., 2011). Refined computational tools found that between human and zebrafish, two species separated by 400 million years, only ~100 lncRNAs have detectable sequence conservation, in contrast to >70% conserved protein-coding genes (Hezroni et al., 2015). These genome-wide comparisons failed to identify orthologs of vertebrate lncRNAs outside this subphylum. The small fraction of sequence-conserved lncRNAs rarely exhibits nucleotide sequence conservation along the whole transcript. Rather, they present **restricted regions of primary sequence conservation** (Fig.2A) (Derrien et al., 2012; Guttman et al., 2009). The alignable sequence between lncRNAs from two different species is five times shorter than mRNAs, usually localized in short stretches and strongest close to the promoter and the 5' end of the gene (Carninci et al., 2006; Hezroni et al., 2015; Necsulea et al., 2014).

However, there are several lines of evidence to believe that the absence of linear sequence conservation does not imply a lack of functional relevance (Omer et al., 2017; Pang et al., 2006). LncRNA turnover is much faster than most mRNAs, but their exons evolve slower than introns of protein-coding genes and intergenic sequences (Cabili et al., 2011; Guttman et al., 2009), indicating the existence of selective pressure. Mutagenesis studies of lncRNA-unrelated ultraconserved regions of the genome, which showed some of them to be dispensable for survival and development, also suggest that sequence constraint is not a faithful reflection of functionality (Ahituv et al., 2007; Snetkova et al., 2022). In like manner, some mouse lncRNAs do not have human orthologs but are reported to have key developmental regulatory roles (Klattenhoff et al., 2013; Maamar et al., 2013; Yin et al., 2015).

Indeed, nucleotide sequence conservation can be considered a restrictive criterion for transcripts that do not have to deal with protein-coding limitations, such as reading frame, codon usage, or disposition of functional elements upon translation. Today, it is clear that lncRNA genes are impacted by different evolutionary constraints. These kinds of transcripts feature less apparent conservation signatures beyond sequence similarities, that have been discovered using an evolutionary perspective.

The limitations of BLAST and other sequence alignment algorithms were overcome by taking into consideration the **genomic position conservation** (Fig.2B) and the transcriptional orientation relative to adjacent genes, a characteristic termed SYNTENY. In other words, lncRNA orthologs could be searched not only based on primary sequence similitude but also by examining flanking orthologous

sequences in distal species. These refined computational approaches (reviewed in Ross & Ulitsky, 2022) revealed the presence of thousands of syntenic lncRNAs, also named 'syntologs', conserved across numerous species (Necsulea et al., 2014; Ulitsky et al., 2011; Washietl et al., 2014), even beyond vertebrates (Hezroni et al., 2015). Of note, the majority of lncRNAs in each species showed lineage-specific expression.

The identification of deeply conserved transcripts using synteny-based transcriptome-wide comparisons indicated an evolutionary pressure to retain these lncRNAs and consequently suggested their potential biological significance. Other than equivalent genomic location, lncRNAs can own other conservation properties as well, which may sometimes overlap.

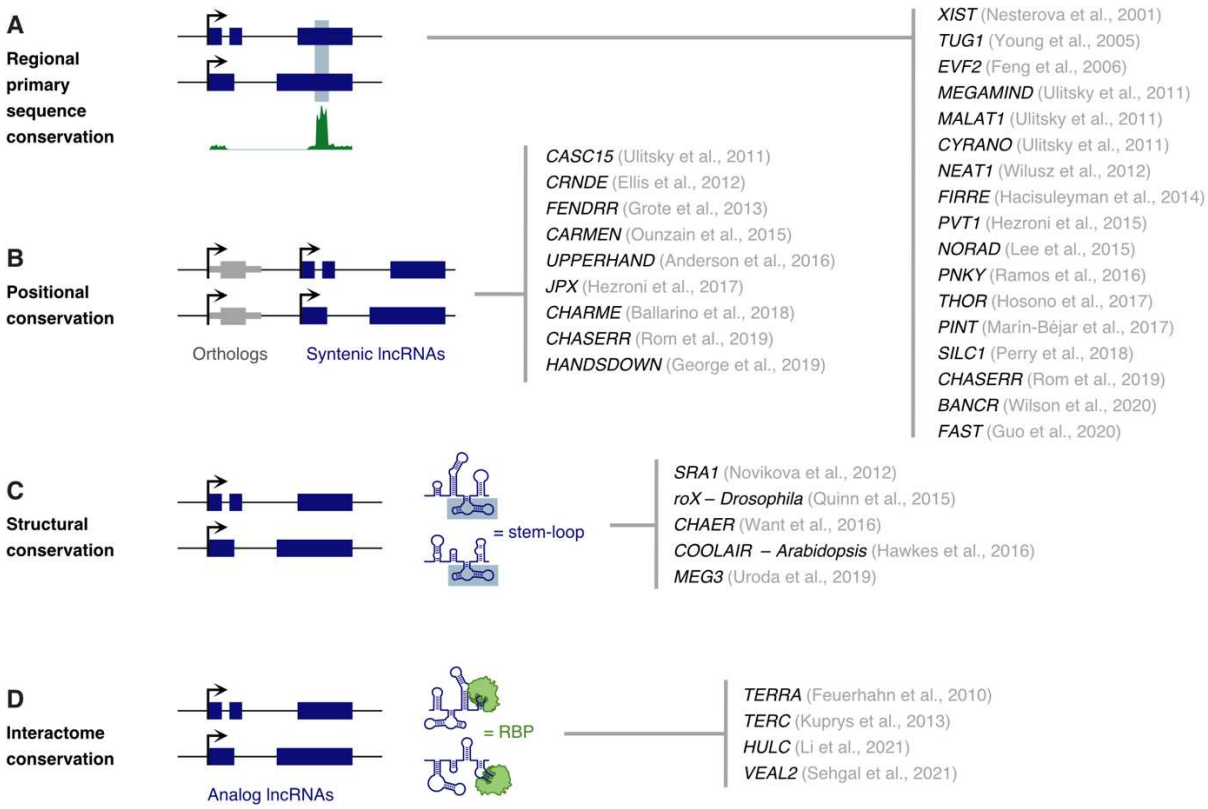


Figure 2. Multidimensional nature of lncRNA evolutionary conservation
LncRNAs mentioned in this manuscript that have been cross-species compared, grouped by their main conserved characteristic. It is worth noting these features are non-mutually exclusive, which is why is hard to classify lncRNA by conservation criteria i.e. most linear-sequence conserved lncRNAs, as well as *MEG3*, also present synteny. *NORAD*, *TERC,* and *CYRANO* share in addition interactome between species (binding to PUMILIO proteins, TERT and miR-7 respectively). *XIST, MALAT1,* and *CYRANO* have some structural preservation throughout evolution as well. Positional conservation varies in depth i.e. restricted to mammals or vertebrates. Sequence conservation length is also heterogeneous: *XIST* – multiple repeats; *TUG1* – along human-mouse sequence; *MEGAMIND* and *CYRANO* – ~300 base pair (bp) human to zebrafish; *MALAT* – ~100bp; *NEAT1* – ~100bp; *FIRRE* – ~150bp motifs; *PVT1* – multiple exons, different in mammals than fish; *NORAD* – 12 repeating units of ~300bp; *PNKY*– 2 fragments of ~100bp; *THOR* – ~200bp; *PINT* – ~400bp; *SILC1* – exon1; *CHASERR* and *BANCR* – all along the transcript; FAST – ~200bp.

Although some lncRNAs are able to perform their function as an unstructured sequence (Davidovich & Cech, 2015), unfolded states are extremely rare. Generally, most RNAs commonly form local stem-loop structures, tertiary structures, and large quaternary assemblies of ribonucleoprotein (RNP) complexes, which can mediate their molecular functions (Ganser et al., 2019). In fact, high-order structures of lncRNAs tend to be more conserved than primary sequences (R. Li et al., 2016; Novikova et al., 2013). These **conserved structural elements** (Fig.2C) could therefore represent functional domains and underlie mechanistic conservation of orthologs with nucleotide-divergent sequences.

Our understanding of RNA structures is still in a rudimentary stage. Various in silico approaches have tried to use evolutionary information to model RNA folding and identify structural preservation throughout lncRNA evolution (Rivas, 2020; Seemann et al., 2017; Smith et al., 2013). However, RNA does not adopt a stable three-dimensional structure but rather a large number of diverse conformations that coexist with different abundance (Bevilacqua & Blose, 2008; Ganser et al., 2019). In addition, changes in RNA structure can be induced by the interaction with RBPs and other molecules, by the act of transcription itself, or by post-transcriptional modifications (Ganser et al., 2019; Leulliot & Varani, 2001). Thus, the computational study of RNA structures poses a significant technical challenge. Several tools have recently been developed to probe conformational states of RNA using experimental in cellulo evidence (Morandi et al., 2021; Spitale et al., 2013; Ziv et al., 2018). Yet, structural characterization is available only for a few lncRNAs (Ross & Ulitsky, 2022).

Cross-species structural comparison has been performed only for certain lncRNAs. *XIST* A-repeat conformation has been determined to be conserved using combinatorial methods (Duszczyk et al., 2008; Nesterova et al., 2001; Smola et al., 2016). *MALAT1* contains a conserved triple helical structure located at the 3' end of the transcript (Wilusz et al., 2012; B. Zhang et al., 2017). Likewise, a local and functional tertiary structure in *MEG3* lncRNA, assessed in vitro and in cellulo, presents mammalian conservation (Uroda et al., 2019). *SRA1* (Novikova et al., 2012) and *CHAER* (Z. Wang et al., 2016) are additional examples of lncRNAs that retained structural components throughout evolution. It is worth stressing that these RNA conformations were probed in one species and were only computationally predicted to be conserved on account of patches of the transcript with a similar sequence.

Only a small number of lncRNAs have been experimentally determined to present structural conservation among different species in absence of linear sequence similarities. The paradigm of this phenomenon are *roX* lncRNAs, which feature stem-loop structures conserved in different *Drosophila* species with no recognizable sequence homology (Ilik et al., 2013; Quinn et al., 2016). Similarly, in vitro conformational probing of *COOLAIR* lncRNA in six plant species revealed highly preserved interhelical junctions without significant sequence similarity (Hawkes et al., 2016). RNA structure mapping is increasingly being used to predict, but also experimentally define, intramolecular interactions and conformations. Nonetheless, because their biological relevance is often not fully assessed, it is still a matter of controversy whether secondary structures of lncRNAs are indeed

evolutionarily conserved or confer true functional domains driving their mechanism (Ponting & Haerty, 2022; Rivas et al., 2016).

Lastly, the **molecular interactome** (Fig.2D) can be shared between orthologous lncRNAs from different species. For example, in order to induce X chromosome inactivation, mouse and human *XIST* recruit the polycomb repressive complex (PRC) (Cifuentes-Rojas et al., 2014; Dixon-McDougall & Brown, 2021). The function of *NORAD* lncRNA, too, is mediated by the binding of PUMILIO proteins both in human cells (S. Lee et al., 2016; Tichon et al., 2016) and in the mouse brain (Kopp et al., 2019). Similarly, *CYRANO* transcript contains a vertebrate-conserved near-perfect miRNA binding site that triggers its target degradation in mice (Kleaveland et al., 2018). These orthologs, however, harbor RNA elements that are resemblant and that drive the interaction with their respective partner molecules. The contrary case is *FIRRE* lncRNA: despite local repeats being partially conserved in mammals, the recruitment of hNRNPU protein occurs in humans but not in mice (Hacisuleyman et al., 2016).

Still, lncRNAs that do not share any nucleotide sequence similarities or synteny could perform equivalent functions partially through the same interactors. Non-coding transcripts from distinct species originating from different ancestor loci could evolutionarily converge to mediate the binding of certain RBPs, ultimately sharing mechanisms and constituting functional lncRNA analogs (Owen & Cooper, 1843). The marsupial *Rsx* transcript is known to regulate transcriptional silencing of X chromosome via binding PRC (Grant et al., 2012; Sprague et al., 2019), and dosage compensation in flies is controlled by *roX* lncRNAs (Meller et al., 1997). These lncRNAs share no sequence similarities with the eutherian *XISTs* and constitute putative analogous transcripts (Ramírez-Colmenero et al., 2020; Vallot & Rougeulle, 2013).

Similarly, two non-conventional lncRNAs, the telomeric *TERC* and *TERRA*, have been found ubiquitously throughout the eukaryotic domain to regulate telomeric homeostasis (Feuerhahn et al., 2010; Kuprys et al., 2013). The molecular mechanism of TERRA remains thus far to be fully understood. In yeast, this lncRNA is regulated by Rap1 protein (Luke et al., 2008), and its recruitment in humans is RAD51-dependant (Feretzaki et al., 2020). In turn, *TERC* analogs bind to TERT ortholog proteins and do so through structural domains and core elements almost identical across species (Podlevsky & Chen, 2016; Qi et al., 2013; Theimer et al., 2005; Y. Wang et al., 2016).

In contrast, two recent studies reported the finding of analogous lncRNAs based solely on their interacting partners, and only after their equivalent functionalities were established. The zebrafish *veal2* lncRNA interacts with PRKCB2 kinase to maintain the endothelial integrity of fish embryos (Sehgal et al., 2021). RNA immunoprecipitation (RIP) on this protein followed by RNA sequencing (RNA-seq) captured a human non-coding transcript that later on displayed the same regulatory functions. Similarly, two mammalian lncRNAs were found to exhibit synonymous molecular mechanisms. The mouse non-coding transcript *Pair* was first described to associate with phenylalanine hydroxylase (PAH) (Y. Li et

al., 2021). Subsequent cross-linking immunoprecipitation (CLIP) on human liver tissues identified *HULC* lncRNA to bind PAH as well. Further functional analysis demonstrated that both transcripts regulate phenylalanine metabolism.

The evidence above supports mechanistic conservation as a source of functional compatibility between lncRNA from different animal species despite rapid sequence variation. Yet, how RNA elements control interactions with RBPs to perform their function is still largely unknown. Understanding the molecular mechanism of action of lncRNA is fundamental to unraveling their function but might as well shed light on their conservation characteristics.

1.4. LncRNA mechanisms of action

The identification of the protein interactome of a lncRNA is a key step in understanding the biological role and characterizing the molecular and cellular mechanisms of these transcripts. Some RBPs interact with common RNA structural elements indistinctly, like polyA-binding proteins or the cap-binding complex (Fig.3A) (Blobel, 1973; Patzelt et al., 1983). Specific RNA recognition can also be driven by other small RNAs (Fig.3B), like microRNAs (miRNA) do with Argonaute proteins (Hammond et al., 2001) or CRISPR RNAs with Cas13 nucleases (Cox et al., 2017). Frequently, RNA binding occurs thanks to intrinsic lncRNA specificity. These substrate-specific RBP-lncRNA interactions require particular binding elements that distinguish target RNAs and off-target transcripts (Burd & Dreyfuss, 1994; Jankowsky & Harris, 2015; Lunde et al., 2007). Primary sequence motifs or defined structural signatures can allow a given RBPs to discriminate them (Guttman & Rinn, 2012; Leontis et al., 2006).

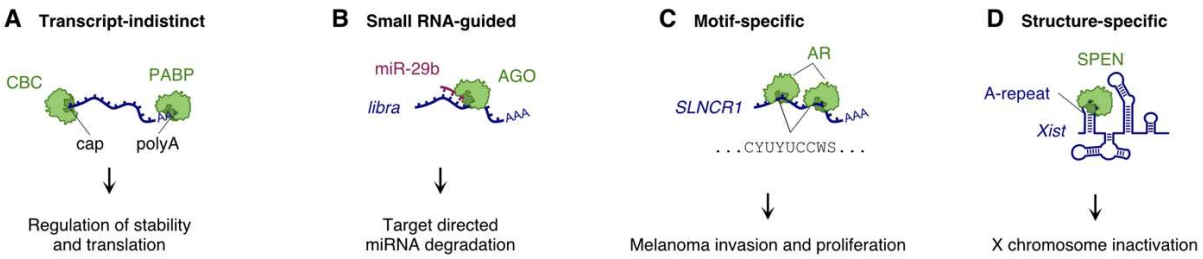


Figure 3. Modalities of lncRNA-RBP interaction specificity
(**A**) Functional interactions can be driven by common RNA components. Cap-binding complex (CBC) recognizes the 5′ 7-methyl-guanosine cap and polyA-binding proteins (PABP) bind to polyA tails. (**B**) miRNAs and crRNAs can guide Argonaute (AGO) or Cas13 proteins to their target, respectively. (**C**) Androgen receptor interacts with a specific RNA sequence motif. (**D**) The binding of PRC2 to the A-repeat region of Xist is mediated by its structural conformation.

Proteins with RNA-binding domains typically bind short, single-stranded, and low-complexity **RNA sequence motifs** (Fig.3C) (Achsel & Bagni, 2016; Cook et al., 2011; Dominguez et al., 2018; Gabut et al., 2008; Zagrovic et al., 2018). Several lncRNAs have been described to harbor small RNA elements

responsible for the recruitment of the RBPs that drive their function (Constanty & Shkumatava, 2021). Prominent examples are the UGURUAUA motif present in *Norad*, which mediates the sequestration of PUMILIO proteins (S. Lee et al., 2016; Tichon et al., 2016) or CYUYUCCWS motif, responsible for the binding of the androgen receptor to *SLNCR1* lncRNA (K. Schmidt et al., 2020). However, computational tools that perform nonlinear sequence comparisons to identify functional RNA motifs, also known as k-mers, are still in an elementary state (Ross & Ulitsky, 2022).

RBPs can also be tethered to a transcript by **secondary or tertiary structural elements** (Fig.3D) (Auweter et al., 2006; Sanchez de Groot et al., 2019; Seemann et al., 2017). For instance, stem-loop structures on *XIST*'s A-repeat drive the recruitment of the RNA binding protein SPEN (Chu, Zhang, et al., 2015; Lu et al., 2016; McHugh et al., 2015; Monfort et al., 2015) and *hFAST* lncRNA contains five individual loops that interact with β-TrCP protein, which in turn maintains the pluripotency of human embryonic stem cells (Guo et al., 2020). In addition, the structural context can determine if a primary sequence RNA binding site is occluded or exposed and able to be bound (Taliaferro et al., 2016).

The abundance of transcripts and RBPs is a major factor of consideration when studying how RNA-protein complexes are assembled. In particular, the expression level of most lncRNAs begs the question of how they can outcompete other target RNAs at such low stoichiometric concentrations. Quantitative studies (M. Wu, Yang, et al., 2021), along with research in phase-separation compartmentalization and condensate formation (reviewed in Unfried & Ulitsky, 2022) help explain how non-abundant lncRNAs can modulate the functions of their interacting partners.

From a protein-centric standpoint, only ~15% of human RBPs have defined RNA-binding domains (Cook et al., 2011; Gerstberger et al., 2014). These are often discrete, with limited ability to interact with RNA alone (Maris et al., 2005). Different RNA-binding domains can coexist in the same RBP, which enhances the specificity of the interaction (Cléry & Allain, 2012; Lunde et al., 2007). The dynamics of the known interactions between RBPs and their targets have been extensively studied (Corley et al., 2020; Hentze et al., 2018; Singh et al., 2015) and their sequence specificities, if present, often show evolutionary conservation (Ray et al., 2013). In fact, advances in structural biology determined the enrichment of inherently disordered regions in RBPs (Baltz et al., 2012; Castello et al., 2016). In addition, it should be mentioned that the RNA-bound proteome extends beyond RBPs: lncRNAs functionally interact with chromatin-associated factors and DNA-binding proteins like TFs (Conrad et al., 2016; He et al., 2016; Hudson & Ortlund, 2014; Khalil et al., 2009). Thus, unconventional RNA binding deserves particular consideration.

The idea of a non-coding 'RNA code' that allows for predicting protein binding, and as a result functionality, is an appealing one (Rinn & Chang, 2020). But RNA-protein interfaces are flexible for major types of RBPs, and RNA motifs are usually deleterious, making the existence of such code questionable (Auweter et al., 2006). In spite of this, *SLNCR1* motif discovery offers hope for lncRNA

functional prediction by extrapolation of a molecular mechanism. The androgen receptor RNA binding element was identified first in the *SRA1* transcript (Novikova et al., 2012). Based on its known function in this RNA and the presence of the same motif in *SLNCR1* lncRNA, its interaction could be predicted and subsequently confirmed (K. Schmidt et al., 2016).

Uncovering the non-coding RNA syntax would help understand the functional conservation of sequence-divergent lncRNAs. The fast progress in molecular biology granted scientists advanced tools to perform functional interrogations of these genes, placing reverse genetics at the core of lncRNA research.

1.5. The complex task of inactivating a non-coding transcript

The use of reverse genetics to answer the question of the functionality of non-coding transcripts is challenging. LncRNA loci are complex, where intricate gene regulatory components and different transcripts frequently overlap. Thus, the disruption of other elements present in the gene might interfere with the interpretation of the cause of the phenotype observed. The mechanisms that drive the function of lncRNA loci comprise (i) enhancers and other DNA-dependent elements, (ii) the act of transcription itself, and (iii) RNA transcript-derived functions (Fig.4A). To discern between these causal factors, a careful conception of the lncRNA targeting strategy is required.

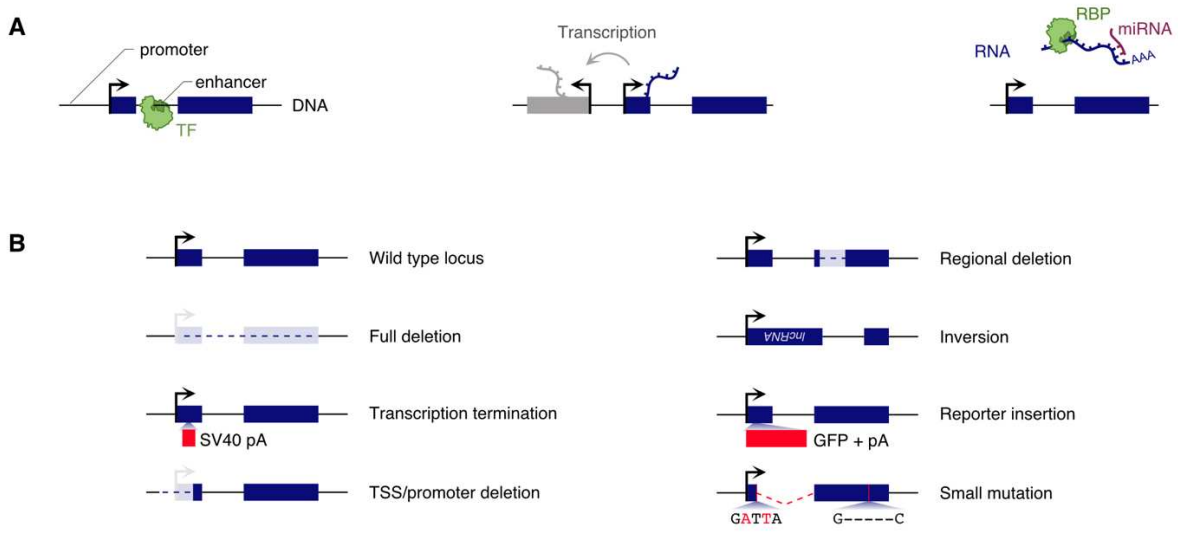


Figure 4. Genetic inactivation strategies to study the role of lncRNA loci
(**A**) Functions of lncRNAs can be DNA, RNA, or transcription-driven. (**B**) Approaches for lncRNA null allele generation. Polyadenylation (pA) insertion can also be multiple like *Evf2* (Bond et al., 2009), *Fendrr* (Grote et al., 2013), *Upperhand* (Anderson et al., 2016), *ThymoD* (Isoda et al., 2017) and *Maenli* $^{pA/pA}$ (Allou et al., 2021) mouse alleles. Depicted examples for small mutations correspond to the *Pair* $^{-/-}$ mouse model, with a 2nt deletion in the splicing site (Y. Li et al., 2021), and the *cyrano* zebrafish allele consisting of 5 nt deletion in the seed binding site of miR-7 (Kleaveland et al., 2018).

The existing approaches to perturb lncRNAs and study their functionalities are multiple. Acute loss-of-function (LOF) methods, like RNA interference (RNAi) based methods or antisense oligonucleotides (ASOs), are useful strategies for a specific and transient targeting of the RNA even though their reproducibility, efficiency, and off-target effects might hinder result interpretation (Gil & Ulitsky, 2020; Sioud, 2011). In turn, gene editing, which results in a uniform cell population or a robust animal model, offers precise and reliable means to investigate lncRNA functions. Several strategies (summarized in Andergassen 2021; Gao 2020) are depicted and explained in detail below (Fig.4B).

The first and obvious rationale to inactivate a lncRNA and elucidate its functions is the **(i) full deletion of the lncRNA locus**. This strategy proved successful at the beginning of lncRNA biology when *H19* was discovered (Leighton et al., 1995) and was applied in vivo for classical lncRNAs like *Xist* (Marahrens et al., 1997). Whole-gene ablation was useful for recent mouse studies such as *Pnky* (Andersen et al., 2019), *Firre* (Lewandowski et al., 2019), *Norad* (Kopp et al., 2019), and in several zebrafish lncRNA alleles (Goudarzi et al., 2019). The complete knockout of the locus cannot be applied, evidently, in cases where some exons are overlapping with other genes. In addition, with lncRNA loci often spanning several kilobases, the potential removal of overlapping regulatory DNA elements encoded in the gene becomes probable, limiting the inference based on the phenotype observed. Although DNA motifs and RNA functional components are not mutually exclusive, this approach strongly relies on complementary evidence compared to other minimally invasive strategies.

An orthogonal strategy to terminate transcription while preserving nearly the entire DNA locus is **(ii) the deletion of the lncRNA promoter**. Complete transcription abrogation can be achieved through the disruption of the lncRNA transcription start site (TSS), examples being *Silc1* (Perry et al., 2018), and *Teshl* (Hyeon Hong et al., 2021), as well as through the excision of upstream regulatory regions (Fitzpatrick et al., 2002; B. Zhang et al., 2012). The first caveat of this approach arises from a potential residual lncRNA expression that can occur from alternative tissue-specific TSSs or cryptic constitutive promoters, resulting in hypomorphic mutants (Lavalou et al., 2019). Furthermore, promoter competition is an important issue to consider: the possibility of a phenotype resulting from a cis-regulatory transcript-independent effect of a promoter cannot be understated (Engreitz, Haines, et al., 2016).

Another alternative editing approach consists of **(iii) the transcription termination with the insertion of an early polyA DNA sequence** to the locus of interest. *Evf2* (Bond et al., 2009), *ThymoD* (Isoda et al., 2017), or *Charme* (Ballarino et al., 2018) mouse alleles demonstrate the utility of this strategy that avoids deleting any DNA sequence. Nevertheless, obstructing the progression of the RNA polymerase complex and terminating the act of transcription itself might affect the transcription of neighboring genes and confound the interpretation of the functional mechanism of the loci, like in the case of *Upperhand* (Anderson et al., 2016; Han et al., 2019). In addition, although polyA-terminator cassette insertion is a powerful strategy to disrupt transcription without nucleotide elimination, transcription inhibition resistance or polyA signal read-through has been previously observed for some lncRNA loci.

Chaserr (Engreitz, Haines, et al., 2016; Rom et al., 2019) and *Handsdown* (Ritter et al., 2019) expression was not completely abolished upon premature termination and serve as a lesson for this phenomenon.

On some occasions, previous knowledge of the lncRNA properties might help direct the efforts toward **(iv) small or partial deletions of sequences** believed to be crucial. That is the case for linear sequence conserved transcripts, such as *PINT* (Marín-Béjar et al., 2017) or *thor* (Hosono et al., 2017). Likewise, some transcripts have exons known to be essential for the role of the lncRNA, like *Chaer* (Z. Wang et al., 2016), *ALAL-1* (Athie et al., 2020), and *BANCR* (K. D. Wilson et al., 2020). Minimal perturbations can also affect the stability of the transcript and eliminate its expression, like the recently described 2nt splicing site mutation in *Pair* mouse lncRNA (Y. Li et al., 2021).

Each one of the genetic manipulations described above confers advantages as well as drawbacks for result interpretation. Individual studies that include complementary alleles for the same lncRNA gene are becoming more and more common. Notably, the *libra/Nrep* study (Bitetti et al., 2018) was built upon two complementary LOF zebrafish alleles, a full deletion and a partial inversion, together with an additional mouse model containing a scrambled miRNA target site. To dissect *cyrano*'s molecular mechanism, the authors deleted its conserved region but also generated six alleles with minimal perturbations in the miRNA binding site (Kleaveland et al., 2018). For the investigation of the *LncGata6* function, four mouse mutants were generated: a constitutive and conditional deletion of the near-full gene body, an insertion of an SV40 polyA module into the promoter, and 4 nt mutation in a functional domain (P. Zhu et al., 2018). In a similar fashion, a comprehensive work with multiple mouse alleles successfully determined the mechanism of action of *Maenli* (Allou et al., 2021). The authors elegantly combined a whole-gene deletion, a promoter deletion, a sense and antisense polyA termination, a GFP reporter insertion, and individual exon and intron deletions.

Using evidence from combinatorial methods for lncRNA inactivation represents a solid approach when attributing a phenotype to a lncRNA loss and uncoupling it from DNA elements or transcription effects. It must be emphasized that these modes of action are not mutually exclusive. Some lncRNAs serve as a model of multifaceted lncRNA-mediated functions. *Tug1* locus contains a DNA-based element that negatively regulates the expression of neighboring genes but also has a trans-acting RNA-driven role (Lewandowski et al., 2020). Likewise, the act of transcription of *Evf2* gene impacts nearby genes and, in parallel, *Evf2* transcript controls the transcription of other loci in trans (Bond et al., 2009). In short, the adequate selection of one or multiple inactivation strategies will lead to the understanding of the mechanism of action of the lncRNA locus (Fig.4A) and the correct interpretation of the LOF phenotype.

1.6. Unveiling lncRNA regulatory roles

Over the past two decades, the systematic functional interrogation of lncRNA allowed the characterization of their vital roles in various physiological processes, such as development, homeostasis, or cell differentiation (Y. G. Chen et al., 2017; Fatica & Bozzoni, 2014; Hezroni et al., 2019; Mallory & Shkumatava, 2015; Perry & Ulitsky, 2016). LncRNAs also play key regulatory roles in the pathogenesis of multiple human diseases (Huarte, 2015; Salta & de Strooper, 2017).

In contrast, some mutagenesis screens reported no in vivo phenotypes when selected lncRNA loci were disrupted, concluding that the genes were non-functional. One laboratory found that only fifteen out of twenty mouse lncRNA knockouts were viable and showed no apparent phenotype (Lai et al., 2015; Sauvageau et al., 2013). Another study depleted 12 lncRNAs in mice, and 11 mutants were born at the expected Mendelian ratio and displayed no obvious abnormalities (Han et al., 2018). 32 deletion alleles for numerous zebrafish non-coding transcripts led to equal conclusions (Goudarzi et al., 2019). Similar large-scale lncRNA knockout studies done in *Drosophila* (Schor et al., 2018; K. Wen et al., 2016) and *C. elegans* (Akay et al., 2019; Wei et al., 2019) revealed significant fractions of mutants that lacked phenotypic changes in the analyzed traits, like fertility or growth rate.

One may argue whether the interpretation of these results should be more cautious. First, because this outcome is not exclusive to lncRNA genes as many protein-coding knockout alleles yield no overt phenotype (Birling et al., 2021; Dickinson et al., 2016; Pearson, 2002). Second, because these multiple knockout studies have several caveats that require special mention.

First and foremost, the choice of allele is not trivial when assessing a mutant model, as extensively discussed in the previous section. Besides, it is well known that phenotypes depend on the animal strain and the standard laboratory conditions (Crabbe et al., 1999; Jaric et al., 2022). Compensation mechanisms are also another important factor, as shown by Meller et al. While *Drosophila roX1* and *roX2* individual lncRNA knockouts were viable, double mutant lethality underlined their functional redundancy (Meller & Rattner, 2002). Special consideration has to be given to the context-dependency of the phenotype since the effects of the lncRNA mutation might appear only after stimulus or stress exposure. Three independent *Malat1* mouse alleles (Eißmann et al., 2012; Nakagawa et al., 2012; B. Zhang et al., 2012) and one zebrafish mutant (Lavalou et al., 2019) reported no evident phenotypes. However, it was later demonstrated that this lncRNA is a key regulator of cancer progression (Arun et al., 2016; Gutschner et al., 2013), which does not naturally occur in mice.

Finally, certain phenotypes can go unnoticed unless appropriate assays are carried out. Many lncRNAs display discrete, fine-tuning functions whose depletion causes modest phenotypes. *Ldair*, for instance, is a light-regulated transcript that controls stress-response in fish, and its expression changes upon transition from short- to long-day conditions (Nakayama et al., 2019). Another example is *Neat1*

knockout mice: initially, no phenotype was observed in mutant mice (Nakagawa et al., 2011), and only after subtle abnormalities were reported, like defects in the formation of corpus luteum (Nakagawa et al., 2014) or in mammary gland development (Standaert et al., 2014).

Phenotypes might not be obvious at the organismal level, but the inactivation of lncRNAs can still have an impact at the cellular or molecular level. For example, *Cyrano* post-transcriptionally regulates miR-7 via target-directed miRNA degradation, enabling the accumulation of a circular RNA (Kleaveland et al., 2018), yet depletion of *cyrano* in zebrafish and mice does not overtly impact viability or development (Goudarzi et al., 2019; Kleaveland et al., 2018; Lavalou et al., 2019). Even if careful phenotypic scrutiny is performed, some phenotypes might go unobserved under optimal laboratory conditions. It is not unreasonable to assume that the absence of some lncRNAs could confer a disadvantage only outside the laboratory, in a natural environment.

In essence, the obtainment of precise lncRNA LOF alleles and the rigorous inspection of phenotypic changes upon the loss of the transcript are fundamental for the understanding of lncRNA roles. Some of these functions are at times specific to one species (Guo et al., 2020), though function can also represent another layer of lncRNA conservation.

1.7. Cross-species functional conservation of lncRNAs

Whether orthologous or syntologous lncRNAs retain function throughout evolution is an intriguing question that still remains largely unresolved, as such experiments are technically demanding. Several lncRNAs have been reported to display similar or related phenotypes among different species. *NEAT1* role in the control of paraspeckle formation is retained across species (Cornelis et al., 2016; Sasaki et al., 2009); *CARMEN* is required for cardiomyogenesis both in humans and mice (Ounzain et al., 2015); *PINT* orthologs control cell invasion and are transcriptionally regulated by p53 (Marín-Béjar et al., 2013, 2017); *libra*/*Nrep* limit miR-29b expression in cerebellar granule neurons and regulate behavior in mice and zebrafish (Bitetti et al., 2018); mammalian *GAPLINC* lncRNAs have synonymous functions as inflammation modulators (Cortez Vollmers et al., 2021); *CHARME* has also been suggested to have a conserved role in mammalian myogenesis (Ballarino et al., 2018). Indeed, phenotype resemblance can inform about or indicate functional conservation.

Rescue experiments constitute a fundamental validation step toward attributing a phenotype to a LOF allele (Andergassen & Rinn, 2021). They have been used for dissecting cis-/trans-based in vivo roles of lncRNAs (Andersen et al., 2019; Grote et al., 2013; Lewandowski et al., 2019; Wei et al., 2019; K. Wen et al., 2016), even though they were always designed to re-introduce the sequence previously inactivated, that is, from the same species. In turn, cross-species rescue experiments represent the gold standard experiment to prove functional conservation throughout evolution. To date, few studies have tested if a lncRNA from one species can functionally replace its syntologous or analogous lncRNA.

Some experimental evidence comes from heterologous transgene expression in cells. Human *XIST* can silence the X chromosome in mouse cells (Heard et al., 1999); mice and human *CHAER* rescue transcriptional de-regulation in *Chaer*-depleted rat cardiomyocytes (Z. Wang et al., 2016); human *JPX* functionally complement the loss of its mouse counterpart, in absence of sequence or structural conservation (Karner et al., 2020); and human *FIRRE* transgene expression partially restores methylation marks on the inactive X chromosome upon mouse *Firre* depletion (Fang et al., 2020).

In turn, only a handful of heterologous lncRNA expression experiments have been performed in vivo. A gain-of-function experiment was performed to assess if human *THOR* retained its oncogenic capabilities across a long evolutionary distance, by expressing the transcript in zebrafish (Fig.5A) (Hosono et al., 2017). Human and lizard *EVX1AS* lncRNAs were also subjected to a functional comparison and both were determined to regulate gene expression in the same manner (Fig.5B) (Olazagoitia-Garmendia et al., 2022). The aforementioned *roX* lncRNAs from two *Drosophila* species separated by 40 million years were probed for functional conservation through a transgenic cross-species rescue experiment (Fig.5C) (Quinn et al., 2016).

RNA-based in vivo rescue experiments have the ability to test for synonymous functions in a particular context. For example, in vitro transcribed human *VEAL2* could complement the heterozygous loss of its zebrafish analog and partially restore a phenotype in zebrafish embryos (Fig.5D) (Sehgal et al., 2021). Similarly, *Pair* knockout mice treated with the human analog *HULC* RNA mimics recovered phenylalanine metabolism (Fig.5E) (Y. Li et al., 2021).

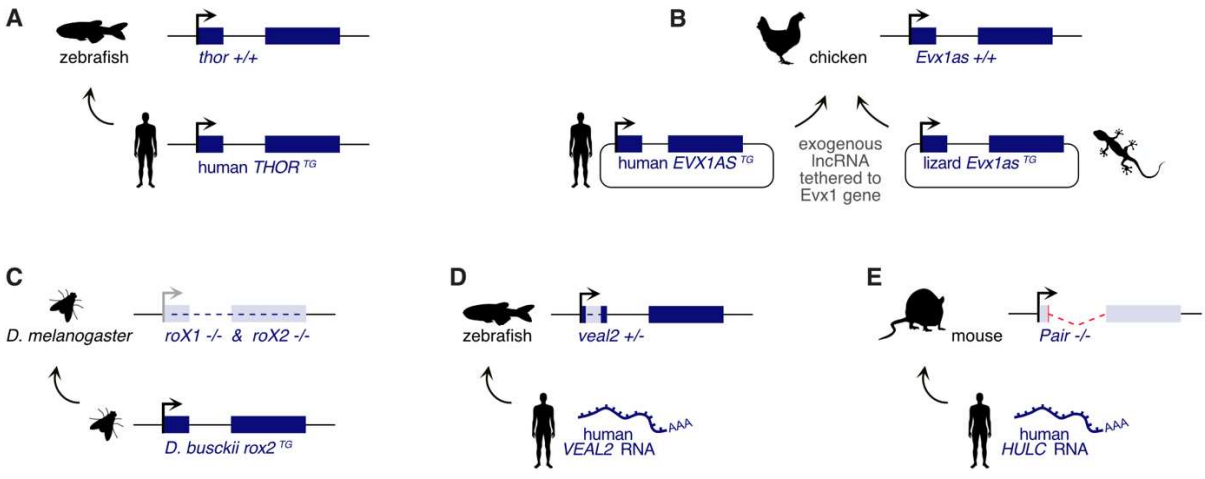


Figure 5. In vivo cross-species comparison experiments describing functional conservation.
(**A**) Male lethality in *Drosophila melanogaster roX1* and *roX2*-null allele was partially rescued upon transgenic heterologous expression of syntenic orthologue *Drosophila busckii roX2* lncRNA. (**B**) *EVX1AS* lncRNA from human and gecko lizard were able to induce EVX1 expression in chicken embryo when recruited to the locus, despite non-homologous sequences. (**C**) *THOR* human syntenic orthologue was not introduced in zebrafish *thor*-/- background where the conserved region was excised, but in a zebrafish *thor* wild type background. Therefore, strictly speaking, it cannot be considered a rescue experiment. Yet, the human lncRNA recapitulated the oncogenic in vivo function of the zebrafish transcript. (**D**) Heterozygous embryos for an 8 nt deletion disrupting the expression of *veal2* were complemented with in vitro transcribed human analog *VEAL2*, which rescued the hemorrhage phenotype. (**E**) Injection of human *HULC* RNA mimic to *Pair*-/- mouse (Fig.4B) reduced phenylalanine concentrations in blood, rescuing the phenylketonuria-like phenotype. TG = transgene.

Performing rescue experiments in vivo is particularly challenging for lncRNAs, as one has to deal with numerous variables. The success of the transgenesis depends on the choice of the lncRNA isoform if multiple ones have been described, but also the level of expression of the heterologous transcript (Grote et al., 2013), the time window of expression in the case of developmental or precise time-dependent phenotypes (Ramos et al., 2015; Sehgal et al., 2021), and the subcellular localization of the transcript or the locus where the gene is inserted, key for cis-regulatory lncRNAs (Gil & Ulitsky, 2020; Olazagoitia-Garmendia et al., 2022).

1.8. The lncRNA *CASC15* is functionally conserved among vertebrates

The human lncRNA CAncer Susceptibility Candidate 15 (*h.CASC15*) spans a ~500 kilobase (kb) -long locus. It is located downstream of SOX4, a transcription factor (TF) known to regulate epithelial to mesenchymal transition and to have crucial functions in cell differentiation, zebrafish and mice embryonic development and oncogenesis (Lefebvre et al., 2007; Lourenço & Coffer, 2017; Potzner et al., 2010; Tiwari et al., 2013; W. Wen et al., 2015). The human expression profile of these two neighboring genes is dissimilar: SOX4 is ubiquitously expressed, whereas in general *h.CASC15* is lowly expressed and only enriched in the ovary and testis (The GTEx Consortium, 2020). RefSeq annotation indicates that this lncRNA has 12 exons that produce a ~2 kb transcript, although the locus is annotated in GENCODE to encode a multitude of isoforms (Frankish et al., 2019).

Based on the literature so far, *h.CASC15* has an inhibitory role in melanoma and neuroblastoma progression. Lessard et al. identified this nucleus-localized lncRNA frequently altered in patients with metastatic melanoma, playing a role in the cancer proliferation to invasion switch. They described the existence of multiple isoforms being expressed, and by absolute quantification, they showed that the most abundant exons were in the 3' end of the transcript (Lessard et al., 2015). The expression of a shorter isoform of *h.CASC15* was later identified in neuroblastoma (Russell et al., 2015). High expression of this isoform was associated with improved cancer survival.

Another layer of complexity of the 6p22.3 locus is the presence of an antisense lncRNA between the exons 9 and 10, the NeuroBlastoma Associated Transcript 1 (*NBAT1*), that represses tumor progression in this cancer type (Mitra et al., 2021; Pandey et al., 2014). The detailed interplay and mechanism of *h.CASC15* and *NBAT1* lncRNAs in neuroblastoma have also been established (Juvvuna et al., 2021; Mondal et al., 2018), determined to be driven by the interaction with USP36 protein. In this case, the lncRNA is transcribed, processed, and exerts its function elsewhere in the cell. These so-called trans-acting lncRNAs act independently of their transcription site.

Alternatively, work done in mouse cell lines suggests that mouse *Casc15* regulates Sox4 expression (Fernando et al., 2017). Cis-acting lncRNA functions depend on the loci they are transcribed and

regulate nearby genes. However, the inactivation strategy used in this latter study was to target the TSS, which might impact upstream transcription in an RNA-independent manner (see Section1.5).

Despite the overall agreement in the tumor inhibitory function of *h.CASC15*, no null allele human cell lines have been generated and no animal models had been designed for its study in vivo. An evolutionary perspective of the function of this lncRNA was, thus, lacking. A former Ph.D. student Shkumatava laboratory, Perrine Lavalou, aimed at filling this knowledge gap and investigated the role of *CASC15* in vivo using *Danio rerio*. The zebrafish model appeared ideal since its genome contains a lncRNA syntenic to *h.CASC15* (Fig.6A) (Ulitsky et al., 2011), initially named *lnc-sox4a* and hereafter referred to as *zf.casc15*, allowing the study of this gene in two evolutionary-distant species. Indeed, these syntenic lncRNAs present a similar expression pattern in both species (Fig.6A). In addition, powerful reverse genetic approaches for zebrafish genome engineering have been established (Bedell et al., 2012; Hwang et al., 2013; Sander et al., 2011) and this animal model has been extensively used to study melanoma (Frantz & Ceol, 2020; Hosono et al., 2017; Kaufman et al., 2016; Patton et al., 2005).

Cutaneous melanoma develops upon the oncogenic transformation of melanocytes, melanin-producing neural crest-derived cells located in the epidermis. The great majority of melanoma tumors originate from mutations in either BRAF and RAS kinases which lead to an overactivation of the mitogen-activated protein kinase family (MAPK) signaling pathway (Alkallas et al., 2020; Davies et al., 2002). Zebrafish melanocytes are also derived from the neural crest, and the transcriptional program that drives this differentiation is highly conserved across evolution, mainly owing to SOX10 and the Melanocyte Inducing Transcription Factor (MITF) (Baggiolini et al., 2021; Mort et al., 2015). Yet, there are differences between the melanocytes of these two species. For instance, mammalian melanocytes secrete pigment-containing vesicles called melanosomes to transfer melanin to adjacent keratinocytes, whereas zebrafish melanocytes retain their melanosomes (Aspengren et al., 2008).

Important to highlight is the fact that zebrafish do not spontaneously develop melanoma. In order to establish a melanoma model, the Zon laboratory generated a zebrafish allele expressing human $BRAF^{V600E}$ oncogene under the melanocyte-specific promoter mitfa (Patton et al., 2005). However, the animals only developed nevi. In fact, it is only in a $tp53^{-/-}$ background zebrafish that this and other oncogenes such as $NRAS^{Q61K}$ are able to generate aggressive melanoma, histologically resemblant to human tumors (Frantz & Ceol, 2020). One oncogene that is sufficient to give rise to malignant melanoma without carrying any complementary LOF is $NRAS^{G12D}$ (Casar et al., 2018). Because embryonic expression of mutant NRAS is lethal in mice (You et al., 2021) and zebrafish (Runtuwene et al., 2011), the generation of melanoma inducible zebrafish is mediated by tol2 transposition, which creates mosaic animals while enabling robust cell-specific integration of the transgene.

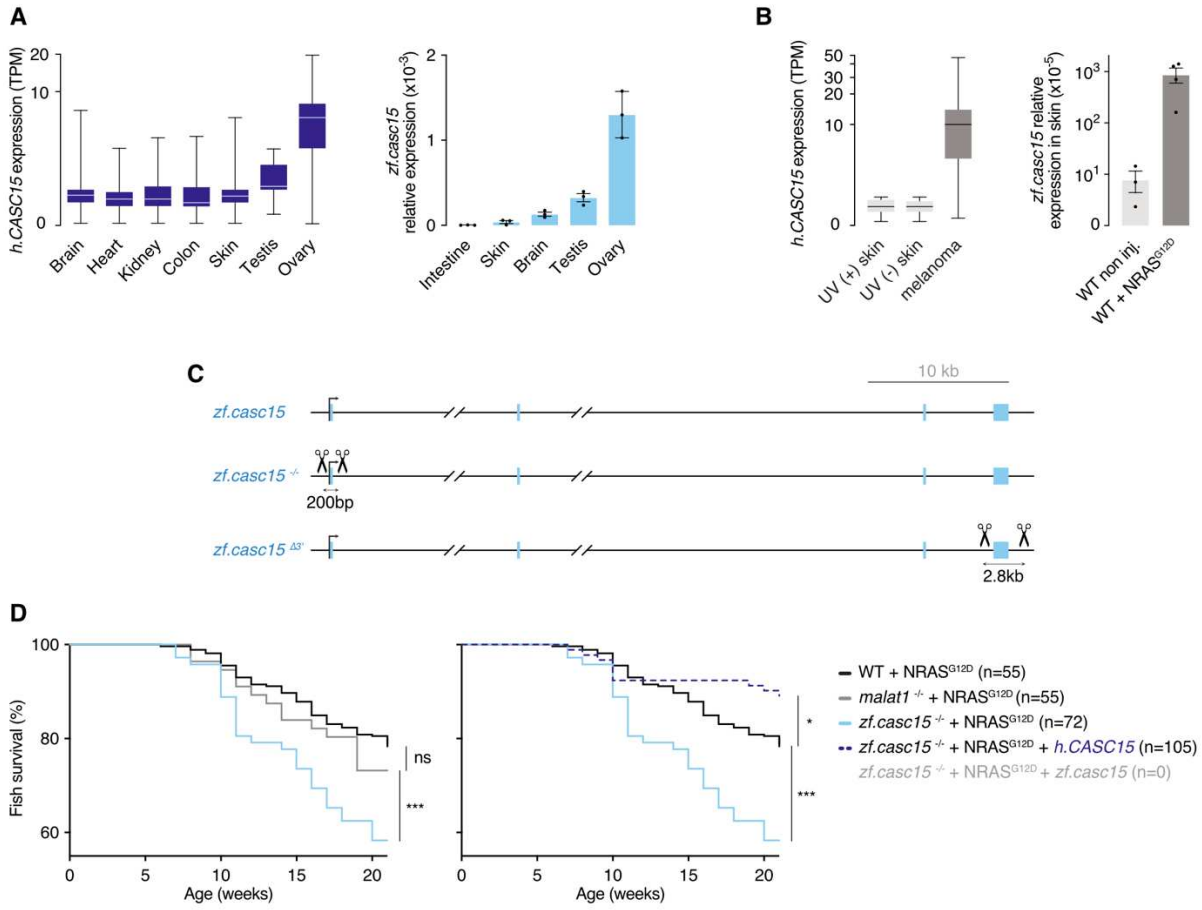


Figure 6. *CASC15* ortholog functional conservation study in vivo
(**A**) (LEFT) *h.CASC15* expression in adult organs in transcript per million (TPM), from Genotype-Tissue Expression (GTEx) project. (RIGHT) *Zf.casc15* expression in different organs of wild type zebrafish detected by qRT-PCR, relative to eef1α1. (**B**) (LEFT) Expression of *h.CASC15* in UV unexposed skin (n=265), UV exposed skin (n=394) and melanoma samples (n=36) in transcript per million (TPM), obtained through the Pan Cancer Analysis of Whole Genome project. (RIGHT) *Zf.casc15* expression in wild type zebrafish uninjected or injected with the melanoma inducible construct, detected by qRT-PCR, relative to eef1α1. (**C**) Gene architecture of the zf.casc15 locus and two mutant alleles generated in the laboratory (Lavalou et al. 2019). (**D**) Percentage of fish survival between week 5 and 21 post-injection in WT and *casc15*-/- background. Data were analyzed with a Mantel-Cox test. Unpublished results obtained by Perrine Lavalou (Shkumatava laboratory).

To uncover the role of *zf.casc15*, this melanoma inducible strategy was used. First, an upregulation of the expression of the lncRNA was observed upon heterologous expression of the NRAS oncogene, mirroring the misregulation of *h.CASC15* in patients under melanogenic conditions (Fig.6B). To investigate whether *zf.casc15* had protective or aggravating effects in the context of melanoma in vivo, a LOF allele was generated targeting the TSS (Lavalou et al., 2019). The resulting *lnc-sox4*$^{\Delta TSS}$ zebrafish (Fig.6C), here referred to as *zf.casc15*$^{-/-}$, were viable and fertile and lacked overall morphological defects. The expression of the lncRNA was abolished in all the tissues except the brain, where a tissue-specific alternative TSS (aTSS) maintained the expression of the 3' end of the transcript. An alternative strategy, in which exon 4 was deleted (Fig.6C), failed to generate homozygous fish.

In order to examine the impact of *zf.casc15* loss in melanoma progression in vivo, wild type and mutant zebrafish were injected with the mitfa:NRAS^{G12D} vector. Survival and tumor aggressivity were monitored. Upon melanoma induction, *zf.casc15*$^{-/-}$ fish showed an earlier appearance of melanoma (data not shown) and decreased survival as compared to *malat1*$^{-/-}$ or wild type controls (Fig.6D). To determine if *zf.casc15* and *h.CASC15* had synonymous functions, genetic rescue experiments were set up, expressing *h.CASC15* transcript in the melanocytes of NRAS^{G12D} *zf.casc15*$^{-/-}$ fish. This experiment allowed to address if *h.CASC15* could restore melanoma inhibition in *zf.asc15* depleted fish to the level observed in NRAS^{G12D}-expressing wild type animals. The heterologous expression of *h.CASC15* in *zf.casc15*$^{-/-}$ melanocytes slowed down melanoma onset and progression (Fig.6D), indicating that despite the absence of linear sequence conservation, *zf.casc15* and *h.CASC15* have conserved biological functions in melanoma.

1.9. Objectives

The work described above established the importance of *zf.casc15* in regulating melanoma aggressiveness and progression in zebrafish, and the equivalent melanoma-inhibitory functions of *zf.casc15* and *h.CASC15* in vivo. However, these syntenic lncRNAs lack alignable primary sequence conservation. In this chapter, I wanted to tackle the question of which mechanism allows the two orthologous transcripts to perform the same function.

Since RBPs define and transmit the function of lncRNAs, I speculated that both transcripts interact with a similar set of proteins that instruct a compatible biological role in the context of melanoma. I hypothesized that specific short sequences contained in both syntologs, not detected by classical aligning BLAST-based algorithms, would be responsible for the recruitment of conserved protein partners. If true, conserved RNA-sequence motifs present both in *zf.casc15* and *h.CASC15* would ultimately drive their functional conservation. The experiments outlined below aim at answering those question marks.

2. RESULTS

Using the zebrafish model organism, previous work in our laboratory established the functional equivalence of human and zebrafish *CASC15* orthologs. In order to facilitate the molecular study of this lncRNA, and because I could not isolate zebrafish melanocytes, I turned to a simplified cellular system. The investigation of *CASC15* in human melanoma cell lines allows the application of highly maneuverable gene editing tools to characterize this locus from an evolutionary perspective. However, no homozygous genetic inactivation of this locus had been previously obtained. To mechanistically understand the function of *CASC15*, I first generated a null allele cell line and examined the implication of the deletion at the behavioral and transcriptomic levels. Based on these results, I comprehensively interrogated *CASC15* syntologs for RBP binding and performed an unbiased cross-species comparison of their interactome. Subsequent functional analysis of the protein partners of the two transcripts enabled the identification of two proteins that we believe drive the functional conservation between *h.CASC15* and *zf.casc15*.

2.1. Organization and conservation of the *h.CASC15* locus in melanoma cells

The Shkumatava laboratory first identified 501 human melanoma (501mel) cell line, in which *h.CASC15* is robustly expressed (data not shown). While deciding on the best genetic strategy to inactivate the transcript, I detected a 48 nt deletion this cell line carries in the TSS (Fig.7A). I also confirmed the status of the main melanoma oncogenes: $BRAF^{V600E}$ mutated and $NRAS^{G12}$ wild type (Fig.7B).

As mentioned above, *CASC15* presents poor primary sequence conservation along the transcript. BLAST global alignment of the exonic sequences between human and mouse CASC15 reveals 51% of identities, much fewer than *MALAT1* (79%) or *XIST* (67%). Sequence identities are reduced to 49% when human and zebrafish transcripts are compared. Only a ~90 base pair (bp) -long stretch of the sequence is highly similar in mammals (Fig.7C), which is restricted to a 20 bp motif in zebrafish. This sequence is located in the 5' end of the transcript and likely represents a DNA regulatory element conservation rather than an RNA-related conserved region. Indeed, sequence conservation in the promoters and 5' terminus of the transcript is a feature that many lncRNAs share (Carninci et al., 2006; Hezroni et al., 2015; Necsulea et al., 2014). The overall poor nucleotide conservation throughout vertebrate *CASC15* syntologs can also be visualized by the PhastCons score (Fig.7D).

To mechanistically examine *h.CASC15*, I first sought to generate a LOF allele in melanoma cells in order to gain precise insights into its function in a cellular context. To date, only heterozygous deletion of the lncRNA or acute LOF was obtained in different cancer cell lines. Several strategies could be followed to engineer an allele that abrogates the expression of the transcript in 501mel cells. However, early polyA termination or removal of the TSS were not considered mainly because many isoforms are annotated in h.CASC15 locus (Frankish et al., 2019).

Eventually, I proceeded with a deletion of the 3' part of the transcript based on (i) the inability to obtain a homozygous deletion of the *h.CASC15* TSS in neuroblastoma cell lines (Mondal et al., 2018); (ii) the presence of the transposon-derived ANGEL elements in the last exon of the zebrafish transcript (Kapusta et al., 2013) indicating the putative presence of functional sequences; (iii) the lethality of the 3' homozygous deletion allele in zebrafish and the appearance of aTSS when the main one is deleted (Lavalou et al., 2019); and (iv) the exon usage of *h.CASC15* isoforms in different tissues (Fig.7E) (The GTEx Consortium, 2020), which suggests exon 12 is common in most transcript isoforms.

Additionally, I hypothesized that the deletion of the last exon, where the endogenous transcription termination site is located, would impact the stability of the transcript and contribute to the complete elimination of expression. The obtention of a knockout allele proved fundamental for the understanding of the role and mechanism of h.CASC15 in melanoma cells.

Figure 7. Human CASC15 locus overview.
(**A**) Outline of h.CASC15 locus and detail of the TSS. (**B**) Sanger sequencing trace of 501mel cell line in BRAF (1785-1809 nt) and NRAS (25-45 nt). (**C**) Conservation plot relative to the human locus. (**D**) PhastCons conservation score in 100 vertebrates. (**E**) GTEx Transcript Browser isoform heatmap of *h.CASC15* (ENSG00000272168.6). Tissues are indicated in rows, exons in columns. Chr = chromosome

2.2. Deletion of *h.CASC15* exon 12 inactivates the lncRNA in melanoma cells

To engineer clones with a homozygous loss of *h.CASC15* exon 12, I used CRISPR/Cas9 technology. Since 501 melanoma cell line (501mel) had never been used in our laboratory before, I optimized the transfection and selection protocols for efficient genetic inactivation, using 96-well plate GFP cell sorting (see the Methods section for details). Genotyping confirmed that 3 out of ~200 screened clones were homozygous for exon 12 deletion (Fig.8A, 8B).

Detailed scrutiny of the locus, using transcriptomic data of 501mel cells later obtained, revealed that the main isoform in this human cell line uses an aTSS downstream of the main TSS, resembling the *zf.casc15* brain-specific isoform (Lavalou et al., 2019). The resulting shorter transcript only contains the last three exons and is similar to the ones previously described as *CASC15-S* or *CASC15-004* (Lessard et al., 2015; Mondal et al., 2018; Russell et al., 2015) (Fig.8C), referred herein as *h.CASC15-short*. CAGE-seq data of two other melanoma cell lines supports the generalized usage of an aTSS in this type of cancer cell (Fig.8D).

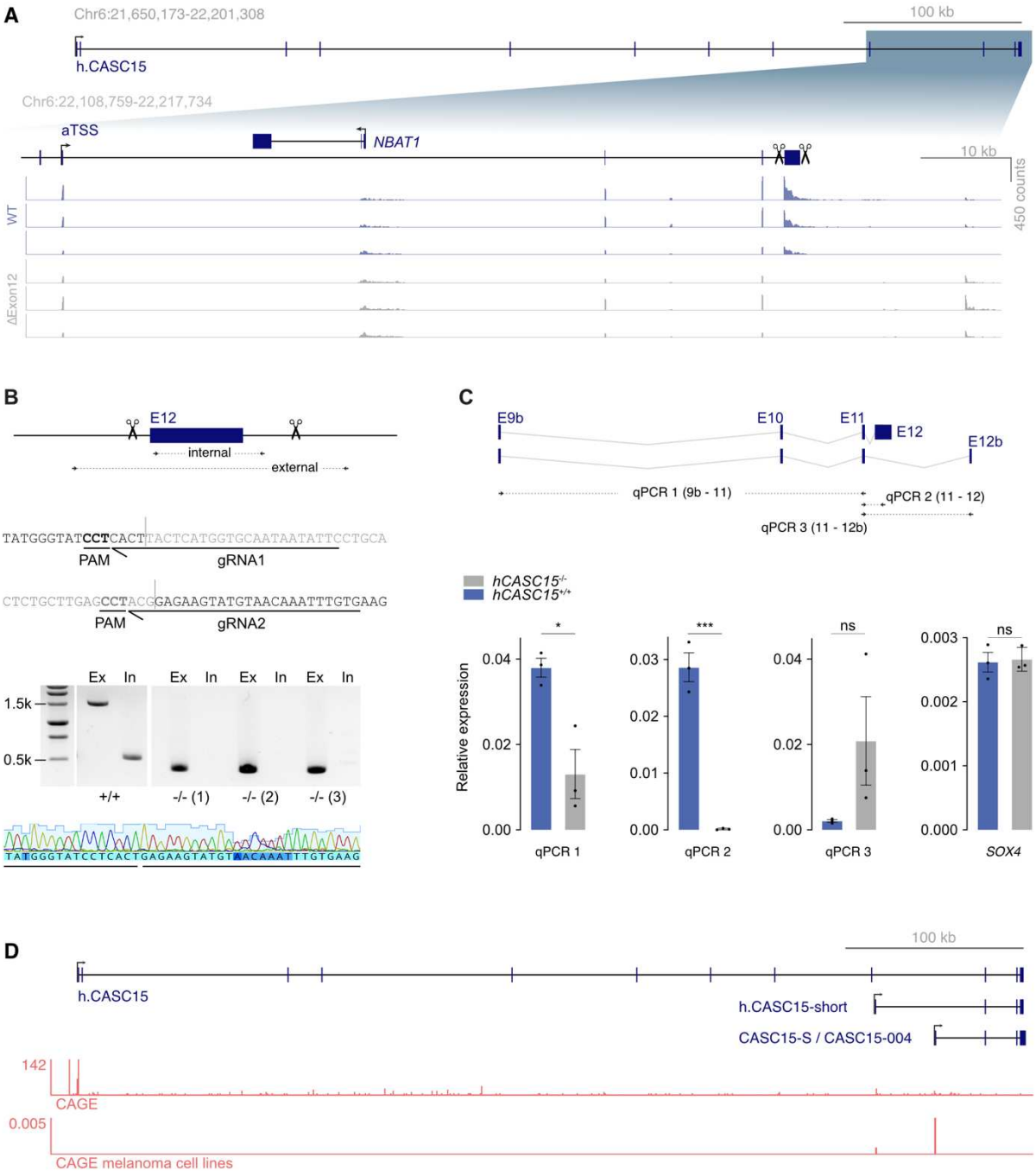


Figure 8. Deletion of exon 12 generates a hypomorphic *h.CASC15* mutant in 501mel cells.
(**A**) Outline of h.CASC15 locus and detail of the 3' end. (A) KO of h.CASC15 Exon 12 using CRISPR/Cas9 and RNA-seq coverage in the 3' end of h.CASC15 locus in *h.CASC15*+/+ and *h.CASC15*-/- 501mel cells, normalized together to the same scale. aTSS = alternative transcription start site. (**B**) (TOP) Schematic representation of the design to generate the null allele by targeting exon 12. gRNAs are indicated below, and primers used for colony screening are marked by arrows and listed in the Supplementary Table. In = Internal. Ex = External. (BOTTOM) Representative image of genotyping results for *h.CASC15*+/+, *h.CASC15*+/- and *h.CASC15*-/- clones, and sequencing trace of one *h.CASC15*-/- clone. (**C**) *h.CASC15* expression in *h.CASC15*+/+ and *h.CASC15*-/- in 501mel cells relative to ß-actin detected by RT-qPCR using specific primers targeting exons indicated in the scheme. *P* values were calculated by unpaired t-tests. **P* < 0.05; ****P* < 0.001; ns = not significant. (**D**) CAGE-seq counts for two melanoma cell lines (COLO-679 and G-361) at the h.CASC15 locus. CASC15-S / CASC15-004 represent the short CASC15 isoforms identified by Mondal et al. and Russell et al. Chr = chromosome

Regarding the effects of the deletion of exon 12 in the transcript stability, I observed an incomplete abrogation of the *h.CASC15* expression. A residual truncated transcript is expressed at ~30%, generating a hypomorphic *h.CASC15* mutant, hereafter referred to as *h.CASC15*$^{-/-}$ (Fig.8C). The expression is not entirely abrogated likely due to the usage of exon 12b, inefficiently spliced in unedited conditions, eventually stabilizing the transcript after the removal of exon 12.

LncRNAs and transcription factors located in the genomic vicinity tend to regulate each other at the transcriptional level (Gil & Ulitsky, 2020). Therefore, I investigated a potential local cis-regulation function of *h.CASC15* transcript. Upon deletion of *h.CASC15* exon 12, no changes in expression were detected in the SOX4 protein-coding gene (Fig.8C) recapitulating what was observed in *zf.casc15*$^{-/-}$ zebrafish (data not shown). The Tol2-dependent random insertion of *h.CASC15* during in vivo rescue experiments (Fig.6D) also supports the trans-acting function of this lncRNA.

Together, these data show that exon 12 deletion largely abrogates *h.CASC15* expression in melanoma cells. These experiments also suggest that *h.CASC15* acts independently of SOX4, excluding a cis-acting role of this lncRNA in melanoma cells. Likewise, the presence of a short *CASC15* isoform is a recurrent feature in different contexts, both in human and zebrafish.

2.3. Conservation of the ectoderm-specific *CASC15-short* isoform throughout vertebrates

Melanocytes reside, among other organs, in the epidermis. The gene expression programs of melanocyte differentiation slightly vary between species. However, in all of them this cell type migrates through the developing embryo from the neural crest into the skin (reviewed in Mort et al., 2015). Therefore, I wondered if the expression of *CASC15-short* is exclusive to zebrafish brain and human melanoma and neuroblastoma cell lines or if a smaller transcript is prevalent in ectoderm-derived tissues across different vertebrates. To that end, with the help of Igor Ulitsky (Weizmann Institute of Science), we examined *CASC15* tissue expression features throughout several species using publicly available RNA-seq data as well as transcriptomic data generated by our group (Fig.9).

In human, mouse, dog, and chicken the main isoform is transcribed from CASC15 loci starts near SOX4 and has a dozen of exons, and it is expressed primarily in the ovary and thymus. In the human brain, an aTSS between exons 9 and 10 is used, resulting in the expression of a similar transcript to 501mel cells, and coinciding with GTEx data (Fig.7E). This second, shorter isoform is also present in mouse, dog, and chicken brain, and is often inefficiently spliced. It is noteworthy that between mammals and chicken there is also sequence conservation in the region where this short isoform starts (Fig.7D). *CASC15* transcripts in lizard and *Xenopus* have fewer exons and transcription starts at the 5' end of the loci. Of note, one additional exon at the 3' end is used in *Xenopus* brain, despite the earlier exons also being present.

As mentioned above, previous work in the Shkumatava laboratory described the usage of a brain-specific aTSS upon deletion of the main TSS (Lavalou et al., 2019). This result indicated an active mechanism to maintain a basal expression of the 3' end of the transcript in the zebrafish brain. Nevertheless, and consistent with this publication, several publicly available RNA expression datasets from adult zebrafish brain show a low expression of *zf.casc15* in this tissue. In turn, in early stages of development, this locus shows active transcription. These findings suggest *zf.casc15* might play a role in zebrafish brain development.

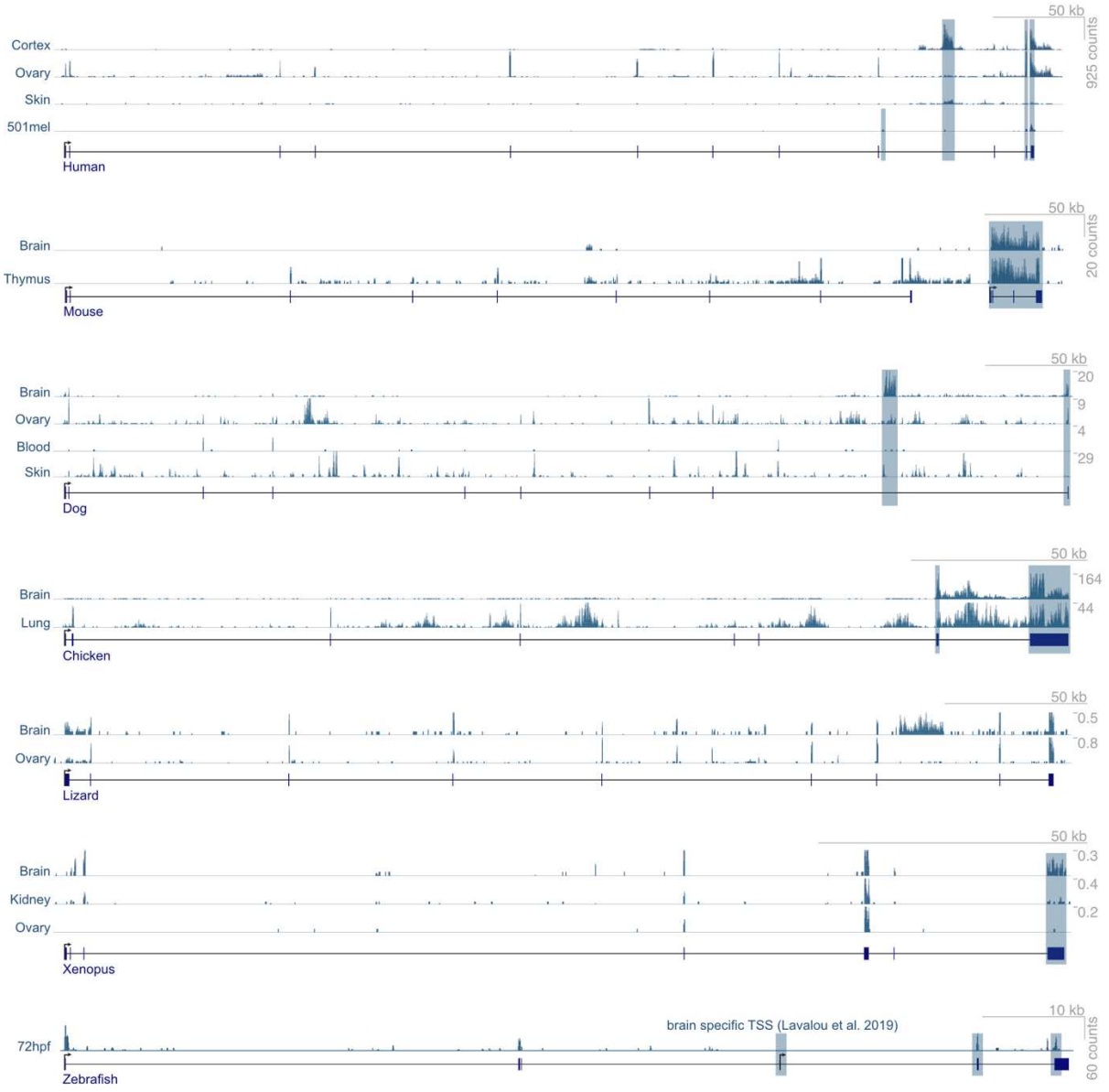


Figure 9. *CASC15-short* isoform is present in ectoderm-derived tissues across different species.
Genomic loci of CASC15 orthologs in the indicated vertebrate genomes. RNA-seq data are from HPA (human) (Fagerberg et al., 2014), ENCODE (mouse), SRP009687 (dog), SRP016501 (chicken), SRP009831 (lizard), SRP039546 (*Xenopus*) and unpublished data from Shkumatava laboratory (zebrafish).

2.4. Behavioral characterization of *h.CASC15*$^{-/-}$ melanoma cells

Transient depletion of *h.CASC15* expression in melanoma cells previously uncovered a role of this transcript in proliferation and invasiveness capabilities, which differed between cell lines (Lessard et al., 2015). To further characterize the role of *h.CASC15* in 501mel cells, I examined the impact of the exon 12 deletion on the oncogenic properties of the newly generated cellular allele. To do so, I subjected 501mel cells to a series of standard cellular assays. I first tested the effect of the deletion on the proliferative behavior of 501mel cells. However, no difference was observed when comparing *h.CASC15*$^{-/-}$ and *h.CASC15*$^{+/+}$ cells (Fig.10A).

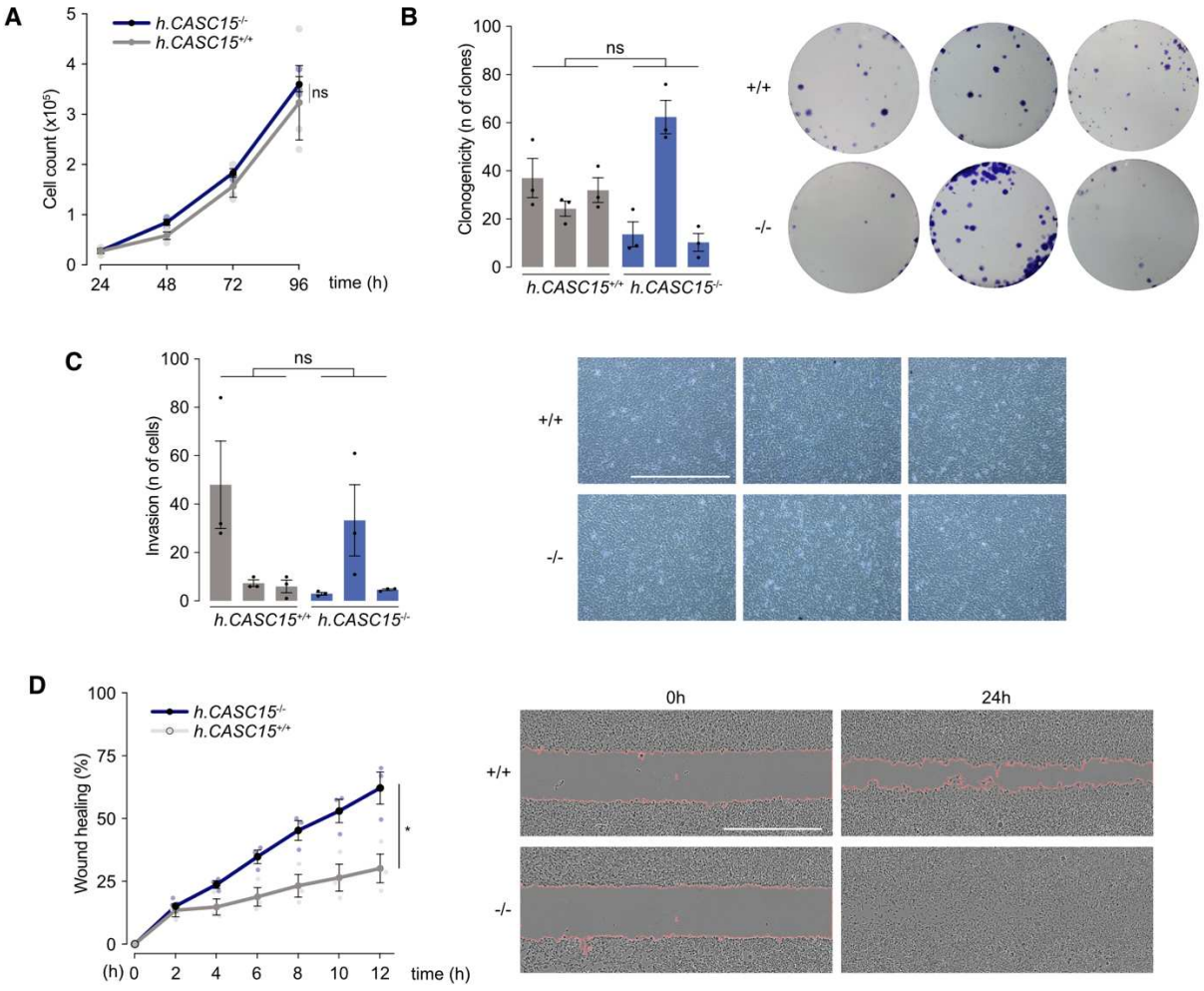


Figure 10. Cell migration is increased in *h.CASC15*$^{-/-}$ 501mel cells.
(**A**) Proliferation assay. (**B**) Colony formation assay quantification and representative images. (**C**) Invasion assay quantification and representative images (scale bar = 0.5 mm). (**D**) Wound healing assay quantification and representative images (scale bar = 1 mm). Data from three independent *h.CASC15*+/+ and *h.CASC15*-/- clones replicates are presented as mean ± s.e.m.; *P* values were calculated by two-way ANOVA; **P* < 0.05; ns = not significant.

Later, I investigated the clonogenic and invasive potential of *h.CASC15*$^{-/-}$ and *h.CASC15*$^{+/+}$ cells. Different cell clones from the same genotype showed different clone formation capacities (Fig.10B) as well as invasiveness when placed in Matrigel-coated Boyden chambers (Fig.10C). Thus, no clear effect on those behavioral features could be attributed to *h.CAS15* exon 12 deletion. It is worth mentioning that intrinsic heterogeneity of the parental melanoma cell line may have an impact on the divergence of results between clones of the same genotype.

Finally, I used the wound healing test to examine cell migration. In this assay, two populations of cells are seeded 500 µm apart and recorded in order to measure the time of wound closure, indicative of migration capabilities. *h.CASC15*$^{-/-}$ cells displayed an increased migratory behavior compared to the *h.CASC15*$^{+/+}$ (Fig.10D), reminiscent of the effect of *zf.casc15* inactivation on zebrafish melanoma progression. Taken together, this result indicates that *h.CASC15* inhibits cell migration in the melanoma context.

2.5. Transcriptomic characterization of *h.CASC15*$^{-/-}$ melanoma cells

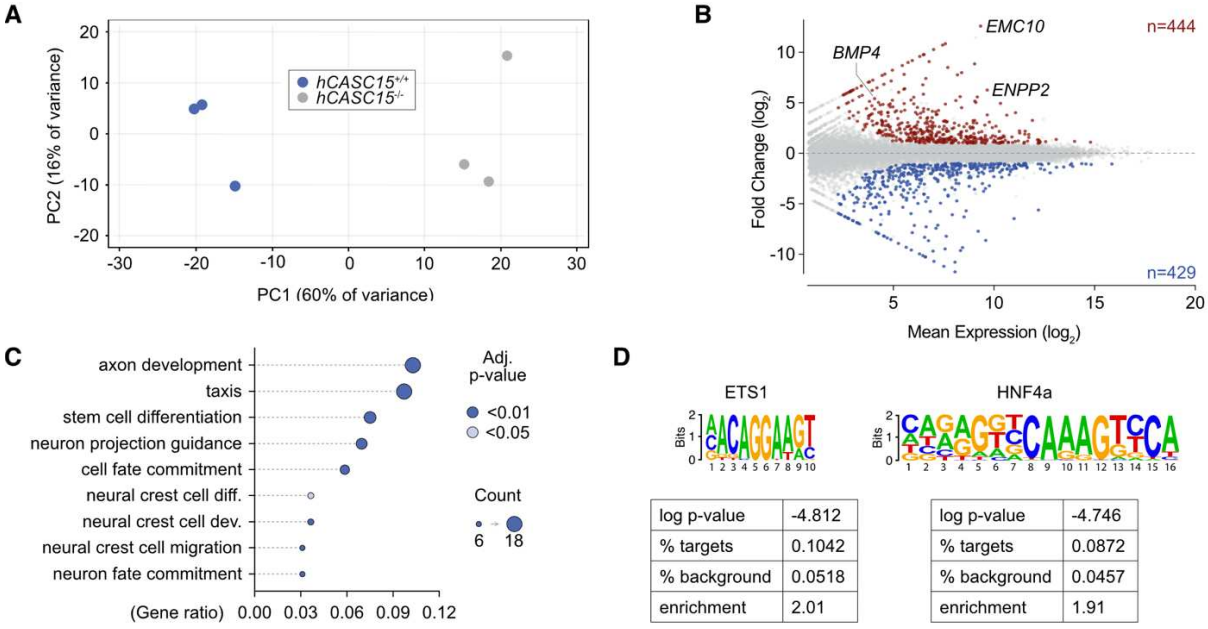
Many lncRNAs are known for being key regulators of gene expression (Statello et al., 2021). Because *h.CASC15* is located in the nucleus (Lessard et al., 2015; Mondal et al., 2018; Russell et al., 2015), I hypothesized that this RNA transcript would regulate gene expression, and that the loss of *h.CASC15* has an impact on the transcription of certain melanoma-related genes or a global effect on gene regulation of the cells.

To investigate the effect of *h.CASC15* inactivation in gene expression of 501mel cells, I performed comparative transcriptomic analysis. Three *h.CASC15*$^{+/+}$ and three *h.CASC15*$^{-/-}$ independent clones (see the Methods for details) were sent for RNA-seq and clustered together (Fig.11A). Together with Florian Constanty, a Ph.D. student of the Shkumatava laboratory, I identified 873 differentially expressed genes (Fig.11B). Biological process analysis of upregulated genes revealed neural crest cell and migration processes as highly enriched terms (Fig.11C). For instance, BMP4, EMC10, and ENPP2 have been implicated in cell migration and motility in different developmental and cancer models (Frisca et al., 2016; Reboll et al., 2017; Woo Nam et al., 2000; S. Zhou et al., 2021).

To identify potential regulators of gene expression changes, we performed DNA motif analysis (Heinz et al, 2010) allowing the identification of conserved TF binding DNA elements enriched in promoters of differentially expressed genes (Fig.11D). The analysis of upregulated genes revealed enrichment of the HNF4a DNA binding motif in the dataset, suggesting a subset of the genes were under the transcriptional control of the HNF4a transcription factor. HNF4a has been reported to have important functions in different cancer types (Brunton et al., 2020; Lucas et al., 2005; Teeli et al., 2021; Z. Wang et al., 2020) but has not been yet implicated with melanoma.

The top-ranked motif for downregulated genes was the highly conserved ETS DNA-binding domain, containing a core GGAA/T sequence (Macleod et al., 1992). In the melanoma-specific context, ETS1 has been reported to function in the activation of mesenchymal-epithelial transition (Kubic et al., 2015), to drive recurrent mutagenesis (Mao et al., 2018), and to participate in phenotype switching (Wouters et al., 2020). The ETS family of transcription factors is also implicated in melanoma invasion (Rothhammer et al., 2004).

In conclusion, the transcriptomic analysis of our mutant cells is consistent with the behavioral changes observed in the *h.CASC15* mutant cells, revealing a reactivation of neural crest and migration gene signatures.



log p-value	-4.812
% targets	0.1042
% background	0.0518
enrichment	2.01

log p-value	-4.746
% targets	0.0872
% background	0.0457
enrichment	1.91

Figure 11. Gene expression changes upon *h.CASC15* exon 12 deletion.
(**A**) PCA plot of RNA-seq samples from *h.CASC15*+/+ and *h.CASC15*-/- in 501mel. (**B**) Volcano plots showing changes in gene expression in *h.CASC15*+/+ 501mel cells compared to *h.CASC15*-/-. Color points correspond to adjusted *P* value <0.05 (**C**) Bubble plot with biological process enrichment of upregulated genes. (**D**) Motif analysis of promoter region of downregulated (left) and upregulated (right) genes.

2.6. Dissection of the molecular mechanism of action of *CASC15*

I then wanted to address the molecular mechanism behind the functional conservation of *h.CASC15* and *zf.casc15*. The transcript does not have predicted miRNA binding sites and previous experiments constitute in vivo and in cellulo evidence that *CASC15* orthologs regulate melanoma progression in *trans*. Thus, I hypothesized that both lncRNAs would bind to equivalent RBPs that would drive their function. To functionally characterize *CASC15* orthologs, I sought to unbiasedly identify the full protein interactome of the two syntenic transcripts.

Available techniques to identify cellular lncRNA-RBP interactions rely on RNA affinity purifications, which are often low efficiency and are at risk of contamination by nonspecific RBPs. In turn, in vitro methods might not recapitulate the appropriate cellular conditions for binding. To overcome these caveats, I used the technique developed in the laboratory named in cell protein-RNA interaction or incPRINT, (Graindorge et al., 2019; Sabaté-Cadenas & Shkumatava, 2020) (Fig.12A). It is based on a high-throughput immunoprecipitation of RBPs followed by luciferase-based detection and quantification of their interactions with the RNA of interest. This technique allows the screening for in-cell binding of 3,000 human proteins including RBPs, transcription factors, and chromatin modifiers.

Several incPRINT experiments had previously been run in laboratories with facilities adapted to high-throughput technologies, not the case for our laboratory in Institut Curie. Therefore, with the help of Allison Mallory, I set up protocols for high-throughput bacteria culturing, DNA prepping, and DNA concentration measurement in a 96-well format. We adapted as well the core incPRINT steps to be performed in a robot-independent manner, for the in-house implementation of this technique.

I later interrogated *h.CASC15* and *zf.casc15* with the library of human RBPs using incPRINT. As expected and as shown for other transcripts non-related to the project, the majority of proteins did not bind to the tested RNAs. Different sets of proteins were identified to interact specifically with *h.CASC15* or *zf.casc15*. When comparing the two datasets, 15 proteins were detected to interact specifically with both *h.CASC15* and *zf.casc15* (Fig.12B). Remarkably, many of the identified proteins have been reported to be involved in DNA damage response, metastasis, migration, tumorigenicity, apoptosis or angiogenesis among others.

Next, I sought to determine if any of the 15 common protein interactors is critical for the lncRNA function. To test this, I individually depleted each of the RBPs in *h.CASC15*$^{+/+}$ 501mel cells and tested their migratory behavior with the wound healing assay. I hypothesized that, if functionally relevant, the protein knock-down and subsequent disruption of the lncRNA-RBP interaction would cause a significant increase or decrease in the migration of the cells compared to the control. Depletion of DDX50 and DUSP12, but not the rest of the interactors, recapitulated the increased cell migration of *h.CASC15*$^{-/-}$ 501mel cells (Fig.12C, 12D), suggesting that these proteins are relevant for *h.CASC15* to perform its melanoma inhibitory function. Together, these findings demonstrate that DDX50 and DUSP12 are functional partners of *h.CAS15* and *zf.casc15*.

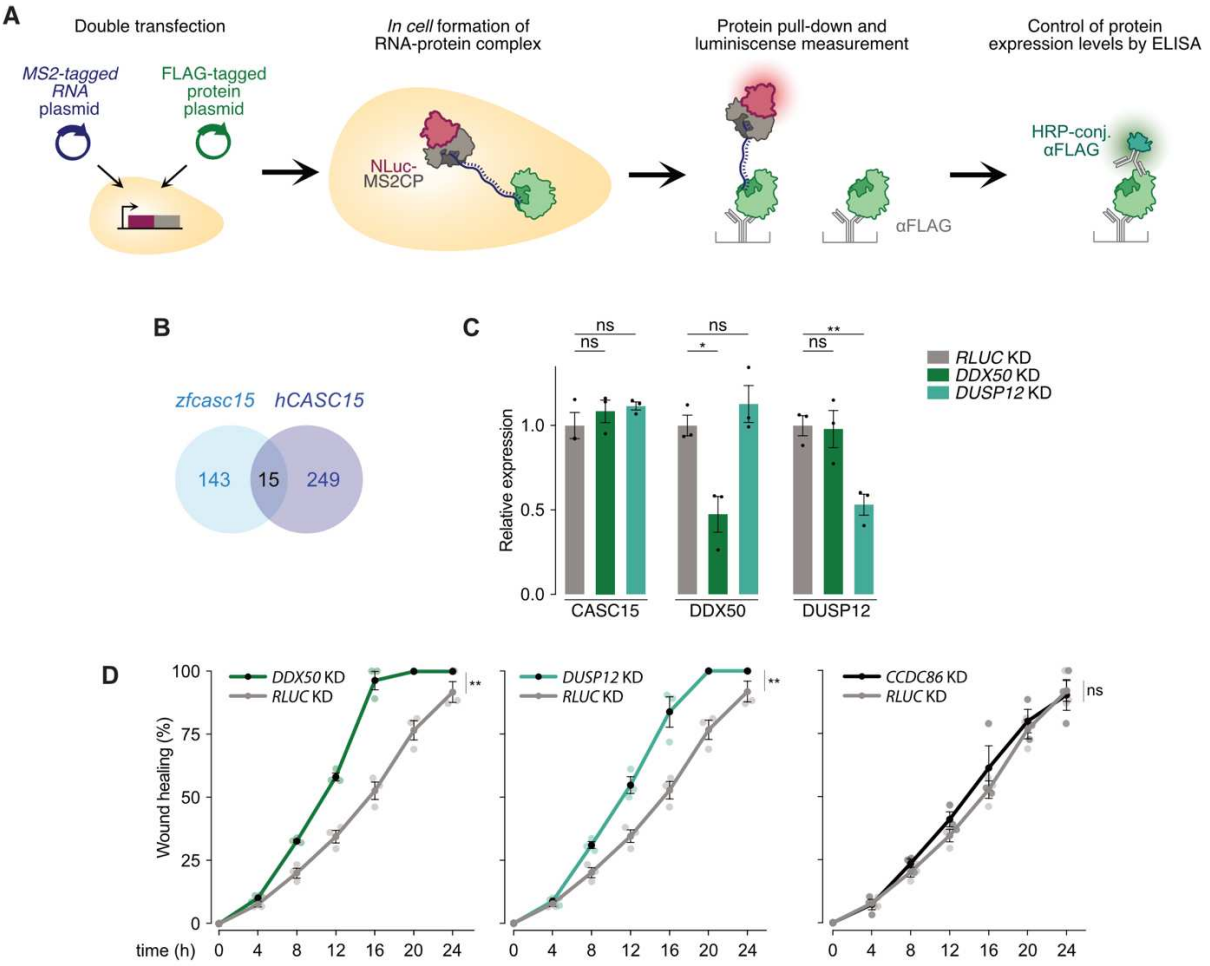


Figure 12. DDX50 and DUSP12 are functional protein binders of *h.CASC15*.
(**A**) Overview of the incPRINT framework. (**B**) Venn-diagram depiction of the overlap for the interactors identified testing zebrafish *casc15* and *h.CASC15* against a library of 3,000 RBPs using incPRINT. (**C**) *h.CASC15*, *DDX50* and *DUSP12* expression after 24h of endoribonuclease-prepared siRNA (esiRNA) treatment detected by RT-qPCR, relative to ß-actin and normalized to RLUC negative control. Data from three biological replicates are presented as mean ± s.e.m. *P* values were calculated by unpaired t-tests (**D**) Wound healing assay with *h.CASC15*+/+ 501mel cells treated with esiRNAs for DUSP12, DDX50, CCDC86 or RLUC negative control. Data from three independent transfections are presented as mean ± s.e.m. *P* values were calculated by two-way ANOVA. **P* < 0.05; ***P* < 0.01; ns = not significant.

The interactions identified by incPRINT occur upon exogenous co-expression of the RNA and proteins of interest. With the aim to know if the binding happens in melanoma native conditions as well, I preceded with an orthogonal validation of the interaction under endogenous conditions. To facilitate the molecular study of the proteins, I sought to use CRISPR/Cas9 to insert a tripartite HA-PreScission-His tag (Pérez-Rico et al., 2020), hereafter referred as to HA-tag, in frame after the start codon of DDX50 or DUSP12. The design of the strategy (Fig.13A) and genotyping results (Fig.12B) are shown below and illustrate the successful obtention of two 501mel cell lines with endogenous RBPs N-terminally fused with an HA epitope, named DDX50^{HA} and DUSP12^{HA}.

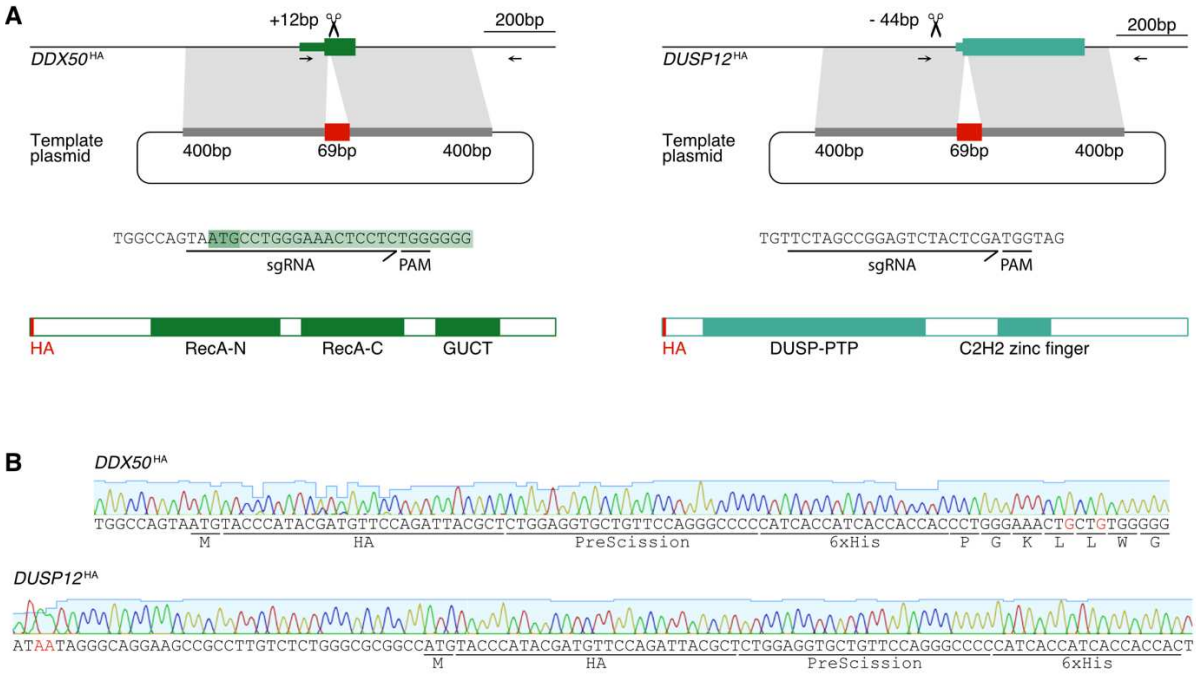


Figure 13. DDX50 and DUSP12 are endogenously tagged.
(**A**) Targeting strategy for the insertion of an HA-tag into the 5' terminus of (LEFT) DDX50 and (RIGHT) DUSP12. A schematic of the donor vector and CRISPR-mediated homologous recombination targeting strategy is shown. Primers used for genotyping of the knock-in alleles by PCR are indicated with arrows and correspond to the primers listed in the Supplementary Table. Below are indicated the sgRNA sequences used and the protein sequence scheme tagged with the HA epitope (see section 3.4 for details on the domains) (**B**) Representative sequencing trace of the tagged region of heterozygous clones. In red, nucleotides changed to inactivate the template without changing the amino acid sequence. GUCT = 'Gu' -patient C-terminal domain. DUSP = dual-specificity phosphatase. PTP = protein tyrosine phosphatase. C2H2 = Cysteine 2 Histidine 2.

To test if the endogenous DDX50^{HA} and DUSP12^{HA} interact with the endogenous *h.CASC15*, I performed RNA immunoprecipitation (RIP) followed by RT-qPCR analyses. Using this approach, I identified enrichment of the *h.CASC15* transcript upon DDX50^{HA} protein pulldown, confirming the interaction under native conditions (Fig.14A, 14B). However, no specific band on the DUSP12^{HA} input or after RIP could be detected by Western Blot (Fig.14C) using an α-HA antibody. The absence of suitable commercially available antibodies prevented us from using this strategy to validate DUSP12^{HA} – *h.CASC15* endogenous interaction.

Because I speculated that the tagging could affect the stability of the protein, I generated a 501mel cell line in which the HA-tag was fused at the C-terminal end of DUSP12. No protein could be detected either with whole-cell lysates or after RIP (data not shown). As an alternative strategy, I aimed at exogenously expressing DUSP12^{HA} protein in 501mel cells to test for binding to the endogenous *h.CASC15*. I amplified DUSP12^{HA} from cDNA of both N- and C-terminally tagged 501mel cell lines and cloned them into a vector (Fig.14D). Both DUSP12^{HA} proteins were successfully detected by Western Blot (Fig.14E). Subsequent RIP experiments will be performed to test for *h.CASC15* enrichment upon DUSP12^{HA} immunoprecipitation.

Additionally, to test if DDX50^{HA} binds specifically to the exon 12 of *h.CASC15*, I used *h.CASC15*$^{-/-}$ 501mel cells to generate an additional cell line that contained the deletion and DDX50 fused to HA-tag. However, mutant cells appeared to be more sensitive to a second genetic manipulation, since changes in morphology, low growth rate, and massive cell death were observed (data not shown). To overcome that, DDX50^{HA}-containing plasmid will also be generated, and together with DUSP12^{HA} plasmid, RIP will be performed in both *h.CASC15*$^{+/+}$ and *h.CASC15*$^{-/-}$ 501mel backgrounds. This experiment would ultimately allow the assignment of RBP binding to the exon 12 of *h.CASC15* or elsewhere.

Together, these data confirm that the melanoma-inhibitory function of *h.CASC15* is mediated via binding to DDX50 helicase. Binding to DUSP12 under endogenous conditions remains to be further examined, as well as the precise location of the binding site of the two RBPs along *h.CASC15* transcript.

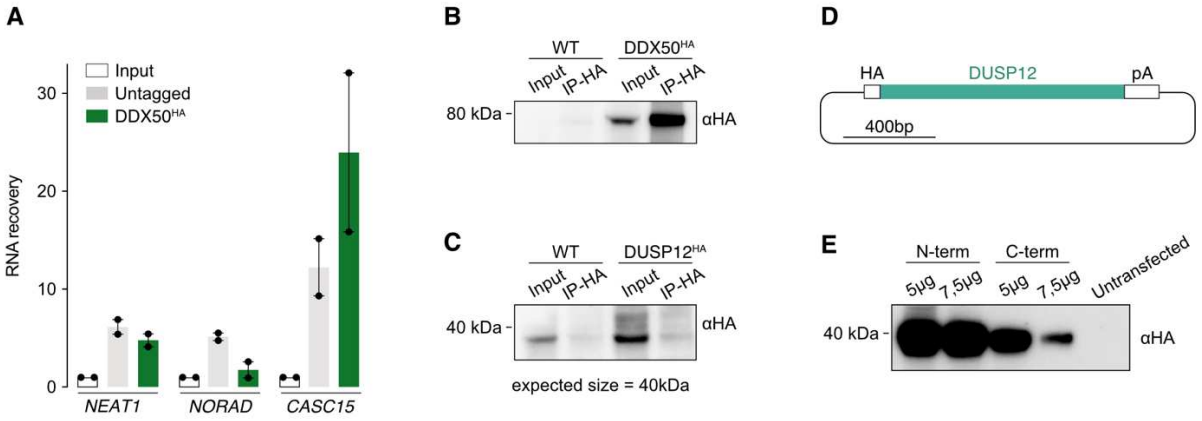


Figure 14. Endogenous DDX50 interacts with *h.CASC15*.
(**A**) RT-qPCR using primers targeting the indicated regions following IP with the indicated antibody. Data shown as relative expression of each RNA to ß-actin and fold change relative to input, n = 2. (**B,C**) Western blot control after IP using α-HA antibody on the indicated sample. 1% input, 20% IP. (**D**) Schematic representation of one of the pCDNA3.1 plasmids containing DUSP12-HA that will be used for exogenous expression and subsequent RNA immunoprecipitation. (**E**) Western blot after transfection of indicated µg of DUSP12-HA containing plasmid, detected using α-HA antibody. WT = wild type

2.7. Region-specific interaction of *h.CASC15* protein binders

The use of RIP in combination with HA-tagging and *h.CASC15* mutant cell lines could reveal if exon 12 is responsible for DUSP12 and DDX50 recruitment. However, I aimed at using a combinatorial approach to answer the very same question. Previous work of our team showed that incPRINT's enables the identification of RNA region-specific RBPs (Graindorge et al., 2019). Having done a comprehensive description of the interactome of *h.CASC15* testing the full-length transcript against our library of RBPs (Fig.12B, Fig.15A), I wanted to bring incPRINT one step further and exploit its capabilities to detect region-specific interactions.

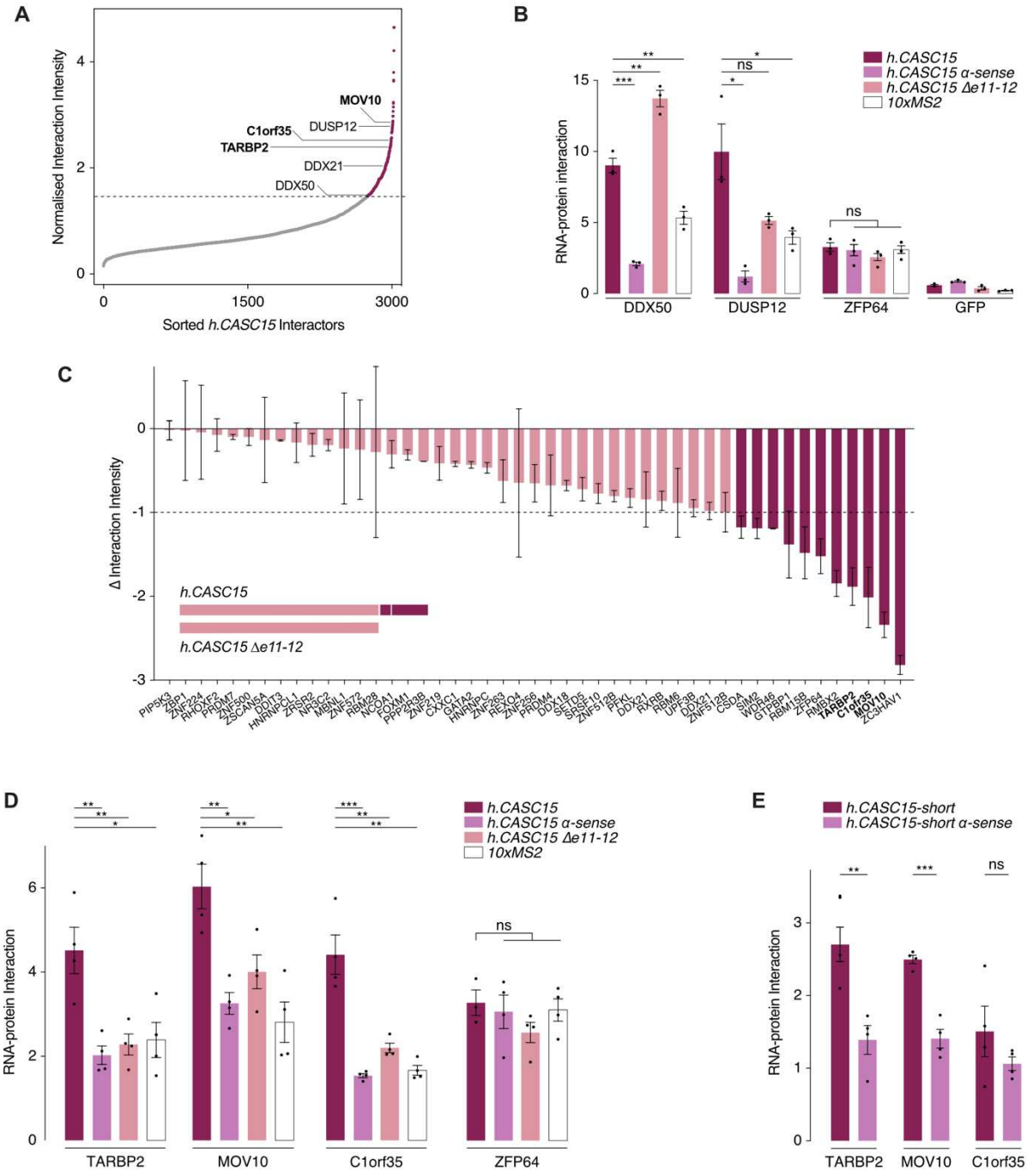


Figure 15. Distinct *h.CASC15* protein interactors bind specifically to the 3' end of the transcript.
(**A**) Normalized RNA-protein interaction intensities averaged from two biological replicates detected by incPRINT, sorted in increasing order. The horizontal dotted line represents a cut-off used for the classification of interactors. See methods for normalization. (**B**) Interaction intensities of DDX50, DUSP12, ZFP64 and GFP negative control tested against the indicated 10xMS2-tagged transcripts in the legend. Data from three biological replicates. (**C**) Difference in interaction intensity of each protein between *h.CASC15* full-length and *h.CASC15* lacking exon 11 and exon 12. Data from two biological replicates. Only the top 70 *h.CASC15* interactors were tested. Shown are the proteins whose interactions with the truncated transcript decrease. (**C & D**) Interaction intensities detected by incPRINT. Transfected RNA plasmids are indicated in the legend, protein plasmids are indicated on the x axis. Data from four biological replicates. All values are given in Relative Light Units (x10^5) as mean ± s.e.m; *P* values were calculated by unpaired t-tests; **P* < 0.05; ***P* < 0.01; ****P* < 0.001; ns = not significant.

First, I generated a *h.CASC15-MS2* antisense (α-sense) transcript to use as a negative control. I tested the interaction of DDX50 and DUSP12 and compared it to *h.CASC15-MS2* sense, which revealed a significant difference in binding (Fig.12B). ZFP64 protein was used as a negative control and GFP as background. To identify if the RBPs interact with the 3' part of the transcript, I used an MS2-tagged *h.CASC15* transcript that lacks exon 11 and exon 12. A significant increase in DDX50 binding was identified and a decrease in DUSP12 interaction was detected, although not significant (Fig.12B). Thus, these results are not conclusive to assign a region-specific binding site to our conserved functional protein interactors.

Later, I sought to repeat a similar test using additional *h.CASC15* interactors. I selected the top 70 *h.CASC15* binders and used *h.CASC15-MS2 Δe11-12* to test for regional-specific binding. A dozen of RBPs whose interaction decreased the most in the truncated transcript (Fig.15C) were selected for further specificity tests. TARBP2, MOV10, and C1orf35 displayed a significantly higher interaction to the full-length *h.CASC15-MS2* transcript compared to the control RNAs (Fig.15D), suggesting that functional sequences are located at the 3' end for recruitment of these protein interactors. A similar strategy using *h.CASC15-short* sense and antisense RNAs confirmed the findings for TARBP2 and MOV10 (Fig.15E). Further functional analysis using acute LOF and wound healing assessment (Fig.12D) would confirm if these two protein interactions, specific for human *CASC15,* are relevant for the melanoma inhibitory function of the lncRNA.

Together, these results demonstrate that incPRINT allows assigning RBPs interactions to precise segments of an RNA molecule. Of note, additional optimizations and complementary MS2-tagged RNA controls are needed to infer some hard-to-interpret lncRNA-RBP interactions.

2.8. Identification of short, conserved RNA sequence motifs in *CASC15* orthologs

The interaction of common RBPs to *CASC15* syntologs occurs despite a lack of alignable nucleotide sequence conservation. Thus, I hypothesized that short conserved sequences representing functional RNA motifs in the two orthologous lncRNA transcripts are responsible for the binding of RBPs, enabling their equivalent functionality. In order to identify RNA elements that are conserved from *zf.casc15* to *h.CASC15*, together with Caroline Ross and Igor Ulitsky (Weizmann Institute of Science, Rehovot, Israel), we performed RNA motif analysis by LncLOOM (lncRNA linear ordered conserved motifs) (Ross et al., 2021). This recently developed algorithm uses the sequences of the syntenic lncRNAs from different species to search for short, ordered, evolutionarily conserved k-mers.

Curation of *CASC15* orthologous sequences from eight vertebrate species, located using RNA-seq data, and subsequent exploration using LncLOOM revealed four conserved RNA sequence motifs (Fig.16). The super-conserved stretch located at the first exon (Fig.7C) is likely to interfere with the finding of relevant elements distributed along the transcript. Noteworthy, *h.CASC15-MS2* transcript used for the

zebrafish rescue experiment (Fig.6D) and in incPRINT (Fig.12B) did not contain the first exon. We then performed LncLOOM analysis of *CASC15* orthologs prior to the removal of the first exon in all species, resulting in no motifs conserved up to zebrafish (data not shown).

In an effort to refine the search, we sought to use *h.CASC15-short* as well as the equivalent shorter ectoderm-specific isoforms in other species (Fig.9). However, the fact that some isoforms were inefficiently spliced, like the 20 kb mouse *Casc15*, made it impossible for LncLOOM algorithm to detect conserved RNA sequence motifs. Together, the results above indicate the current impossibility to identify conserved RNA motifs between *h.CASC15* and *zf.casc15* due to the algorithm limitations.

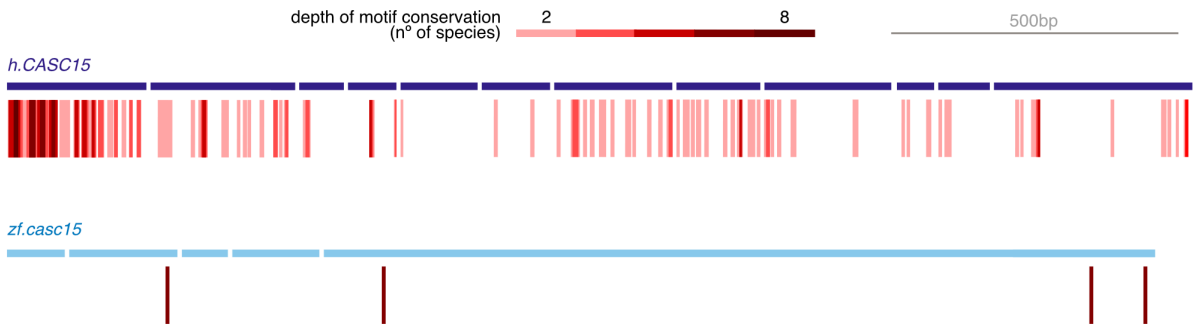


Figure 16. CASC15 orthologs show poor ordered conservation of RNA motifs.
Distribution of the conserved RNA elements (in red) in the exons (in blue) of *h.CASC15* and *zf.casc15* when the whole transcript is considered.

2.9. In cellulo and in vivo rescue experiments with *h.CASC15-short*

Lastly, I sought to establish if *h.CASC15-short* could rescue the zebrafish melanoma progression phenotype and the cellular changes in the migration of 501mel mutant cells. Towards that end, I performed a series of additional rescue experiments both in vivo using *zf.casc15*$^{-/-}$ zebrafish (Fig.17A) and in cellulo utilizing *h.CASC15*$^{-/-}$ cells.

I first generated additional mitfa:*h.CASC15-short* NRAS^{G12D}-containing vectors. The anti-sense transcript was cloned to be used as negative control. By heterologously expressing the human lncRNA in the mutant zebrafish, I expected to recapitulate the results of previously performed rescue experiments (Fig.6D). Despite injected zebrafish being selected for vector insertion, most animals displayed early stages of the melanomagenesis, developing nevi, and only few of them had tumors (Fig.17C). We attributed the poor efficiency of tumor induction to a low-quality tol2 transposase mRNA used during injections.

To examine if the exogenous expression of *h.CASC15-short* from a vector would be sufficient to rescue the increased cell migration of h.CASC15 mutant cells, I performed a pilot experiment and used MS2-tagged plasmids containing *h.CASC15-short* and its antisense counterpart. However, no reduction in cell migration could be identified in *h.CASC15*$^{-/-}$ cells re-expressing *h.CASC15-short* compared to the controls (Fig.17D). Additional experiments are required to determine the transfection efficiency and the appropriate time and concentration of the lncRNA. Likewise, heterologous expression of *zf.casc15* could be performed to determine if conservation of function also occurs at the cellular level.

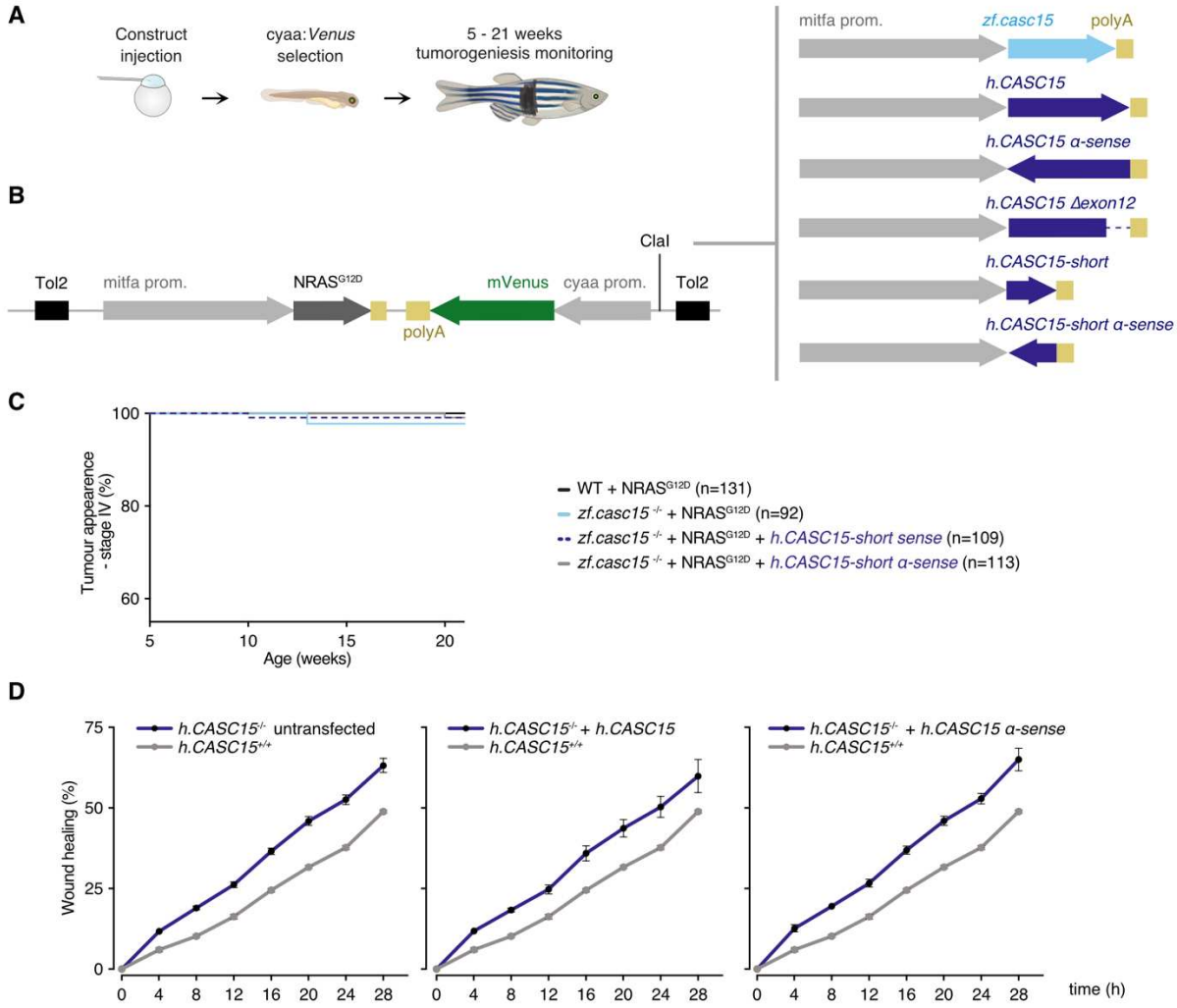


Figure 17. Cross-species rescue experiments in-cell and in vivo require optimization.
(**A**) Schematic of the injection procedure to analyze the impact of lncRNA loss on cutaneous melanoma tumorigenesis. A minimum of 50 individual with YFP positive lens were generated. (**B**) Schematic diagram of the Tol2 construct used to induce melanoma in zebrafish and the rescue constructs that have been cloned. (**C**) Percentage of fish with tumors. Fish were between week 5 and 21 post-injection in wild type and *zf.casc15*-/- background. (**D**) Wound healing assay with *h.CAS15* exogenously expressed. One clone from each *h.CASC15*+/+ or *h.CASC15*-/- genotype was used.

3. DISCUSSION

Twenty years ago, the ENCODE program revealed that a substantial fraction of the human genome is transcribed though less than 2% of this RNA is protein-coding (The ENCODE Project Consortium, 2012). Out of these majoritarian non-coding transcripts, lncRNAs represent one of the most heterogeneous groups. The most recent curation projects, such as FANTOM (v5) and GENCODE (v40), estimate that the human genome contains 27,919 and 18,805 lncRNA genes, respectively (Frankish et al., 2019; Hon et al., 2017), although the numbers vary using less conservative approaches (Uszczynska-Ratajczak et al., 2018). Of these lncRNAs, only a small proportion have been functionally characterized so far and, in most cases, they are not studied under the lens of evolution.

One such lncRNA is *CASC15*, a transcript known to be implicated in melanoma and neuroblastoma (Lessard et al., 2015; Mondal et al., 2018; Russell et al., 2015). As described above, human *CASC15* has syntologous lncRNAs in many vertebrate species (Ulitsky et al., 2011). Despite poor sequence conservation at the nucleotide level, the Shkumatava laboratory described that zebrafish and human *CASC15* orthologs are functionally conserved, as both inhibit melanoma initiation and progression in vivo. My work tackles the origin of the functional equivalence, hypothesizing that both transcripts perform the same function through a conserved molecular mechanism of action.

3.1. *h.CASC15* inhibits cell migration in melanoma cells

To investigate the role of *CASC15* in human melanoma cells, I used reverse genetics to inactivate the lncRNA. Because of the complexity of the locus, I considered several strategies and chose to delete the exon 12 (Fig.8A/Section2.2), aiming to simultaneously remove functional sequences and destabilize the transcript. The removal of the exon caused an overall reduction in the expression of *h.CASC15*, though I observed the presence of a residual transcript. Thus, the deletion of exon 12 generated a hypomorphic allele (Fig.8C). Comparable results were described for *ALAL-1* allele, in which a similar single-exon mutagenesis approach was followed (Athie et al., 2020).

Inactivation of *h.CASC15* caused a consistent increase in melanoma cell migration (Fig.10/Section2.4). This result was coherent with the in vivo phenotype of *zf.casc15* mutant zebrafish, in which an aggravation of the tumor initiation and progression was observed. In an ideal situation, combinatorial methods can be envisioned to corroborate these findings and ensure an RNA-dependent effect in the enhanced migratory behavior (see introduction, Section1.5). One such strategy would be the deletion of the aTSS downstream exon 9, with the aim to achieve a complete inactivation the transcript. Of note, I carried out the design of the allele, and the generation of the additional cell line is ready to be executed.

Additional tests for invasion, proliferation, and clonogenicity behavior on the cell lines I generated revealed that both deleted and *h.CASC15*-unmodified cells behave heterogeneously. Lessard et al. already observed some variability between cell lines as well as heterogeneity dependent on the acute LOF strategy used. In our case, such diversity issue likely derives from an intrinsic cell-to-cell variation of the parental melanoma cell line, exacerbated by our clone picking strategy. To overcome the clone heterogeneity, one could follow recently developed deletion methods, where gene modifications are introduced in cell pools (Meng et al., 2022).

Transcriptionally, *h.CASC15* mutant cells exhibited an upregulation of genes implicated in cell migration and neural crest biological processes (Fig.11/Section2.5). This result was evocative of a neural crest expression pattern in zebrafish initiating melanoma tumors (Kaufman et al., 2016). Although these findings were also consistent with our cell behavior assays, the transcriptional dysregulation observed was rather mild, with relatively few miss-regulated genes. Indeed, *CASC15* might control gene expression in a direct manner, recruiting RBPs and controlling transcription of specific genes. However, one should take into consideration that the observed differences in transcription are late in timing, that is, several passages after *h.CASC15* depletion. Melanoma cells could, for instance, develop compensatory mechanisms that would confound primary regulatory targets of this lncRNA. In fact, direct lncRNA-dependent gene expression changes can occur in minutes upon depletion (Much, Lasda, et al., 2022). To infer the early and direct regulatory targets of *CASC15*, knock-down or inducible knock-out methods could be used. These strategies would shorten the time scale at which the regulatory function of our transcript of interest is studied.

3.2. What drives the expression of *CASC15* orthologs?

Although this work focused on the repercussion of *CASC15* expression and the understanding of its molecular mechanism, the causal determinants of the lncRNA expression are still elusive. As described above, under physiological conditions the human transcript is expressed in the ovary in its entire length, and a shorter isoform is present in the brain (Fig.9/Section2.3). Single cell RNA-seq of human epidermis shows that *h.CASC15* is barely expressed in normal melanocytes (S. Wang et al., 2020). Yet, an increase in expression of its shorter isoform occurs during oncogenesis, in neuroblastoma cells (Mondal et al., 2018) and our studied melanoma cells. Which mechanism powers the transcription of the *h.CASC15-short* specifically in ectoderm-derived tissues upon oncogenesis is a question that remains to be answered.

One plausible explanation for the differences in isoform expression might come from a cell-specific chromatin reorganization of the locus. To evaluate DNA interactions, I inspected available HiC datasets of different human cell lines with the help of Nicolas Servant (Institut Curie) (Banerjee et al., 2020; Sanborn et al., 2015; The ENCODE Project Consortium, 2012). In neuroblastoma cells, the TSS of the short *h.CASC15* isoform seems to be located in close proximity to the SOX4 locus, which might impact its expression (Fig.18). The low-resolution data on normal keratinocytes and RMPI melanoma cells prevented from inferring the three-dimensional organization of these region. An in-depth study of chromatin accessibility and organization of the locus upon neuroblast and melanocyte oncogenic transformation would provide crucial insights for the understanding of the transcriptional activation of *h.CASC15*.

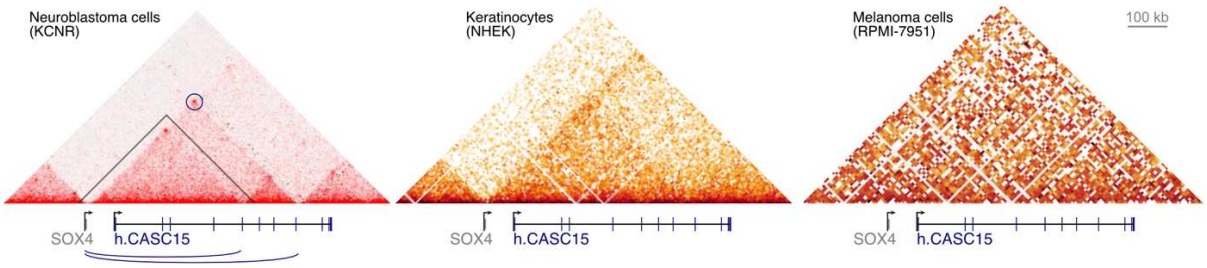


Figure 18. Genome organization of CASC15 locus.
HiC profile of different cell lines in SOX4 and *CASC15* locus. Adapted from Banerjee et al. or visualized through HiGlass. Black square pinpoints a cis-interaction between SOX4 and a known enhancer in CASC15 locus. The black circle highlights a contact that might be relevant for the expression of *h.CASC15-short*. Blue arches help visualize the topologically associated domains on the linear DNA.

Similarly, *zf.casc15* is mainly expressed during development and in the ovary, but not in the brain (Fig.9/Section2.3). Because no zebrafish homozygous for the deletion of the 3' part of the transcript could be obtained (Lavalou et al., 2019), *zf.casc15* might have an important role during development. The expression of the transcript cannot be traced in publicly available single cell RNA-seq datasets of zebrafish early development (Farnsworth et al., 2020) since *zf.casc15* is not annotated. However, re-

analysis of this and other zebrafish melanocyte RNA-seq data (Kramer et al., 2022; Venkatesan et al., 2018) might be of interest to describe the basal state of *zf.casc15* expression in this cell type.

Inspired by the analysis done with *DIRC3* lncRNA (Coe et al., 2019), I wondered whether the CASC15 locus harbored some important DNA elements that could reveal transcriptional regulation by melanocyte-related TFs. ChIP-seq data in 501mel cells (Fontanals-Cirera et al., 2017; Laurette et al., 2015) revealed some binding sites for MITF and SOX10 TFs near peaks of H2K27ac and H2k4me1, indicating their potential involvement in transcriptional control (Fig.19). Whether these TFs modulate the expression of *h.CASC15* and whether these TFs binding profiles are similar in zebrafish remains to be tackled and could be suggestive of other conserved properties of this locus.

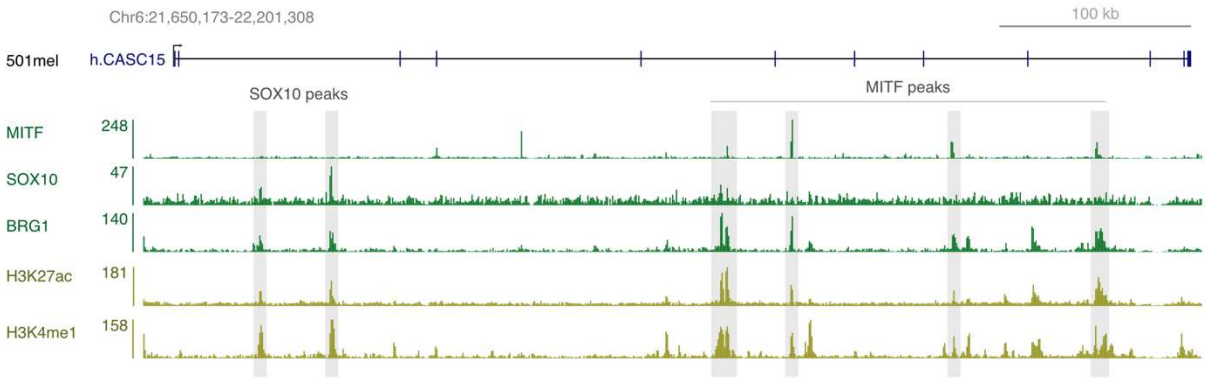


Figure 19. Localization of enhancers in the human CASC15 locus
UCSC genome browser view showing that the *CASC15* locus contains CHIP-seq picks melanocyte transcription factors and histone marks in 501mel cells

3.3. Identification of functional *CASC15* protein interactors

In order to address the molecular mechanism of action of *CASC15* orthologs in a melanoma framework, I aimed at identifying RBPs interacting with *zf.casc15* and *h.CASC15* which could ultimately underlie their conserved functionality. Several methods have been developed in order to identify RNA-protein interactions, and their advantages and caveats have been extensively discussed (Barra & Leucci, 2017; Chu, Spitale, et al., 2015; McHugh et al., 2014; Ramanathan et al., 2019). In order to perform an RNA-centric interactome discovery while overcoming the low efficiency of RNA-pulldowns, I performed incPRINT. This high-throughput method relies on FLAG-tagged protein immunoprecipitation, where RNA-RBP interactions are quantitatively measured using a luciferase detector (Graindorge et al., 2019; Sabaté-Cadenas & Shkumatava, 2020).

Through the characterization of zebrafish and human RNA-interacting proteomes and a cross-species comparison, I identified a set of fifteen proteins as common binders of the *CASC15* orthologs (Fig.12/Section2.6). Based on these results, I explored if the binding of any of these proteins was

sufficient to drive *CASC15* function. To do so, I performed a side-by-side knockdown of each RBP in melanoma cells and characterized their migration. I observed that knockdown of DDX50 and DUSP12 was sufficient to cause an increase in cell migration (Fig.12C) and were considered thereupon functional *CASC15*-interactors. A caveat of this screen is the individual depletion of the proteins, which would become false negatives in the case of compensatory mechanisms. Using incPRINT I could identify region-specific interactions for some of the *h.CASC15* protein partners though optimization is required to understand in detail the binding of DDX50 and DUSP12. Finally, the RNA-protein interactions detected by incPRINT depend on the exogenous expression of both the lncRNA and the RBP. To examine if the binding occurs in melanoma cells under endogenous conditions, I am currently finalizing RIP experiments of both DDX50 and DUSP12 (Fig.13,14).

3.4. How do *CASC15* protein partners participate in melanoma inhibition?

Whereas DDX50 and DUSP12 interact with our transcript of interest, the precise mechanism by which these RBPs mediate *CASC15* melanoma inhibitory function remains to be explored. During the coming months, our efforts will be directed toward a precise understanding of this mechanism. Nevertheless, a systematic revision of recent literature already provides some functional insights.

DDX50 is a DEAD-box RNA helicase, also known as RH-II/Guβ (Valdez et al., 2002). Members of the DEAD-box class are present in all domains of life and share a common helicase core composed of two domains homologous to a bacterial single-stranded DNA binding protein, RecA (Caruthers et al., 2000; Story et al., 2001; N. Yao et al., 1997). DDX50 shares 63% of amino acid sequence identity with its paralog helicase, DDX21 or RHII/Guα. These two RNA helicases are the only members of the Gu family, containing a GUCT domain (Marcaida et al., 2020). Of note, the zebrafish genome contains only Ddx21. The two RecA domains of these three helicases are highly conserved, presenting between 70 and 80% of amino acid identities.

Contrary to other RNA helicases of the same family, DDX50 does not show RNA folding or unwinding activity in vitro (McRae et al., 2017; Valdez et al., 2002). Some DDX helicases are believed to participate in the formation of RNA-protein complexes and provide stability to quaternary structures (P. Linder & Jankowsky, 2011; Singh et al., 2015). Their unstructured C-terminal domains might as well mediate their enzymatic activity and protein/RNA interactions, and are composed of different repeats: `R-G-S` for DDX50, `F/P-R-G-Q-R` for the human DDX21, and `G-N-R-S` for the zebrafish Ddx21.

Notably, DDX21 has been shown to interact with another lncRNA through the RecA domain and control ribosome biogenesis by regulating RNA Polymerase I in ovarian and cervical cancer cell lines (Calo et al., 2015; M. Wu, Xu, et al., 2021; Xing et al., 2017). This helicase also has multiple functions both in human melanoma cells and zebrafish melanocytes. DDX21 can sense nucleotide stress through

the progesterone receptor (PGR) and regulate gene expression pre-transcriptionally, binding to chromatin, and post-transcriptionally, engaging mRNA (Santoriello et al., 2020). Mass spectrometry also revealed zebrafish Ddx21 as one of the main interactors of the metastasis-associated phosphatase of regenerating liver 3 (Prl3a), an interaction conserved in humans (Johansson et al., 2020). PRL3 was shown to regulate DDX21 binding and distribution along chromatin, impacting transcription. In vivo, zebrafish embryonic melanocytes were not impacted upon *ddx21* knockdown, though melanocytic stem cell regeneration was affected (Santoriello et al., 2020). The depletion of this helicase was also found to impact neural crest cells both in *Xenopus* and zebrafish (Calo et al., 2018). Importantly enough, to support gene activation of its targets, DDX21 has been shown to bind c-JUN (Calo et al., 2015; Holmström et al., 2008; Westermarck et al., 2002), a major regulator of melanoma progression (Kappelmann-Fenzl et al., 2019)

If DDX50 modulates transcription in a similar manner, chromatin immunoprecipitation and sequencing (ChIP-seq) analysis (Pérez-Rico et al., 2020) in our DDX50-HA 501mel cells would reveal which genes are regulated by DDX50 (Fig.20A). *CASC15* might also drive the interaction of DDX50 with other proteins to form a macromolecular complex. Because other proteins have been known to be bound to DDX50 (Pallett et al., 2022), a mass spectrometry study of the DDX50 co-precipitated proteins could help determine novel functions of this RNA helicase. (Fig.20B).

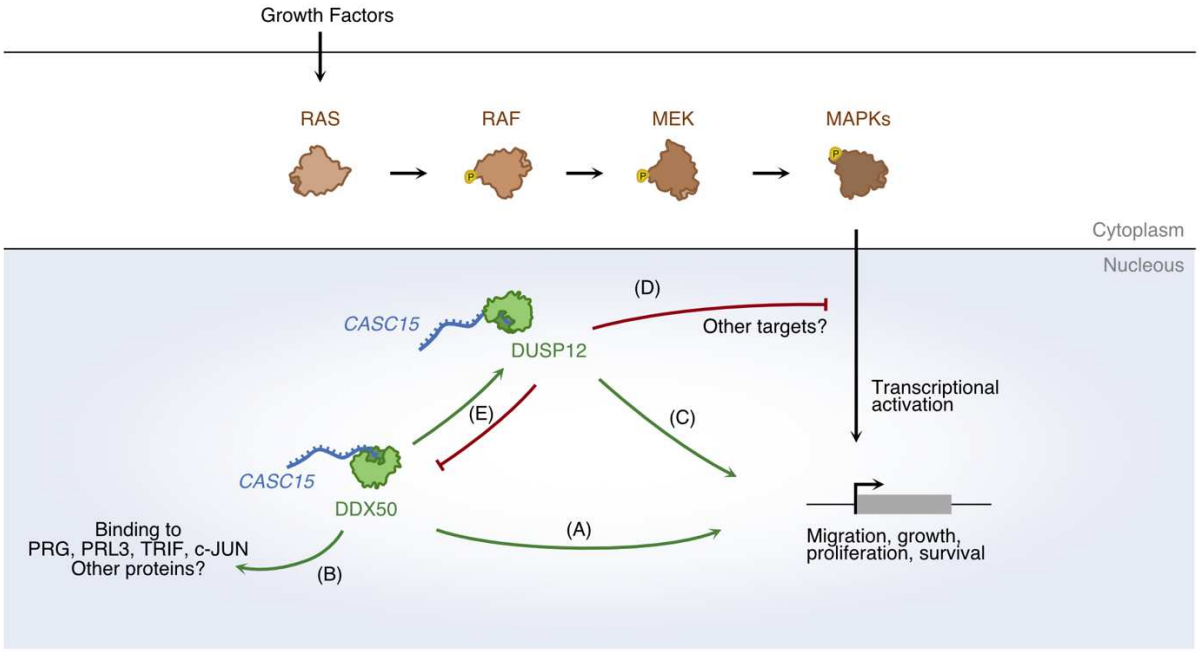


Figure 20. Model for the mechanism of action of *CASC15* in a melanoma context.
The majority of melanoma patients have mutations in kinases of the MAPK pathway (brown) that remains constitutively activated and promotes transcriptional. Illustrated are possible mechanistic models for DDX50 and DUSP12 function: (**A**) CASC15 might regulate transcription of specific genes via directing the binding of DDX50 to chromatin or (**B**) in an indirect manner, facilitating the binding to other proteins. (**C**) Like (A), DUSP12 co-transcriptional regulation may occur through binding to DNA via C2H2 domain. (**D**) DUSP12 could regulate in a phosphorylation-dependent or independent fashion the activity of MAPKs and other pathways. (**E**) *CASC15*, DUSP12, and DDX50 might cooperate, via phosphorylation, ribonucleoprotein complex stabilization, or subcellular localization guiding.

The second functional *CASC15* interactor I identified is the dual specificity phosphatase 12, or DUSP12. Also referred to as hYVH1, this protein belongs to an evolutionarily conserved family of enzymes that contains a common phosphatase domain, whose catalytic site is composed of conserved `D`, `C`, and `R` amino acid residues (Muda et al., 1999). As the 'dual' name pinpoints, DUSPs are protein tyrosine phosphatases (PTP) also capable of dephosphorylating serine and threonine residues. DUSP12 is primarily localized in the nucleus and has been found to participate in ribosomal maturation among other processes (Lo et al., 2010). Besides, genome-wide association studies identified DUSP12 as a low-risk susceptibility locus for neuroblastoma (Nguyễn et al., 2011), another cancer type where *CASC15* is found highly expressed.

On one hand, DUSP12 is the only phosphatase known to harbor a C2H2 zinc-binding domain (Muda et al., 1999), known to interact both with DNA and RNA (Iuchi, 2001). Indeed, DUSP12 has been found to assist in ribonucleoprotein complex formation (Geng et al., 2017). On the other hand, this protein belongs to the class of 'atypical' dual phosphatases as it does not contain the domain that targets MAPKs in its N-terminus, known as Rhodanese or Cdc2 homology 2 domain, which carries the MAP kinase-binding or kinase-interaction motif (C. Y. Huang & Tan, 2012; Jeffrey et al., 2007). Despite predicted to dephosphorylate other substrates, it has been suggested that atypical DUSPs can regulate MAPKs in absence of this binding domain (Alonso et al., 2003; Cho et al., 2017; Marti et al., 2001; Rahmouni et al., 2006; Zama et al., 2002).

Because DUSP12 contains a zinc finger domain, it is reasonable to hypothesize that the protein might participate to transcriptional regulation through chromatin binding activity. Thus, ChIP-seq analysis could also be applied to study the DNA targets of this protein (Fig.20C). Regarding its catalytic activity, since ~30% of the proteins of eukaryotic cells are thought to be phosphorylated, an educated guess of the cellular targets of DUSP12 does not seem feasible. Phospho-specific antibodies coupled with mass spectrometry could be employed to identify which proteins are post-translationally modified by our phosphatase of interest (Fig.20D).

3.5. Are DDX50 and DUSP12 cooperating to drive the function of *CASC15*?

Beyond their independent function, another question worth highlighting is if DDX50 or DUSP12 regulate each other (Fig.20E). Co-immunoprecipitation experiments would assess if a physical interaction exists between the two RBPs. Importantly, DUSP12 has been shown to bind DDX50 in breast cancer cells (Falcão Monteiro & Forti, 2019). If true, a plausible hypothesis could be that *CASC15* serves as a platform for this interaction. Whether the binding of these RBPs is *CASC15*-dependent could be tested using our 501mel *h.CASC15*-depleted cell line. It is also tempting to speculate that DUSP12, or even one of the MAPKs, may target DDX50. In fact, several residues of this helicase are

phosphorylated in vivo (Hornbeck et al., 2015), and the phosphorylation of its paralog, DDX21, has been shown to regulate its nucleolar localization (Calo et al., 2018; Mialon et al., 2008).

In this sense, *CASC15* may control the spatial distribution of DUSP12 and DDX50 in the cell. The lncRNA-dependent recruitment of these RBPs could impact their nucleolar/nucleoplasm localization or direct their location in the chromatin for a target-specific function, as has been shown for other nuclear lncRNAs (Unfried & Ulitsky, 2022). Alternatively, *CASC15* might exert its function through the control of the abundance of the RBP interactors, affecting the availability of DDX50 or DUSP12 for functional binding elsewhere (M. E. Bordeleau et al., 2006; Elguindy & Mendell, 2021). Immunofluorescence could be coupled with RNA fluorescence in situ hybridization for co-detection of RBPs and *h.CASC15* lncRNA (McHugh et al., 2015).

3.6. What is the extent of the mechanistic conservation between *CASC15* syntologs?

To confidently state that the molecular mechanism of action is conserved between zebrafish and human *CASC15* orthologs, more experiments are required to fully characterize *zf.casc15*. Whether the zebrafish lncRNA subcellular localization differs from the human, like *FAST* or *FIRRE* orthologs (Guo et al., 2020; Much, Smallegan, et al., 2022), remains to be described. With regards to the protein interactors, one limitation of this study is the use of human RBPs for the identification of the interactome of lncRNAs from other species (Graindorge et al., 2019). Because zebrafish lack a DDX50 ortholog, the in vivo interaction between *zf.casc15* zebrafish Ddx21 paralog still needs to be studied. The same holds true for zebrafish dusp12. Cloning the zebrafish proteins into the pCDNA3.1 FLAG-tagged plasmid and testing the interaction of zebrafish orthologs by incPRINT would help address this caveat. Additionally, assessing the migratory behavior of 501mel cells upon depletion of DDX21, also found as a top *h.CASC15* interactor (Fig.15), could tackle the differences between the two paralogs in human cells.

3.7. Which feature underlies the conserved protein interactions?

Following the identification of conserved RBP interactors, I wondered how *zf.casc15* and *h.CASC15,* two positionally-equivalent lncRNAs with modest primary sequence conservation, can be bound by the same proteins. Given that the RNA-bound proteome interacts with short, non-complex RNA sequence elements, I hypothesized that the two transcripts are comparable in terms of their constituent protein binding sites. To test that, I followed an RNA motif discovery approach.

To perform a nonlinear comparison of lncRNA sequences and to detect similarities in k-mer content, several algorithms can be used (Ross & Ulitsky, 2022). The Multiple EM for motif elicitation software, or MEME (Bailey et al., 2009), was initially designed for TF motif discovery and has been recently

shown to distinguish lncRNA sequence motifs responsible for RBP interaction (Jin et al., 2021). Similarly, the Sequence evaluation through k-mer representation algorithm, or SEEKR (Kirk et al., 2018), compares and quantifies patterns of k-mer abundance using a statistical approach without considering the positional relationship of lncRNAs. By utilizing this in silico method, similar RNA motifs were detected between *Xist* and *Rsx* (Sprague et al., 2019) and the analog of the human *EVX1AS* lncRNA could be identified among lizard lncRNAs (Olazagoitia-Garmendia et al., 2022).

Recently, lncLOOM algorithm was developed in order to identify primary sequence motifs that appear in the same order in sequences from orthologous lncRNAs (Ross et al., 2021). The framework identified the primary sequence motif `AAAUGGA` was conserved in *CHASERR* syntologs and was found responsible for the binding of DHX36 protein. Because numerous species have been found to harbor a lncRNA loci downstream SOX4 protein coding gene, constituting *CASC15* syntologs (Ulitsky et al., 2011), I used lncLOOM for the identification of functional RNA motifs (Fig.16/Section2.8). This strategy revealed unsuccessful at detecting conserved k-mers between *h.CASC15* and *zf.casc15*. Yet, the presence of conserved RNA sequence elements that lncLOOM might have overlooked should not be discarded.

LncLOOM's assumption for linear order conservation of the k-mers might not hold for our transcript of interest. Some lncRNA might function in a modular fashion, like *bags of motifs* that do not require precise positioning. Lessons can be learned from the *billboard model* for TF binding motifs, which postulates that regulatory functions of enhancers are independent of the relative order of the TF binding motifs (Kulkarni & Arnosti, 2003). In fact, the conservation of order of TF motifs can change while retaining enhancer functionality (Wong et al., 2020). Additionally, it is important to consider that RBP binding sites are deleterious, that is, they are permissive to sequence variation. Thus, the requirement for a perfect conservation of motif sequence and order might be too constraining for *CASC15* syntologs. Finally, because the functional conservation was demonstrated only between the zebrafish and human transcripts, one could argue that the other syntologous lncRNAs might have lost certain RNA motifs and together with them, the melanoma inhibitory functions of *CASC15*, consequently confounding lncLOOM search.

In conclusion, other algorithms might be required to detect conserved RNA motifs in *CASC15* syntologs. The secondary structure determination of human and zebrafish *CASC15* might as well contribute to the definition of key conserved modules responsible for protein interactions. In this respect, incPRINT could be useful for mutation-driven RNA motif discovery helping assign RBP interactions to specific RNA sequences. Similar to what was done for *h.CASC15* structure-function analysis (Fig.15), conserved k-mers could be mutated in incPRINT constructs for an exhaustive structure-function analysis and the ultimate decryption of the non-coding 'RNA code' for protein interaction (Auweter et al., 2006; Rinn & Chang, 2020).

3.8. Relevance and perspectives

LncRNAs are untapped therapeutic targets and might offer opportunities for novel drugging strategies. This is especially relevant for oncogenic or cancer-promoting transcripts. The triple helix structure of *MALAT1* has been shown to be targetable leading to destabilization and downregulation of the lncRNA (Donlic et al., 2018). Targeting *SAMMSON* lncRNA has also been shown to suppress the growth and proliferation of a patient-derived xenograft melanoma model (Leucci et al., 2016). But despite being regarded as drug targets for a long time, no lncRNA-targeting drugs have entered clinical trials yet (Winkle et al., 2021).

In the new era of RNA medicine (W. Xie et al., 2021), the potentiality of using *CASC15* transcript as an RNA therapeutic to slow down melanoma progression cannot be understated. An outstanding example of lncRNA-based therapy is *HULC*. This transcript was found to facilitate the activity of phenylalanine metabolism enzymes, and the treatment with a *HULC*-based RNA therapeutic agent reduced excessive phenylalanine in a mouse phenylketonuria model (Y. Li et al., 2021). In this sense, given the melanoma-inhibitory function of *CASC15*, one could envision a *CASC15* RNA-based therapy topically administered (Sallam et al., 2021) while awaiting resection of the tumor, that could be used as well as a preventive treatment. Admittedly, this strategy would have to be evaluated in melanoma mouse models (Dankort et al., 2009).

Another feasible approach to slow down melanoma progression would be to target *CASC15* – RBP interactions. This is well illustrated by *GAS5* lncRNA and UPF1 protein, whose interaction induces the degradation of the transcript. Blocking the binding of this RBP increased the half-life of *GAS5*, ultimately stimulating the insulin receptor and inducing glucose uptake (Y. Shi et al., 2019). In light of our findings, a compound could theoretically inhibit melanoma progression by stimulating the binding between *CASC15* and its protein partners DDX50 or DUS12. Such RNA-RBP interaction stabilizers have already been described for other RNA helicases and splicing factors (M.-E. Bordeleau et al., 2005; Sivaramakrishnan et al., 2017). In addition, the absence of *CASC15* expression in normal melanocytes would ensure a target-specificity of this strategy. As we will discuss in deep in the following chapter, the targeting of interactions between RNA and proteins is emerging as a promising therapeutic scenario.

CHAPTER 2

Targeting the m^{6}A reader YTHDF2 with small molecules

4. INTRODUCTION

Dysregulated RNA-protein interactions have been associated with or, more importantly, are causative agents of many types of human pathologies, including genetic diseases (Gebauer et al., 2021), cancer (Minuesa et al., 2019; Paris et al., 2019), autoimmune disorders (Fitzgerald & Kagan, 2020), kidney diseases (Seufert et al., 2022) and neurodegenerative diseases like spinal muscular atrophy (Hua et al., 2011), amyotrophic lateral sclerosis (R. H. Brown & Al-Chalabi, 2017) or myotonic dystrophy (Udd & Krahe, 2012). Similarly, in order to facilitate the viral life cycle, genomic RNAs from viruses are known to associate with RBPs. These proteins can be produced from the viral genome, like the human immunodeficiency virus (HIV) transactivation response element (*TRE*) – Tat interaction (Dingwall et al., 1989) and the Rev responsive element (*RRE*) – Rev interaction (Malim et al., 1989), or have a host origin, like the case of SARS-CoV-2 (Flynn et al., 2021; N. Schmidt et al., 2021).

Despite initial skepticism, the therapeutic target potential of the RNA binding and modifying machinery has attracted the attention of the drug discovery community (Boriack-Sjodin et al., 2018; Julio & Backus, 2021; Mohibi et al., 2019; P. Wu, 2020). In this sense, RNA modifications, which often fall into non-coding regions of transcripts, and their orchestrating proteins have become appealing targets in recent times.

4.1. N^6-methyl-adenosine transcriptional modification

N^6-methyl-adenosine (m^6A) is the most abundant RNA post-transcriptional modification within protein-coding and long non-coding RNAs in eukaryotes (Desrosiers et al., 1974). The reversible methylation of diverse cellular transcripts adds another layer of sophisticated regulation to the RNA primary sequence. In fact, m^6A modification is involved in almost all aspects of the metabolism of different RNA species, including processing (Haussmann et al., 2016; Pendleton et al., 2017; Xiao et al., 2016), miRNA maturation (Alarcón et al., 2015), transport (Roundtree et al., 2017; G. Zheng et al., 2013), stability (Vu et al., 2017; X. Wang et al., 2014; Zhao et al., 2017), and translation efficiency (Barbieri et al., 2017; X. Wang et al., 2015; J. Zhou et al., 2015).

Mapping of transcriptome-wide m^6A distribution revealed that this mark is enriched in the 3'UTRs near the stop codon (Dominissini et al., 2012; Meyer et al., 2012). The m^6A moiety is added to the conserved [G/A/U] [G>A] m^6A [C] [U>A>C] motif (Fu et al., 2014), also termed DRACH, in a specific manner: most m^6A -modified RNAs contain only a single m^6A, despite most transcripts having many potentially methylated DRACH sequences. In addition, despite the wide prevalence of DRACH elements, only specific RNAs acquire a modification (B. Linder et al., 2015; Schwartz et al., 2014). However, the site- and transcript-specific elements that dictate m^6A deposition remain mostly unknown.

Contrarily, the m^6A modulation mechanism is tightly regulated by RBPs that have been well-characterized (Fig.21) (H. Shi et al., 2019). This modification is reversibly installed by the methyltransferase METTL3/METTL14 multi-subunit 'writer' complex (reviewed in Zaccara et al., 2019), and oxidatively removed by m^6A demethylases, FTO, and ALKBH5, named 'erasers' (Jia et al., 2011; G. Zheng et al., 2013). The precise effects of RNA methylation are determined by the 'reader' proteins, most of them containing YT521-B homology domain (YTHD): YTHDC1, acting in the nucleus, and YTHDC2 and YTHDF1-3 present in the cytoplasm. The YTHDF family behaves redundantly, inducing destabilization and degradation of the same subset of target transcripts (Lasman et al., 2020; X. Wang et al., 2014; Zaccara & Jaffrey, 2020).

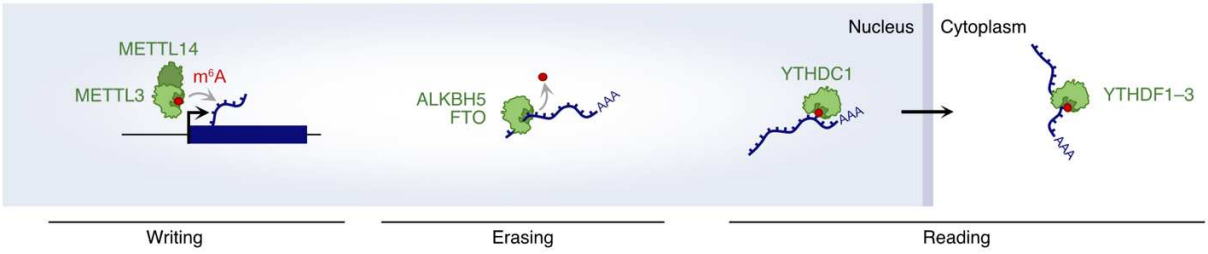


Figure 21. Life cycle of m^6A mark.
m^6A is deposited co-transcriptionally mainly by a writer multiprotein complex (which includes METTL3, METTL14, WTAP, VIRMA, CBLL1, ZC3H13 and RBM15/15B) and removed by 'eraser' demethylases. In the nucleus, the m^6A residue is recognized by YTHDC1. Upon transport to the cytoplasm, m^6A transcripts can be bound by different readers, namely YTHDF1–3, YTHDC2, IGF2BP1–3, eIF3, FMRP and the HNRNP family.

It has been described that m^{6}A increases the affinity of YTH domain-containing proteins to the RNA binding site (Theler et al., 2014), which was determined to be up to ten-fold higher compared to unmethylated targets in vitro (T. Zhu et al., 2014). The binding of reader proteins can also result after local structure conformation changes of the RNA molecule upon m^{6}A -modification, also coined 'm^{6}A -switches', which can displace RBPs or increase their accessibility and subsequent recruitment (Alarcón et al., 2015; N. Liu et al., 2015).

Previous work has demonstrated that m^{6}A modification of RNAs plays a pivotal role during zygotic and embryonic development (M. Li et al., 2018; Mendel et al., 2021; Zhao et al., 2017). Another evolutionarily conserved m^{6}A-dependent process is the regulation of meiosis and fertility: the loss of RNA methylation enzymes has an impact on gametogenesis in mice (Lin et al., 2017), zebrafish (Xia et al., 2018), *Drosophila* (Hongay & Orr-Weaver, 2011) and *Arabidopsis* (Zhong et al., 2008). Likewise, m^{6}A erasers and readers control murine spermatogonia and oocyte growth and maturation (Wojtas et al., 2017; G. Zheng et al., 2013). Besides germline development, the m^{6}A pathway has been linked as well to tumorigenesis, tumor invasion, and metastasis (Delaunay & Frye, 2019). One disorder in which m^{6}A regulator proteins are ubiquitously implicated is acute myeloid leukemia (AML).

4.2. YTHDF2 – m^{6}A recognition as a therapeutic target for acute myeloid leukemia

One of the best-studied m^{6}A -regulated diseases is AML, an aggressive and highly prevalent adult blood cancer. Leukemic stem cells (LSC) have unlimited self-renewal potential and are considered to be the root cause of treatment resistance. Even though most patients initially respond to chemotherapy, relapse is frequent and more than 70% of them die within five years from initial diagnosis (Döhner et al., 2017). Therefore, the identification of novel and specific therapeutic targets for the elimination of LSCs is an unmet clinical need.

In the past decade, virtually all members of the m^{6}A methylation program have been implicated in myeloid leukemogenesis: writers (Bansal et al., 2014; Barbieri et al., 2017; Weng et al., 2018), readers (Y. Cheng et al., 2021; Elcheva et al., 2020; Paris et al., 2019), and erasers (Z. Li et al., 2017; Shen et al., 2020) have an impact in myeloid progenitor differentiation and promote or suppress malignant hematopoiesis. m^{6}A regulators are therefore a source of therapeutic targets in AML. Since these proteins might also be crucial for other biological processes, the selection of a target RBP that specifically impacts LSCs is the next major challenge.

Latest studies highlight the efforts made in hijacking m^{6}A in order to fight AML. Two inhibitors have been identified to efficiently target FTO and display an anti-leukemic effect with very low toxicity in mice (Su et al., 2020). In addition, a first-in-class inhibitor of METTL3 also proved a potent anti-tumor activity in AML models (Yankova et al., 2021). Nevertheless, it should be stressed that this

methyltransferase has previously been characterized as essential for normal hematopoietic stem cell differentiation and that Mettl3 KO causes bone marrow failure in mice (Y. Cheng et al., 2019; H. Lee et al., 2019; Vu et al., 2017).

Recently, YTHDF2 has been shown to be required for disease initiation and propagation. Genetic inactivation of this RNA binding protein selectively compromises LSC in vivo while preserving normal hematopoiesis (Paris et al., 2019), making YTHDF2 a unique therapeutic target (Fig.22). Despite the key function of YTHDF2 in AML, no specific YTHDF2-targeting drugs have been yet identified.

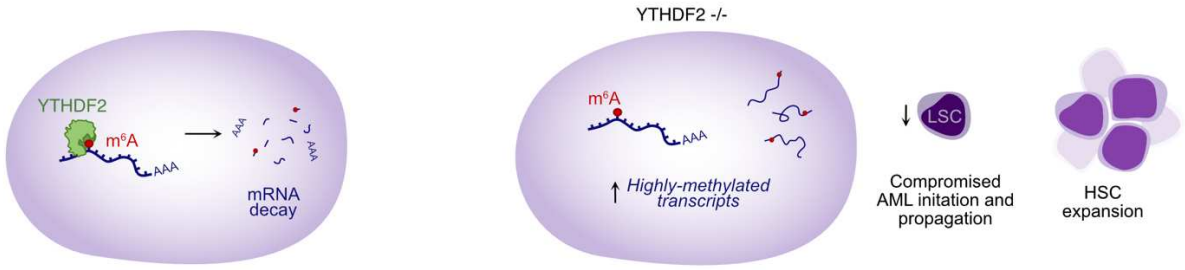


Figure 22. YTHDF2 model.
In absence of YTHDF2, m^{6}A-modified transcripts are stabilized. YTHDF2 is required for disease initiation, and its depletion selectively compromises leukaemic stemm cells (LSC) while promoting normal hematopoietic stem cell (HSC) expansion.

4.3. The complexity of RBP therapeutic targeting

The term UNDRUGGABLE describes a biomolecule that cannot be pharmacologically targeted. Recent examples, though, demonstrate that more appropriate terms for these complex targets are difficult-to-drug or yet-to-be-drugged (Dang et al., 2017; Henley & Koehler, 2021). In the past, RBPs have frequently been deemed undruggable (Gashaw et al., 2011; Hajduk et al., 2005) as they (i) possess a significant fraction of highly disordered regions (Järvelin et al., 2016); (ii) often show considerable structural similarity between individual members of their own family, like the case of the m^{6}A machinery (reviewed in Huang & Yin, 2018); and (iii) present large contact surfaces and lack enzymatic pockets or fold cleft-like motifs typically targeted by inhibitors (Gebauer et al., 2021; Julio & Backus, 2021; P. Wu, 2020). All these features make RBPs very unappealing targets and limit their tractability.

Despite important advances in the drug discovery field, to date, only a handful of RBPs have been described to be directly targeted by SMALL MOLECULES , drugs smaller than 100 atoms. This list includes the translation initiation factor eIF4A (M.-E. Bordeleau et al., 2005), the human antigen R protein (Meisner et al., 2007), the splicing factor SF3b (Kotake et al., 2007), the pattern recognition receptor TLR3 (K. Cheng et al., 2011), MUSASHI proteins (Minuesa et al., 2019), and miRNA-binding LIN28 proteins (reviewed in P. Wu, 2020). Furthermore, the mechanism for RNA binding inhibition or stimulation of these compounds is often dissected in an indirect fashion, and none of them have gained regulatory approval so far. As an alternative to this challenging strategy, relying on an inhibition by an

RNA-competitive binding mechanism, is the blockage of RBP conformational changes required for RNA recognition. The identification and targeting of an allosteric RBP pocket has been proven to be successful for drugging TLR8 receptor to suppress proinflammatory signaling (S. Zhang et al., 2018).

New chemical approaches provide additional opportunities for overcoming the undruggable nature of RBPs. These novel perspectives acknowledge (i) RNAi therapeutics as a reliable drug system, (ii) RNA transcripts as small molecule targets and (iii) the expansion of the chemical space by proving the interaction interface upon the formation of RNA-protein quaternary complex.

4.4. RNA-based therapies targeting ribonucleoprotein complexes

RNA-based interventions have brought hope to the drug development field due to their ability to engage targets that are considered undruggable by classical small molecules (Setten et al., 2019; Winkle et al., 2021). The two main RNA therapeutic approaches, small interfering RNAs (siRNAs) and ASOs, differ in their mechanisms of action. While the former takes advantage of the endogenous miRNA pathway, the latter relies on RNAse H1 cleavage to cause mRNA degradation or acts as a splicing modulator. That is, targeting the complex formed between pre-mRNA and proteins.

Since the discovery of RNAi in *Caenorhabditis elegans* (Fire et al., 1998), alongside the findings showing synthetic siRNAs could mediate targeted gene silencing in mammalian cells (Elbashir et al., 2001), clinical development of RNA therapies has undergone major progress. As of June 2022, eight ASO-based and five siRNA-based therapeutics have been approved by FDA or EMA, and two dozen more RNA drug formulations are in phase II or III clinical development (Winkle et al., 2021). Two prime examples are patisiran (Adams et al., 2018; Coelho et al., 2013), the first-of-its-kind siRNA-based therapy (Fig.23A); and nusinersen (Hua et al., 2010; Mercuri et al., 2018), a splice-switching ASO that constitutes the first drug to treat spinal muscular atrophy (Fig.23B). These cases of success confirm the clinical utility of RNA therapeutics.

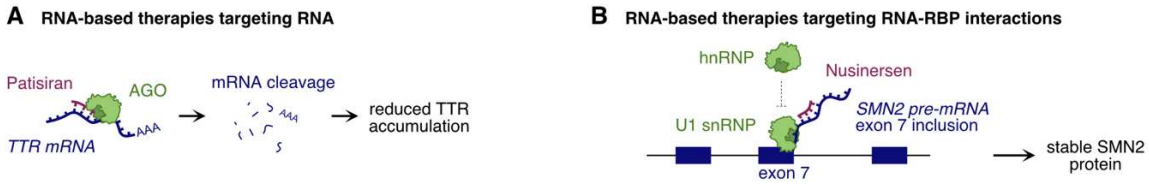


Figure 23. Examples of RNA-based strategies.
(**A**) Patisiran is a 21-nt siRNA that targets transthyretin (TTR) mRNA and prevents its detrimental accumulation (2018, EMA). (**B**) Nusinersen is an 18-mer ASO splicing modulator. It prevents survival motor neuron 2 (SMN2) exon 7 exclusion by occlusion of the splice repressor hnRNPs and the pre-mRNA binding (2016, FDA).

The major advantage of RNA-based therapies is the reliance on nucleotide complementarity for gene silencing. In practical terms, this feature expands the prospective targets to roughly any expressed RNA transcript. In addition, the design and production of lead compounds for novel targets is straightforward, which shortens the time from target identification to preclinical studies, a feature especially attractive for ultra-rare or n-of-1 diseases. This fast-development potential is well-illustrated by milasen, an ASO therapy that could be designed and administered in less than a year after the identification of the patient's mutation (Kim et al., 2019).

The two main roadblocks for RNA therapeutics to become the gold standard treatment lie in (i) their toxicity, concerning immunogenic reactions and molecular, cellular, and tissue off-target toxic effects, (extensively discussed in Setten et al., 2019); and (ii) their delivery, despite significant advancements in recent years, thoroughly reviewed elsewhere (Paunovska et al., 2022; Roberts et al., 2020). Compared to RNA agents, whose pharmacokinetic and pharmacodynamic properties still pose these considerable limitations, small molecules offer up until now unique possibilities to directly target RNA or the interface of ribonucleoprotein quaternary structures.

4.5. RNA is an effective small molecule target

As opposed to RNA-based therapies, small molecules have overall convenient chemical and biophysical tunability, lower toxicity, superior stability, and suitable cellular and tissue permeability. RNA targeting potential has been under-explored until very recently, mainly owing to its four-nucleotide, highly charged, hydrophilic composition (Disney, 2019; Ursu et al., 2019; Warner et al., 2018). However, the last two decades have witnessed encouraging progress and further studies have demonstrated that RNA can be targeted by small molecules.

One such example is HIV: the interactions between *TRE* RNA loop and Tat protein (Mei et al., 1998) and *RRE* RNA element and Rev protein (Kjems et al., 1991) were identified as targets of small molecule inhibitors. Nevertheless, stem-loop motifs are secondary structures that typically contain small or shallow grooves, features that do not meet the common standard of ligand-binding pockets. By contrast, rigorous characterization of RNA structures has revealed druggable pockets for some RNA species. For instance, ribocil targets the bacterial flavin mononucleotide riboswitch (Howe et al., 2015), and X1 is able to recognize the structural *Xist* A-repeat and arrest RBP binding (Aguilar et al., 2022) (Fig.24A). Last, linezolid binds the large subunit of bacterial ribosomal RNA (Wilson et al., 2008), and constitutes the only approved drug class that binds RNA alone to date. The compounds described, however, accommodate complex, buried, cleft-like motifs that might not be present in other RNA transcripts.

Structural information of the RNA of interest, together with computational tools, can accelerate the rational design in drug discovery. More precisely, structure-driven strategies can predict, identify, and even optimize small molecules targeting RNA motifs of interest using 3D structural data (Davidson et

al., 2009; Stelzer et al., 2011), though posterior validation through classical binding assays is required. Nevertheless, these computer-aided methods depend entirely on the availability of the target structure, which is rarely documented in the RNA world. Similarly, in silico approaches can integrate ligand information for virtual screening of novel targeting small molecules (Leelananda & Lindert, 2016; Schreyer & Blundell, 2012). Needless to say, they rely on previous knowledge of compounds that bind to the target of interest.

In fact, as discussed in Chapter 1, most RNA molecules are highly flexible and adopt multiple conformations (Bevilacqua & Blose, 2008; Ganser et al., 2019). This feature might provide transitory structures that might accommodate or be stabilized by small molecules, but also presents difficulties for drug design. Admittedly, taking into consideration the binding of RBPs broadens the targetable chemical space.

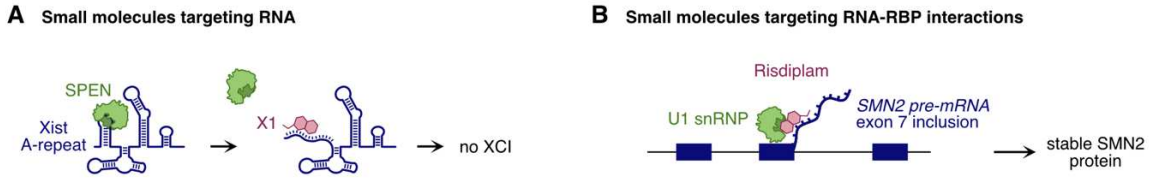


Figure 24. Examples of small molecule approaches.
(**A**) The small molecule X1 is able to disrupt Xist A-repeat, displace SPEN and PRC2 and stop X chromosome inactivation (XCI). (**B**) Risdiplam binds to two distinct sites of the SMN2 pre-mRNA and stabilizes the splicing ribonucleoprotein (RNP) complex, resulting in exon retention (2020, FDA).

4.6. Small molecules target RNA-RBP interactions

The molecular recognition process of RBPs constitutes a substantial source of small molecule target sites. RNA-protein interactions are also very dynamic and structural rearrangements occur in both molecules during binding (Ganser et al., 2019; Hainzl et al., 2005; Leulliot & Varani, 2001). The interaction might perturb RNA stem loops and at the same time expose protein residues (Matthews et al., 2016; Yang et al., 2002). These changes in conformation, both in the RNA and the protein, can create transient pockets that are not discernible when considering only static structures. Thus, targeting the interaction itself, and not its individual components, constitutes an additional approach. The paradigm of small molecules that accommodate within RNA–protein macromolecular complexes are SMN2 splice modulators, like risdiplam (Palacino et al., 2015; Sivaramakrishnan et al., 2017) (Fig.24B). This alternative-splicing modulator selectively binds to a pocket of the quaternary structure formed between SMN2 pre-mRNA and U1 snRNP and stabilizes the interaction.

Despite the success of these proof of concept small molecules targeting RNA and RNA-protein interactions, their discovery does not provide guidance for a rational targeting design. The examples

above were identified without prior knowledge of the nature of the target using phenotypic screens (Howe et al., 2015; Naryshkin et al., 2014; Palacino et al., 2015; Wilson et al., 2008) and only after their functional mechanism was determined.

Novel in vitro systems have recently been described as an emerging alternative for screening RBP-RNA complex formation inhibitors (Ursu et al., 2019; P. Wu, 2020). However, biochemical in vitro methods are limited as they do not recapitulate the cellular environment that affects the binding of RBPs to their RNA targets (Childs-Disney & Disney, 2016). Indeed, the identification of small molecules that target RNA-protein interactions in-cell is currently limited by the absence of unbiased scalable screening methods.

4.7. Objectives

As described above, targeting RNP complexes with small molecules provides a solution to the complexity of drugging RBPs, and constitutes an important alternative to oligonucleotide-based therapies that target RNA in a sequence-specific way. Because of its suitability as a therapeutic target against AML, I aimed to identify small molecules specifically inhibiting YTHDF2 binding to m^6A-methylated RNAs. Since the methods to screen for chemical compounds targeting RNA-protein interactions in-cell are limited, I sought to fill this technological gap by adapting our quantitative incPRINT technique to a drug screening platform. As the inhibition occurs in a cellular system, incPRINT-Drug can overcome the disadvantages of in vitro assays, like the stability of the RNA and the protein or drug crystallization, while providing the cell environment for contextual RNA-protein interactions. In addition, the assay does not rely on phenotypic changes and interrogates a pure molecular interaction, thereby emerging as a promising alternative for the identification of inhibitors of ribonucleoprotein complexes in cellulo.

5. RESULTS

The quantitative nature of incPRINT luciferase assay allows, theoretically, to measure a perturbation of RNA-protein interaction. The disruption of binding either by a mutation in the RNA, in the protein, or due to its inhibition by small molecules would be measured by a decrease in the luciferase activity intensity. However, whether incPRINT has enough resolution to detect the impact of such perturbations had never been tested so far.

To adjust incPRINT to a drug screening method, hereafter referred to as incPRINT-Drug, I sought to find proof of concept RNA-protein interactions that are (i) targetable and (ii) clinically relevant. I started with HIV's *RRE* – Rev binding, a longstanding paradigm of targetable RNA – RBP interaction (Kjems et al., 1991). Later, given the crucial biological role of YTHDF2 in AML development, I assessed the binding between YTHDF2 and m^6A-containing RNAs. Combining several point mutations and disruptions in both the RNA and the protein, I profiled this interaction and selected it as an ideal proof of principle for subsequent experiments. Finally, I followed a standardized drug development pipeline. I first performed a high-throughput screen (HTS) and identified seven hits as candidates for inhibitors of m^6A-modified RNA – YTHDF2 interaction. Later, I validated the hit compounds and tested the specificity of the lead molecules. And as the last step, I tried to some degree a lead optimization phase using fragment-based and an in silico-based approach. This strategy allowed me to demonstrate that YTHDF2 can be targeted by small molecules.

5.1. Optimization of incPRINT-Drug conditions

To perform the initial incPRINT-Drug experiments, I selected the well-studied RRE – Rev interaction as a proof of principle since it has been demonstrated to be targetable using Neomycin B (Zapp et al., 1993). I first cloned the RNA and the peptide to their respective incPRINT plasmids (Fig.25A). However, a 2-hour-long, 200 μM treatment of the cells with this antibiotic led to no detectable changes neither in the interaction intensity nor in Rev expression levels (Fig.25B).

The concentrations and time of incubation I used were chosen based on the literature, despite being applied to other techniques. For this reason, I proceeded with optimization of the duration of the drug treatment step in incPRINT-Drug conditions. Because most of the drugs are dissolved in dimethyl sulfoxide (DMSO), I subjected the transfected cells to a standard 10 μM DMSO treatment for 30 minutes, 2, 16, and 24 hours. Longer treatments resulted in decreased protein expression (Fig.25C), most likely due to cell toxicity. Consequently, I set the maximum drug incubation period at 2 hours.

I then focused on YTHDF2 and the interaction with one of its targets, *tumor necrosis factor receptor superfamily 2* (*Tnfrsf2)* (Paris et al., 2019). Their binding could be efficiently and reproducibly detected by incPRINT, showing a nine-fold difference when compared to GFP (Fig.25D,25E). Aiming to determine the dynamic range of a binding inhibition detectable by incPRINT, I followed a mutagenesis approach. The 3'UTR of *Tnfrsf2* transcript contains up to 32 DRACH motifs, potentially methylated and therefore recognizable by YTHDF2. Using available CLIP data on this protein (Ivanova et al., 2017; Paris et al., 2019), the 15 motifs present within the three peaks found in the 3'UTR of *Tnfrsf2* were selected as effective YTHDF2 binding sites and thereby mutated in our *Tnfrsf2.mut*-MS2 plasmid (Fig.25D). With such mutagenesis-based approach, I was unable to detect any interaction changes between mutated and wild type *Tnfrsf2* by incPRINT (Fig.25E).

Because the m^6A binding site disruption at the RNA level did not prove to be successful, we shifted our attention to a protein-centric mutagenesis approach. It has been demonstrated that W491A point mutation in the three-tryptophan aromatic pocket of YTHDF2, which constitutes the m^6A recognition site, significantly decreases by ten-fold its binding affinity in vitro (Li et al., 2014). Thus, I compared the ability of YTHDF2 wild type ($YTHDF2^{WT}$) or $YTHDF2^{W491A}$ to interact with *Tnfrsf2*, to test if a significant decrease of binding can be detected by incPRINT. The interaction intensity of *Tnfrsf2* and the mutated protein decreased more than two-fold compared to the wild type, while the protein expression levels remained unchanged (Fig.25F). Detected remaining binding of the mutant protein likely results from unspecific binding to 10xMS2 loops. These data demonstrate that *Tnfrsf2* – YTHDF2 constitutes a good proof of concept interaction to assess the application of incPRINT-Drug.

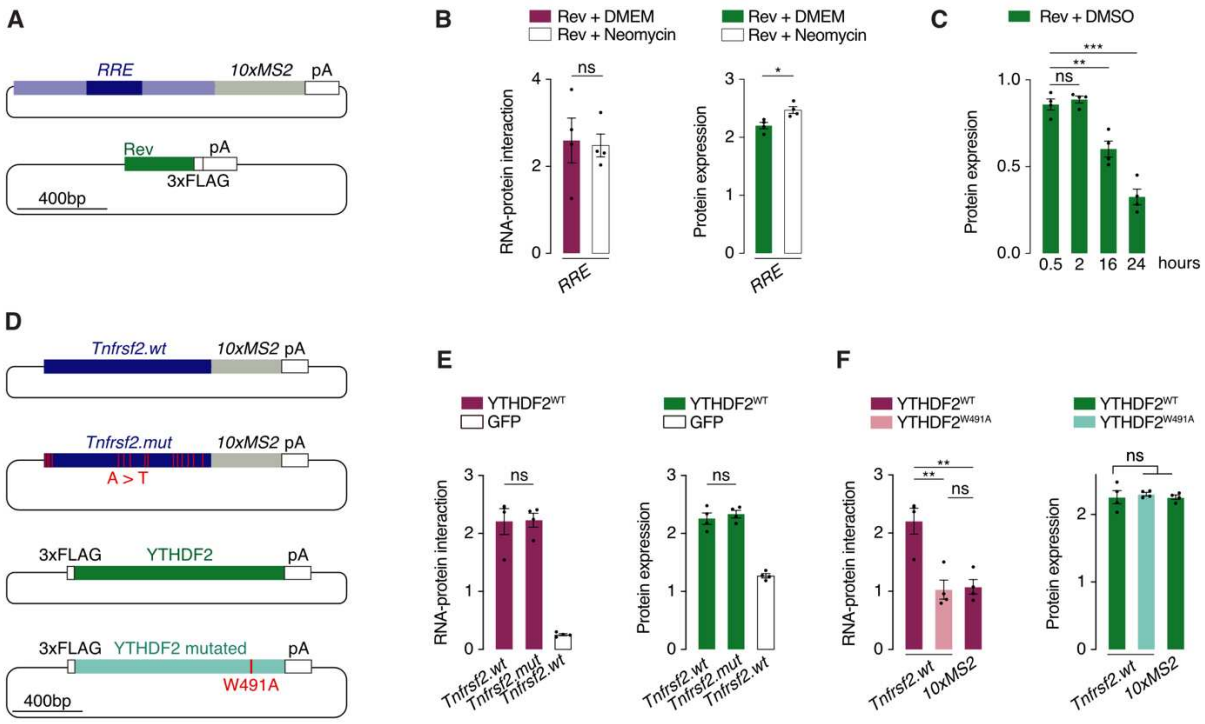


Figure 25. *Tnfrsf2* – YTHDF2 is an optimal interaction for incPRINT-Drug proof of concept.
(**A**) Schematic representation of plasmids containing MS2-tagged *RRE* with 500 bp upstream and downstream adjacent sequence and 3xFLAG-tagged Rev protein. (**B**) Interaction intensity and protein expression levels after 200 μM treatment for 2 h, detected by incPRINT and ELISA respectively. (**C**) Rev protein expression changes upon 10 μM DMSO treatment, for the indicated amount of time detected by ELISA. (**D**) Schematic representation of plasmids containing MS2-tagged *Tnfrsf2* and 3xFLAG-tagged YTHDF2 and their mutated counterparts. m^6A in 15 DRACH motifs were replaced by a T. (**E**) Interaction intensity and protein expression levels, detected by incPRINT and ELISA respectively. Transfected RNA plasmids are indicated on the x axis, and protein plasmids and treatments are indicated in the legend. Data from four biological replicates are presented as mean ± s.e.m. All values are given in Relative Light Units (x10^5). *P* values were calculated by unpaired t-tests. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; ns = not significant.

5.2. HTS of drug inhibitors of *Tnfrsf2* – YTHDF2 interaction

Following the initial tests described above, I established the optimal conditions for incPRINT-Drug screening purposes: I added a 2 h, 10 μM drug treatment step prior to cell lysis (Fig26A). To screen for inhibitors of *Tnfrsf2* – YTHDF2 binding, I used the Prestwick Chemical Library®. This collection of ~1,200 FDA-approved off-patent drugs has high chemical and pharmacological diversity and ensures bioavailability and low cell toxicity. In turn, the reduced dimension of the library allowed me to perform the hole screen in a single week.

Using this approach, I identified seven compounds that showed a decrease in the *Tnfrsf2* – $YTHDF2^{WT}$ interaction comparable to the level of the interaction with $YTHDF2^{W491A}$ (Fig.26B). To compare the data across individual compounds, interaction scores were calculated considering both changes in luminescence –representing interaction intensities– as well as differences in ELISA values –controlling for effects of tested on protein expression levels– (see the Methods section for more details).

Following the high-throughput identification of *Tnfrsf2* – YTHDF2 inhibitors, I validated the identified hits by small-scale incPRINT-Drug (Fig.26C). This experiment confirmed the inhibition specificity and discarded any potential artifacts of data normalization. These data also highlight the specificity of the inhibition since most of the compounds did not inhibit the binding.

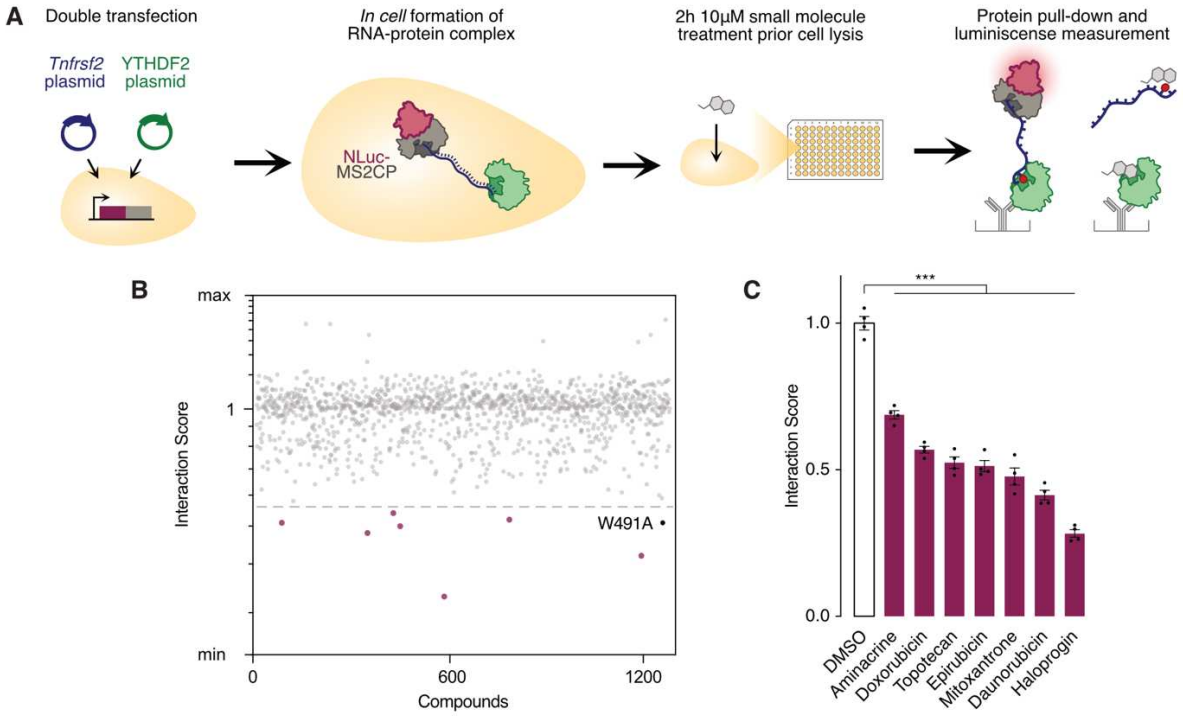


Figure 26. Seven compounds inhibit *Tnfrsf2* – YTHDF2 binding.
(**A**) Workflow of incPRINT-Drug. ELISA step is not shown for visualization purposes. (**B**) Screen of FDA-approved drugs against *Tnfrsf2* – YTHDF2^{WT}. Positive hits are indicated in red and score for *Tnfrsf2*-YTHDF2^{W491A} control is indicated in black. Interaction score averaged from four biological replicates detected by incPRINT-Drug. (**C**) Small scale incPRINT-Drug with the identified hits. Data from four biological replicates are presented as mean ± s.e.m. *P* values were calculated by unpaired t-tests; ***$P < 0.001$.

5.3. Hit-to-lead characterization: Study of the compounds' specificity

To further validate the identified hits and ensure they do not act promiscuously, as well as to understand the nature of their inhibitory effects, I established a series of rigorous controls in each of the incPRINT-Drug steps.

Towards this end, I first performed concentration-dependency assays. A series of different concentrations of each compound were tested, which confirmed a dose-dependency in the effect of five out of seven hits (Fig.27A). Aminacrine and topotecan were discarded as further candidates since their inhibitory effect on the *Tnfrsf2* — YTHDF2 interaction remained the same regardless of the concentration used, suggesting that they are unlikely specific inhibitors of *Tnfrsf2* — YTHDF2^{WT}

interaction. On the other hand, mitoxantrone, doxorubicin, daunorubicin, epirubicin, and haloprogin presented dose-dependent inhibition, confirming their specificity.

Despite being FDA-approved, these small molecules could affect cell integrity. To rule out this potential interference on the assay, I tested the cytotoxicity of the compounds. By observing cell death upon different drug treatments, I identified no changes between the five compounds compared to the DMSO control (Fig.27B). This result corroborated a cell death-independent effect of the decrease of the interaction by the identified small molecules.

Notably, incPRINT-Drug can detect changes in protein expression levels that are posteriorly corrected by the interaction score calculation (see the Methods section for details). However, despite the drug treatment lasting 2h, the compounds could compromise RNA transcription or transcript stability and result in an artifactual hit. A decrease in the amount of the MS2-tagged transcript would impact the luminescence signal and be read out as inhibition of interaction, constituting a false positive. Since the interaction score does not consider changes at the transcript level, I analyzed *Tnfrsf2* levels by RT-qPCR and excluded the effect of the drugs on RNA levels (Fig.27C). Together, these results validated the lead status of five small molecules.

I then took advantage of the versatility of incPRINT-Drug and set RNA and protein specificity controls, providing some insight into the cellular target these compounds engage with. I first assessed RNA specificity by testing other known target and non-target transcripts of YTHDF2 (Paris et al., 2019). As a negative control, I selected the poorly m^{6}A -modified *Ly6A* RNA. Four of the small molecules showed no decrease in *Ly6A* – YTHDF2^{WT} binding (Fig.27D), suggesting an m^{6}A -dependent inhibitory effect. In turn, haloprogin showed a decrease in the interaction intensity, indicating a potential unspecific effect. Next, the highly methylated transcript *Gadd45G* was used as an alternative for *Tnfrsf2* interaction and was expected to behave like such upon drug treatment. The five lead compounds inhibited *Gadd45G* – YTHDF2^{WT} in a similar manner compared to *Tnfrsf2* (Fig.26C,27D).

Later, I tested the drugs using *Tnfrsf2* – YTHDF2^{W491A} as well as the m^{6}A-unrelated *Tnfrsf2* – MOV10 interactions. YTHDF2^{W491A} background interaction intensity was not reactive to the drug treatment of four specific drugs, though haloprogin showed again artifactual decrease in the interaction intensity (Fig.27E). Similarly, MOV10 binding was either mildly or not affected, except for haloprogin. Together, these results indicate that four lead compounds inhibit specifically YTHDF2 – m^{6}A specific inhibition, and that incPRINT-Drug specificity tests can be used to discriminate true hits from false positives.

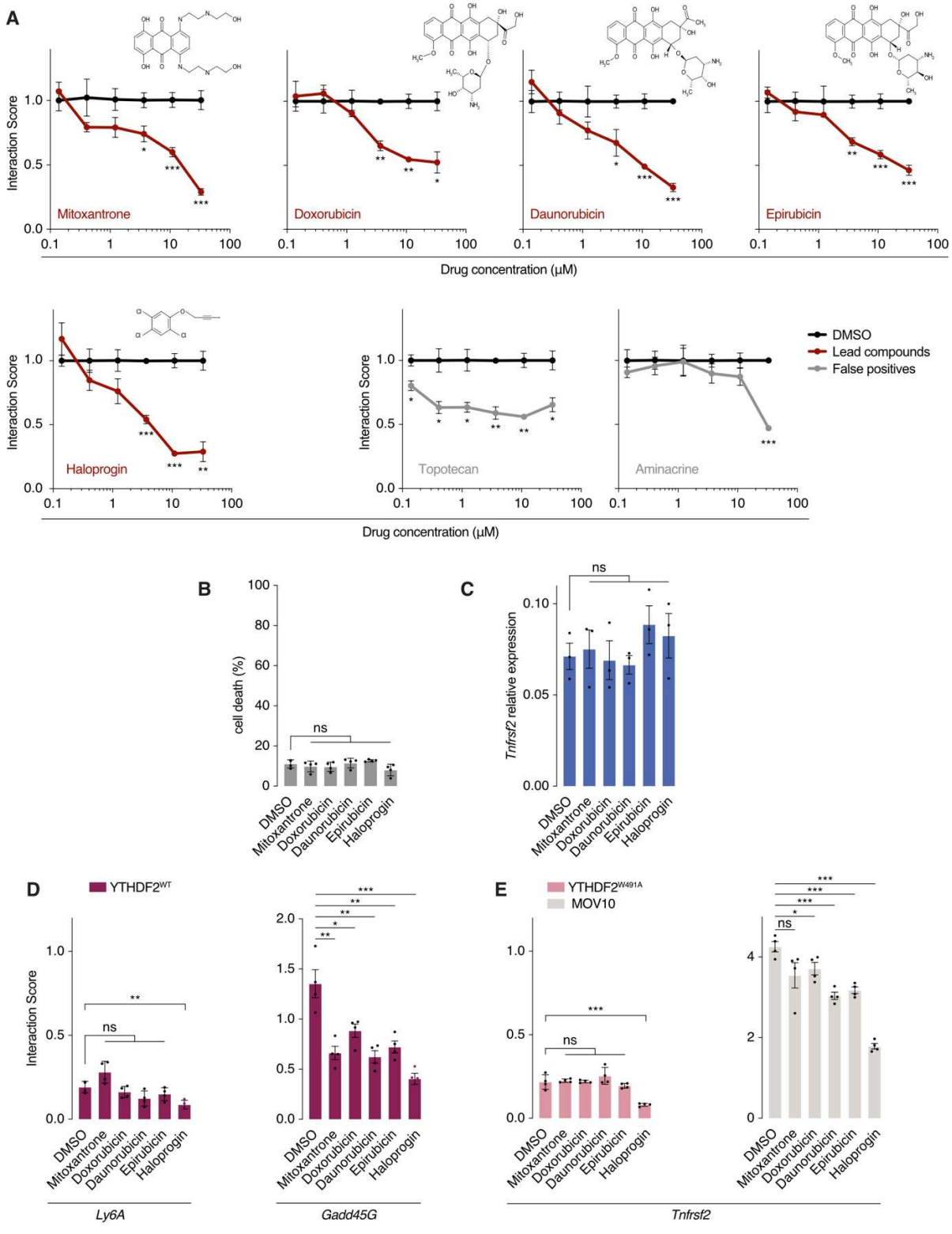


Figure 27. Four small molecules specifically inhibit m^{6}A-modified RNA – YTHDF2 interaction.
(**A**) Dose-dependent inhibition of the interaction upon treatment with different concentrations of the indicated compounds and their molecular structure. (**C**) RT-qPCR comparing levels of expression of *Tnfrsf2* after drug treatment. (**D**) Small scale incPRINT-Drug with selected hits. Compounds are indicated in the x axis, transfected RNA plasmids are indicated below, and protein plasmids are indicated in the legend. Data from at least three biological replicates are presented as mean ± s.e.m. *P* values were calculated by unpaired t-tests. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; ns = not significant.

Importantly, the YTH domain of the three components of the YTHDF family is very similar in amino acid sequence composition (Fig.28A). Structural studies identified the conserved near-identical aromatic cage that recognizes the methyl moiety of m^{6}A modification, which is comprised of three tryptophans (F. Li et al., 2014; Y. Li et al., 2020; Theler et al., 2014). In addition, as discussed above YTHDF paralogs bind m^{6}A transcripts in an equivalent manner. Thus, hypothesizing that the compounds would not only recognize the pocket of all YTH domain proteins, I tested if the interaction of *Tnfrsf2* – YTHDF1^{WT} and *Tnfrsf2* – YTHDF3^{WT} could be targeted by incPRINT-Drug.

I started testing the interaction of wild type and mutated proteins with *Tnfrsf2*, in the same fashion YTHDF2 was first tested. Interaction intensities behaved similarly between the three cases (Fig.28B), suggesting that a drop in the interaction due to drug treatments would indeed be detected. However, a comparison of protein expression levels to YTHDF2 showed a four to seven-fold drop in the expression of the other members of the family in the incPRINT setup (Fig.28B). A difference in the expression of the mutated proteins compared to their respective wild type could also be observed.

Despite these conditions being sub-optimal, I tested if the four compounds could inhibit the interaction with other YTH domains. Treatment of *Tnfrsf2* – YTHDF1^{WT} with three out of the four drugs recapitulated the near-two-fold interaction depletion (Fig.28C). For *Tnfrsf2* – YTHDF3^{WT}, the low protein expression encountered incPRINT-Drug's resolution, hindering the detection of the interaction in the case of mitoxantrone. These results indicate, to a certain extent, that the four lead compounds inhibit the YTH domain interaction with m^{6}A-modified transcripts.

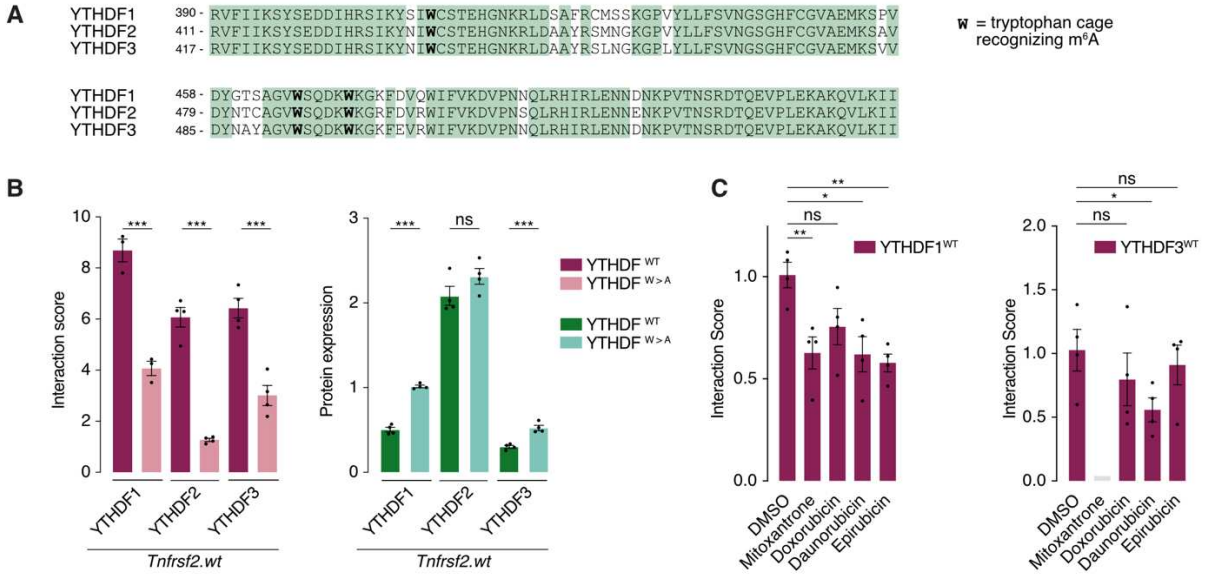


Figure 28. YTHDF1 and 3 protein expression is suboptimal for incPRINT-Drug testing.
(**A**) Comparison of YTH domain sequences of the three YTHDF proteins. In green, conserved amino acid residues. (**B**) Interaction intensity and protein expression levels, detected by incPRINT and ELISA respectively. Transfected protein plasmids are indicated in the x axis, RNA plasmids are indicated below, status of each YTHDF sequence is indicated in the axis. Values are given in Relative Light Units (x10^5). (**C**) Small scale incPRINT-Drug with selected hits. Compounds are indicated in the x axis and protein plasmids are indicated in the legend. All data from four biological replicates are presented as mean ± s.e.m. *P* values were calculated by unpaired t-tests. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ns = not significant.

5.4. Lead optimization: compound modification and fragment screening

Doxorubicin, daunorubicin, and epirubicin belong to the rubicin family of natural compounds, also called anthracyclines, and mitoxantrone is an artificially synthesized drug that resembles rubicins (Fig.27A). To understand if the core structure of the rubicins is responsible for the m^6A – YTHD inhibition described above, I performed an indirect structure-function analysis. By testing additional members of this family of compounds, I aimed to determine if the inhibitory effect was universal to all members, as well as if chemical modifications of the rubicins confer the small molecules a higher inhibitory effect.

I carried out incPRINT-Drug with three extra rubicins congeners. I identified idarubicin's inhibition as comparable to the lead compounds (Fig.29). Although a decrease in the interaction could be seen for pirarubicin treatment, the effect was rather mild. This result excludes the core rubicins structural subgroup as sufficient for *Tnfrsf2* — $YTHDF2^{WT}$ inhibition and confirms that incPRINT-Drug is sensitive to small compound modifications.

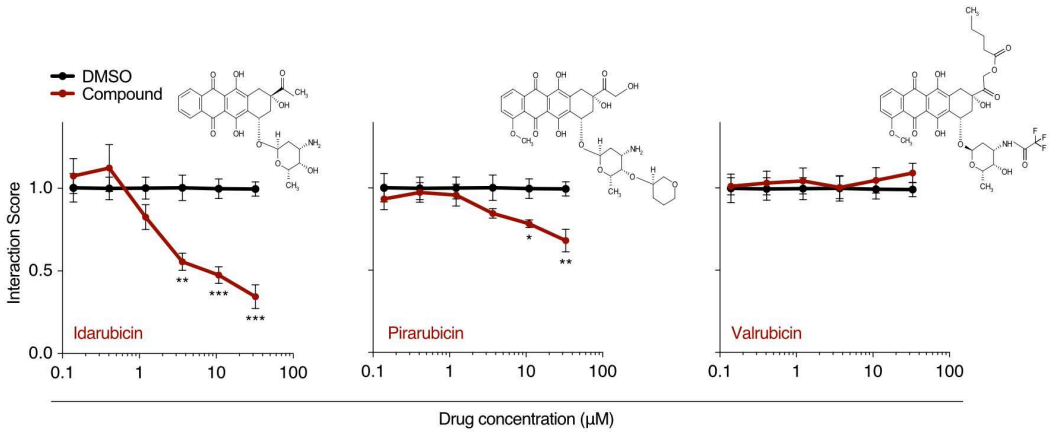


Figure 29. incPRINT-Drug is sensitive to modifications of the same drug family.
Dose-dependent inhibition of the interaction of additional rubicins not included in the drug library and their molecular structure. Data from four biological replicates are presented as mean ± s.e.m. *P* values were calculated by unpaired t-tests. **P* < 0.05; ***P* < 0.01; ****P* < 0.001.

Remarkably, rubicins are used currently as the standard of care for leukemia (Döhner et al., 2017), that has not changed substantially in recent years, although their mechanism has never been related to YTHDF2. Despite controversial, the proposed main target of rubicins is topoisomerase 2 poisoning (Marinello et al., 2018; Nitiss, 2009). With the aim to uncouple the alternative YTHDF2-dependant mechanism we propose from the above-described rubicin target, I followed a fragment-based drug discovery approach. I hypothesized that simpler molecules or FRAGMENTS , defined as drugs that contain less than 20 non-hydrogen atoms, would maintain the inhibition of *Tnfrsf2*-YTHDF2 interaction or even better orientate within the binding site, as well as avoid off-target interactions.

To identify the essential YTHDF2 inhibitory-acting part of the rubicins, I tested ~500 chemotypes, or chemical structure motifs. This Core-Set Fragment Library® includes diverse primary substructure compounds derived from small molecules present Prestwick Chemical Library®. These fragments are smaller and simpler than the ones I previously tested. This *chemotheque* included one fragment derived from our four lead molecules (nº 1884). In addition, I tested five customized rubicin-derivative fragments. None of the fragments from the library or the additional fragments showed an inhibitory behavior comparable to the *Tnfrsf2* – YTHDF2^{W491A} (Fig.30).

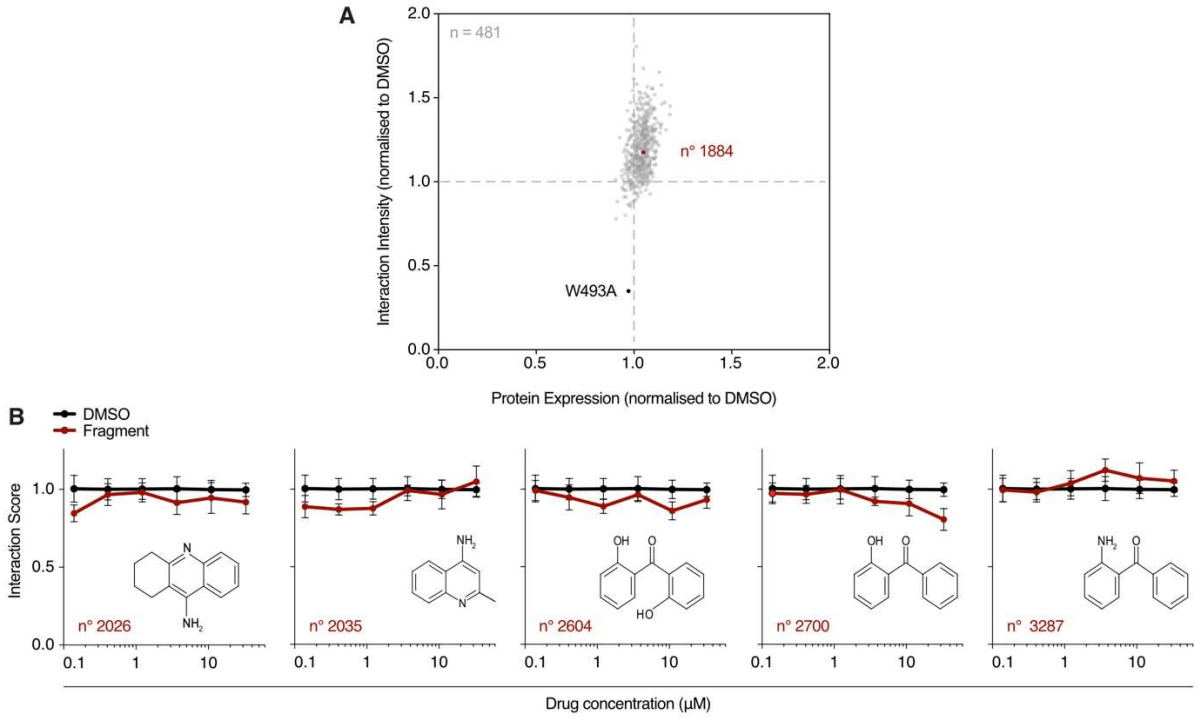


Figure 30. No rubicin-derived fragments inhibit *Tnfrsf2* – YTHDF2 interaction.
(**A**). Scatter plot of Fragment library screened against *Tnfrsf2* – YTHDF2^{WT}. Lead-derived fragment is indicated in red and *Tnfrsf2* – YTHDF2^{W491A} control is indicated in black. Data averaged from four biological replicates detected by incPRINT-Drug and normalized to *Tnfrsf2* – YTHDF2^{WT} treated with DMSO, represented with a dotted line. (**B**) Dose-dependent inhibition of the interaction of additional fragmented compounds not included in the fragment library and their molecular structure. Data from four biological replicates are presented as mean ± s.e.m.

I then adopted a complementary ligand-based drug discovery approach by performing a computational substructure searching. The active rubicin compounds contain a maximal common subgraph (MCS) that can be truncated. To identify possible binding portions of the rubicin and to see if trimming is possible, Steven Shave and Manfred Auer (University of Edinburgh) carried out substructure searching (Schreyer & Blundell, 2012) on the conserved diol ring system and the tetrahydropyran portion. With this in silico approach, 19 compounds were identified to contain a rubicin MCS. Four of them inhibited YTHDF2 – m^{6}A binding in incPRINT-Drug, and only mangiferin and licogliflozin did it significantly (Fig.31). Of note, both inhibitions start to be significant at 33µM, compared to the 4 µM concentration found on the lead compounds (Fig.27A).

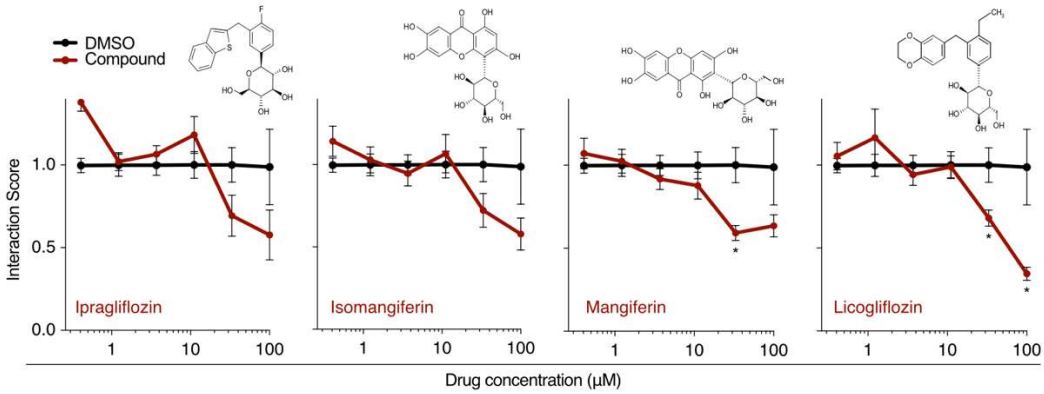


Figure 31. Rubicin-MCS-containing drugs fail to inhibit Tnfrsf2 – YTHDF2 interaction.
Dose-dependent inhibition of the interaction of a compound from additional rubicins not included in the drug library and their molecular structure. Data from four biological replicates is presented as mean ± s.e.m. *P* values were calculated by unpaired t-tests. $*P < 0.05$.

6. DISCUSSION

Since the discovery of the global transcriptomic m^{6}A modification (Dominissini et al., 2012; X. Wang et al., 2014), major efforts have been put into understanding its precise mechanism. It is now well established that a refined ensemble of RBPs regulates m^{6}A deposition, removal, and reading. The m^{6}A interactome has been unexploited as a therapeutic target, mainly due to a lack of technology for drug identification but also because it appeared as clinically relevant not long ago (Cully, 2019). From then on, several biotech companies and academic laboratories are racing to find small molecules to drug RNA modifying proteins (Berdasco & Esteller, 2021; Quattrone, 2020).

Recently, the *Tnfrsf2* – YTHDF2 interaction has been identified as a key therapeutic target for AML treatment (Paris et al., 2019). To further identify compounds inhibiting this binding, several approaches could be followed. Because the crystal structure of its YTH domain is already available (T. Zhu et al., 2014), virtual screening of molecular libraries has already been performed, attempting to computationally target YTHDF2 (Bedi et al., 2020). Phenotypic cell-based screens are powerful drug discovery tools that have brought small molecules targeting RNA-RBP interactions to the clinic (Moffat et al., 2017). However, because of their target-agnostic nature, they cannot be used for the discovery of inhibitors of a precise molecular interaction.

Recently-developed in vitro biochemical techniques, like Förster resonance energy transfer and other fluorescent-based assays, are expanding the capacity of screening for RNA–RBP inhibitors (Ursu et al., 2019; P. Wu, 2020). Likewise, biophysical or structure-based methods like Surface Plasmon Resonance (SPR) or Nuclear Magnetic Resonance (NMR), classically used as validation methods, are being extensively utilized and refined for HTS (Erlanson et al., 2016). Yet, each method has distinct caveats which need to be kept in mind and only recently are being tailored for RNA-protein interaction drugging (D'Agostino et al., 2019). My work presents the customization of an in-cell screening methodology for target-based drug discovery applied to targeting methylated RNA – YTHDF2 interactions, which materialized into the identification of small molecule inhibitors for this RNA-protein interaction.

6.1. Interaction quantification by incPRINT

First and foremost, to quantify the m^{6}A-dependent binding of YTHDF2 to highly methylated RNAs, I conducted a series of small-scale incPRINT assays (Fig.25/Section5.1). A robust and reproducible interaction between *Tnfrsf2* and YTHDF2 was detected, which was nine-fold over the background.

Unexpectedly, the first mutagenesis approach I followed, deleting methylated sites of *Tnfrsf2*, did not result in a decrease in the interaction measurable by incPRINT. There are several explanations for the lack of a decrease in the luminescence signal. In the first place, A>T mutations were introduced solely on the 15 sites identified by m^{6}A immunoprecipitation to be preferentially methylated in pre-leukemic mouse cells (Ivanova et al., 2017; Paris et al., 2019). The 3'UTR of *Tnfrsf2*, as well as GFP cDNA, contains other `DRACH` motifs that can potentially be methylated (Fig.32). In the second place, this chimeric transcript is ectopically expressed and does not reproduce the gene architecture of *Tnfrsrf2*, a feature known to drive methylation patterns (Dominissini et al., 2012; Uzonyi et al., 2022). In the third and last place, although there is general evolutionary conservation of m^{6}A site overlap between mouse and human (Batista et al., 2014; J. Liu et al., 2020; Meyer et al., 2012), the methylation machinery of HEK293T cells used for incPRINT-Drug might have specificities that do not equate those of pre-leukemic cells in mice. My results suggest that a partial loss of the methylation sites of an RNA of interest is not tangible in our system. The depletion of METTL3 methylation enzyme in HEK293T incPRINT cells could circumvent this problem and unequivocally determine the m^{6}A -dependent binding of YTHDF2 and other interactors.

Figure 32. *Tnfrsf2* mutagenesis design.
Mouse *Tnfrsf2* 3'UTR was cloned to pCDNA3.1 MS2-containing vector together with eGFP cDNA to ensure proper export to cytoplasm. 15 DRACH motifs were mutated (light red) based on m^{6}A immunoprecipitation read coverage. The coordinates (mm10 – Tnfrsf1b) of the three picks identified were the following: chr4:145,214,419-145,214,720 (highest peak), chr4:145,214,823-145,215,179 (lowest peak), chr4:145,215,631-145,215,852 (second highest peak). Yet, this chimeric transcript contains 29 non-mutated DRACH motifs (dark red).

On the contrary, by mutating one tryptophan of the m^{6}A binding site of the YTH domain, I was able to effectively detect a decrease by two-fold in the interaction intensity between the wild type and the mutated protein. Beyond establishing a dynamic range of the m^{6}A-dependent interaction for a subsequent small molecule screen, these results also indicate that incPRINT and protein mutagenesis can be combined for the study of the RNA-binding domains of RBPs in cells. In sum, I determined the precise sensitivity of incPRINT in order to quantify RNA-RBP interaction inhibitions.

6.2. Establishment of drug treatment and HTS conditions

The optimization of an appropriate drug treatment time (2 h) was then carried out by small-scale incPRINT-Drug assays. Together with the appropriate drug concentration (10 μM), the suitable conditions to target RNA-RBP interactions without perturbations of cell viability where established for subsequent incPRINT-Drug scalability in a high-throughput format.

The selection of a suitable screening collection of compounds represents one of the major challenges of the hit identification and lead discovery phase in drug discovery studies (Hughes et al., 2011; Macarron et al., 2011). In order to identify RBP-RNA interaction inhibitors, some of our competitors performed an in vitro screen using a library that contained 250,000 small molecules (Yankova et al., 2021), while others started their HTS using a virtual approach to narrow down from 260,000 compounds to few hundreds (Su et al., 2020), which were tested afterward in cells. For our in-cell proof of concept, we prioritized a more compact, manageable library that could ensure low cell toxicity while maintaining a heterogeneous composition. Eventually, we chose an FDA-approved chemical collection of ~1,200 off-patent compounds. Regulatory approval appeared as an advantage more than a limitation: approved molecules sometimes have unknown or multiple targets, and they offer the potentiality of being rapidly repurposed (Ashburn & Thor, 2004; Pushpakom et al., 2018).

Next, I carried out the small molecule screen. The identification of only seven hits, that is, ~0,6% of the compounds screened, substantiated the choice of time and drug concentration (Fig.26/Section5.2). Additionally, the low micromolar range (10 μM) of activity of the identified compounds was in line with RBP inhibitors found in other studies (Lim et al., 2016; Meisner et al., 2007; Minuesa et al., 2019). Of note, some compounds showed an increased luminescence. It would be of interest to examine whether a real boosting of the RNA-protein interactions in fact occurs.

A possible criticism of our approach may be that the IC_{50} of these compounds is not provided. This quantitative measure indicates the concentration of a molecule needed to inhibit a biological process by 50% and is used to compare drug potency (Y.-C. Cheng & Prusoff, 1973; Swinney, 2011). The IC_{50} metric does not apply to incPRINT, as this technique relies on living cells. At high concentrations, where the saturation or maximal effect is detected, the compounds are cytotoxic. To analyze compound affinity in a semi-quantitative fashion, I titrated down the concentration of the drug and determined the effect on the interaction intensity (Fig.27/Section5.3). This experiment allowed to establish the specificity for five of the seven hits, whose inhibition effect did not remain constant upon changes in the concentration.

Altogether, the high-throughput identification of *Tnfrsf2* – YTHDF2 inhibitory compounds demonstrates the fast and scalable potential of incPRINT-Drug.

6.3. Set up of small molecule specificity controls

To ensure that the decrease in the interaction intensity is not simply a product of the incPRINT system, I integrated a hit assessment process (Fig.27/Section5.3). To exclude the possibility of the compounds impacting cell integrity, I tested whether cell toxicity was impacted upon treatment with the small molecules. All hits exhibited equivalent cell death levels. Later, I measured the *Tnfrsf2* relative RNA levels and ensured no variation in the expression would occur between the different compounds and the DMSO vehicle. To gain insight into the specificity of the hits for the m^{6}A – YTHDF2 interaction, I tested them in several combinations of RNA-protein bindings. It should be underlined that the m^{6}A mark enhances the affinity of YTHDF2 to RNA, but this binding can still occur in unmethylated transcripts. Consistent with the fact that our small molecules inhibit the binding of YTHDF2 and methylated RNAs, YTHDF2^{WT} – *Ly6A*, a lowly methylated transcript, and YTHDF2^{W491A} – *Tnfrsf2*, highly m^{6}A-modified, displayed comparably low interaction intensities. Conversely, the drugs inhibited *Gadd45G*, another strongly methylated RNA, in a similar fashion compared to *Tnfrsf2*. Only Haloprogin behaved promiscuously and was discarded as a specific inhibitor.

Worthy of note is the flexibility of this unbiased triage of hits. Additional RNA-RBP interactions can be included to ensure the specificity of the lead compounds. In our case, MOV10 was tested, but complementary m^{6}A unrelated interactions like HNRNPC or PABPC3 could be evaluated as well. Because the YTH domain is essentially identical between the family of proteins, I hypothesized that the identified small molecules would also inhibit YTHDF1 and YTHDF3 binding to methylated RNAs (Fig.28/Section5.3). Despite the decrease in the interaction of YTHDF1 – *Tnfrsf2* being similar to that of YTHDF2, the results highlight a limitation of incPRINT-Drug when FLAG-tagged RBPs are expressed at low levels. Overall, these findings show the successful establishment of specificity controls to discriminate hits from false positives

6.4. Nature of lead compounds and discovery of novel leads

But what is the exact identity of these four hits? As mentioned above, they belong to the same family of compounds, called rubicins or anthracyclines. This class of natural small molecules has been known for decades to act as topoisomerase inhibitors (Arcamone et al., 1961; Marinello et al., 2018; Nitiss, 2009). However, they have also been suggested to harbor alternative non-topoisomerase activity (Gewirtz, 1999; Swift et al., 2006) as well as RNA binding properties (Marcheschi et al., 2009; Velagapudi et al., 2018; S. Zheng et al., 2009). Of note, another compound from the rubicin family, bisantrene, was found to inhibit FTO RNA demethylase (Su et al., 2020). Because many other rubicins have been synthetized since their discovery, I examined the sensitivity of incPRINT-Drug to minor

chemical modifications of the compounds (Fig.29/Section5.4). Three additional rubicins were tested and showed different degrees of inhibitory activity.

Strikingly, rubicins are the current standard of care for AML patients (Döhner et al., 2017). One might wonder, then, what is the relevance of having found drugs to target YTHDF2 – m^6A if they are already being used to treat AML. Individuals with this type of leukemia respond better to such treatment than other cancer patients. Although purely conjectural, it is exciting to consider these compounds might have been naturally selected to treat AML due to previously unidentified targets. Describing alternative targets of approved drugs might not only be crucial for the understanding of the primary and secondary effects these drugs have in AML patients but may also impact research. Some studies in non-coding RNA biology use rubicins to induce DNA damage (Elguindy & Mendell, 2021; Ziv et al., 2021), and this unknown target could have interfered with their results. In short, this novel target of rubicins compounds has important consideration for researchers and clinicians.

Nevertheless, rubicins are far from being an ideal treatment. Patients are at high risk of cardiac damage and some tumors develop resistance (Martins-Teixeira & Carvalho, 2020). In order to discover alternative drugs to adequately address the pressing medical needs of AML patients, we resumed our screening phase in order to identify undescribed compounds inhibiting YTHDF2-*Tnfrsf2* interaction.

First, I tested fragments from approved drugs previously tested (Fig.30/Section5.4). No decrease in the interaction intensity was observed for any of the fragmented compounds. It is known that smaller compounds can form fewer interactions with their targets, leading to lower affinities. Hence, the decision to perform the fragment-based screen at 10 μM might appear too conservative. Subsequent tests with individual fragments with concentrations up to 33 μM did not detect any inhibitory effects of the fragment molecules. To dissipate any doubt, and despite risking obtaining false positives, the library could be used again to screen for inhibitors, augmenting the concentration to 100 μM.

Later, I followed a ligand-based drug discovery approach, using the composition of our four hits as a baseline for the identification of a common subgroup or chemotype that could be present in other compounds. Some of these shared building blocks were pinpointed in 19 additional small molecules, that were subsequently tested on incPRINT-Drug (Fig.31/Section5.4). Together, these results demonstrate incPRINT-Drug can be combined with in-silico approaches for novel compound discovery.

6.5. What is the m^6A – YTHDF2 inhibitory mechanism?

Based on the results obtained, it is hard to infer the mechanism of inhibition of the hits. On incPRINT-Drug conditions, a compound could be deemed a hit on three occasions: (i) if it binds to the intended RNA target, competing with the RBP (Zapp et al., 1993), or disrupting its structure (Aguilar et al., 2022); (ii) if it interacts with the protein RNA-binding site, competing with the RNA molecule (Minuesa

et al., 2019) or an allosteric pocket that blocks a conformational change required for molecular recognition of RNA (Clingman et al., 2014; S. Zhang et al., 2018); and (iii) if it recognizes the RNA-protein interacting interface (Sivaramakrishnan et al., 2017).

To confirm the binding of the rubicins to the YTH domain, we aimed at recapitulating the inhibition using orthogonal methods. In turn, this validation would allow mapping the interaction of the compounds, gleaning information on their mechanism of inhibition. Because YTHDF2 had been previously crystalized (F. Li et al., 2014), we sought to confirm the interaction using X-ray crystallography of the protein soaked in different concentrations of our compounds of interest. However, because no YTHD crystals were achieved with published protocols (T. Zhu et al., 2014) (in collaboration with Dónal O'Carroll and Atlanta Cook laboratories, University of Edinburgh, UK), we decided to pursue an alternative strategy. Another partner laboratory (Schofield laboratory, Oxford, UK) is currently performing a set of secondary validation experiments of the identified small molecules employing Mass Spectrometry-, SPR- and NMR-based methods. If this turns out to be more challenging than anticipated, we will proceed with RIP experiments prior to incubation of our incPRINT transfected cells with the rubicins (Minuesa et al., 2019). These experiments will constitute the final confirmation of the inhibition of the interaction between YTHDF2 and m^6A-modified RNAs.

6.6. Relevance and perspectives

The results above demonstrate the fast and robust capability of incPRINT technique to quantify in-cell RNA-RBP interactions. The luciferase detector system is shown here as a sensitive approach for the detection of changes in RNA-protein binding. Our findings also evidence the first-in-class potentiality of incPRINT-Drug as a rapid in cellulo method for the discovery of small molecules inhibiting RNA-protein interactions.

The same pipeline could be used to identify compounds stimulating the RNA-binding of RBPs, like the anti-tumorigenic interaction between *CASC15* and the proteins DDX50 and DUSP12 (Section3.8). As a matter of fact, these applications are just the tip of the iceberg of a whole wave of alternative implementations incPRINT-Drug might serve as a blueprint for. It is attractive to speculate that this method could be used to screen for nucleotide-based agents targeting specific disease-causing RNA-protein interactions. Theoretically, incPRINT-Drug could also be employed for the study and target a broader spectrum of interactions, like RNA sensing proteins localized in the membrane, also responsible for many pathological processes (G. J. Brown et al., 2022; P. Wu, 2020; S. Zhang et al., 2018).

As regards the work presented here, once scalability and cost-effectivity of incPRINT are resolved, the identification of novel compounds targeting *Tnfrsf2* – YTHDF2 interaction and subsequent pharmacological inhibition assessment in vivo will certainly become a reality.

METHODS

Cell culture. HEK293T cells expressing NanoLuc-MS2CP fusion protein (Graindorge et al., 2019) were maintained in DMEM (Gibco, #41965062) and 501mel cells (Lionel Larue laboratory, Institut Curie) were cultured in RPMI 1640 (Gibco, #21875091). Both medias were supplemented with 10% fetal bovine serum (Gibco, #10270106) and 1% penicillin/streptomycin (Gibco, #15140). Cells were grown at 37 °C in a humidified incubator with 5% CO2 and tested regularly for mycoplasma contamination by PCR. Transfected plasmids or endoribonuclease prepared siRNAs (esiRNA) were resuspended in Opti-MEM I reduced serum medium (Gibco, #31985).

Genome editing of melanoma cells. All 501mel cell line alleles were generated using CRISPR/Cas9 gene editing technology. For h.CASC15 knockout generation, two single guide RNAs were cloned into pSpCas9 (BB)-2A-eGFP vector (Raphaël Margueron laboratory, Institut Curie, Addgene plasmid #48138). Cells were lipofected with 500 ng of each sgRNA with lipofectamine 2,000 (Invitrogen, #11668). 48 hr after transfection the GFP high population of cells was collected by FACs and individual cells were seeded to each well of a 96 well plate. Sorted cells genotyped as wild type were used as a negative control.

To generate HA-tagged DUSP12 or DDX50 cell lines, an HA-PreScission-6xHis tag (Pérez-Rico et al., 2020) was inserted in-frame after the start codon for C-terminal tagging and before the stop codon for N-terminal tagging. 501mel cells lipofected with the DUSP12 or DDX50 HA template vectors, containing the tag flanked by two 400nt homology arms, cloned into pUC57 (synthesized by GenScript), together with the pSpCas9(BB)-2A-Puro vector (Addgene plasmid #62988) containing a gRNA sequence targeting each of the proteins, promoting homologous recombination. Transfected clones were selected with 1 μg/mL of puromycin for 48 h.

In both cases, individual clones were expanded and genotyped for the integration of the tag at the correct genomic location by PCR and DNA sequencing to ensure no deletions or mutations in the protein-coding sequence, or for the deletion. Genomic DNA was extracted by lysing the cells with lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 10 mM NaCl, 0.5% SDS) supplemented with 0.1mg/ml proteinase K (Roche, #03115828001) at 55 °C for 3 h, followed by ethanol precipitation and resuspension in nuclease-free water. All relevant oligonucleotides and tag sequences are listed in the Supplementary Table.

CRISPR guide RNA cloning. Guide RNAs were designed using CRISPick portal from Broad Institute or using the IDT gRNA design tool (Schubert et al., 2021) and cloning to the respective plasmids was done following Zhang Lab General Protocol (Zhang Lab, 2013).

Cellular assays. For the proliferation assay, 30,000 cells were plated in a 35 mm dish at day 0 and counted (Vi-cell XR, Beckman Coulter) each day until day 5. For the clonogenic assay, 35 mm dishes were seeded with 500 cells. After 10 days of growth, colonies were fixed with 4% PFA, stained with

Crystal violet in 10% ethanol, and counted. For invasion assay, transwell Boyden chambers with 8 µm pore size (Falcon, #353097) were coated with GFR Matrigel matrix (Corning, #354230). A cell suspension of 200,000 cells in serum-free media was added to the Matrigel-coated upper chamber and FBS-supplemented media was added to the lower chamber as chemoattractant. Cells were allowed to invade for 24 h, then fixed with 4% PFA, and stained with Crystal violet in 80% methanol, 10% formaldehyde, and counted. For the wound healing assay, 30,000 cells were seeded on each side of silicone inserts (Ibidi, #80209). After 24 h, inserts were removed, and migration was recorded by imaging cells every 1 hour for 24 hours using an IncuCyte Live Cell Imaging System (Essen, Bioscience) with a 4× magnification. Wound closure was evaluated by measuring the size of the wound with ImageJ software using an edge detection macro (Gallagher et al., 2019). All experiments were performed in 3 biological replicates of each mutant and wild type clones in technical triplicates.

Generation of zebrafish melanoma model. pDEST-Tol2-mitfa:humanNRAS^{G12D} plasmid (Adam Hurlstone laboratory, Manchester University) was modified by inserting in ClaI site the mitfa:*h.CASC15* amplified from FirstChoice Human brain Total RNA (Ambion, #6050) and mitfa:*h.CASC15-short* amplified from 501mel cDNA, or their antisense counterparts with Gibson assembly (New England Biolab, #E2621). An artificial intronic sequence was inserted at cDNA exon/exon junction to insure the correct subcellular localization of the transcript. *h.CASC15Δexon12* was cloned by Mattieu Petitjean. To produce Tol2 transposase mRNA, pCS-TP plasmid (Adam Hurlstone laboratory, Manchester University, UK) was digested by NotI and RNA synthesized using HiScribe SP6 RNA Synthesis Kit (NEB, #E2070) and RNA was purified with the RNeasy Mini Kit (QIAGEN, #74104).

Zebrafish embryos were co-injected at one-cell stage with 50 ng/µL Tol2 mRNA, 50 ng/µL of respective melanoma induction plasmid, and PhenolRed (0.01%). A minimum of 2 independents injections were performed to generate a population of fish presenting melanoma in each condition.

All zebrafish were bred and maintained at the Institut Curie, Paris in accordance with the current European Directive 2010/63. Experimental procedures were specifically approved by the ethics committee of the Institut Curie (APAFIS#28477-2020120112502053-v2) in compliance with the international guidelines.

Melanoma monitoring procedure. Injected embryos were selected for *mVenus* expression in the eye lens at 72 hpf. Zebrafish larvae were bred in standard conditions (28-29 °C under a 14 h light, 10 h dark cycle in a filtered freshwater recirculation system). Starting from 5 to 21 weeks post-fertilization, melanoma-induced zebrafish were individually observed each week and classified according to their melanoma progression stage: No Lesion; Stage I = pigmentation defects; Stage II = localized nevi; Stage III = extended nevi (radial growth progression); Stage IV = tumor (vertical growth progression).

All animals presenting tumor, nodule, or hyperplasia covering more than 10% of body surface, epidermic lesion, weight loss, and/or feeding/swimming troubles were sacrificed immediately.

esiRNAs. Gene knockdown was achieved using esiRNAs (Kittler et al., 2007) (Eupheria Biotech) directed against indicated proteins, using RLUC as a negative control. Cells in each of the two sides of the IBIDI inserts were lipofected with 0.05 µg of esiRNAs and 0.5 µl of lipofectamine 2,000 24 h before the start of the migration assay.

RT-qPCR. Total RNA from zebrafish tissue, esiRNA-treated or plasmid-transfected 501mel human cells or RIP samples was isolated using TRIzol (Invitrogen, #15596) according to manufacturer's instructions. Extraction was followed by DNAse treatment (TURBO DNA-free, Invitrogen #AM2238) and ethanol precipitation. 900 ng of total RNA or 200 ng in case of RIP samples were reverse-transcribed using SuperScript IV kit with oligo-dTs (Invitrogen, #18091), followed by qPCR using SYBR Green (Applied biosystems, #4767659) according to manufacturer's instructions. RNA expression levels were normalized to *GAPDH, β-actin* (human cell line studies) or *eif1α* mRNA (in vivo studies) using the comparative $2^{-\Delta\Delta CT}$ method. The number of biological replicates is stated in the figure legends, each biological replicate was run with four technical replicates. Primers are reported in the Supplementary Table.

Generation of incPRINT expression constructs. Customized pCDNA3.1 plasmid containing MS2 loops (Graindorge et al., 2019) was modified by inserting in BstBI site *h.CASC15* amplified from FirstChoice Human brain Total RNA (Ambion, #6050), *zf.casc15* amplified from zebrafish embryo cDNA and *h.CASC15-short* amplified from 501mel cDNA, and their truncated and antisense counterparts with Gibson assembly (New England Biolab, #E2621). RRE was amplified from pTRIP vector (Nicolas Manel, Institut Curie). Rev protein was cloned by Antoine Graindorge and YTHDF wild type and mutated proteins, *Tnfrsf2* wild type and mutated, *Ly6A* and *Gadd45G* were cloned by Azzurra Laura De Pace. YTHDF proteins were cloned to the N-term 3xFLAG containing plasmid. All relevant oligonucleotides are listed in the Supplementary Table.

incPRINT. incPRINT was carried out as described previously (Graindorge et al., 2019): 384-well plates (Greiner Bio-One, #781074) were coated overnight with the anti-FLAG M2 antibody (Sigma, #F1804; 10 µg/ml in 1× PBS) and blocked for two hours in 1% BSA, 5% sucrose, 0.5% Tween-20 in 1×PBS. Plasmids encoding RNA-MS2 and 3×FLAG-tagged test proteins were co-transfected using Polyethylenimine (PEI) (MW 40,000, Polysciences, #24765) into a HEK 293T stable cell line expressing the NanoLuc luciferase fused to MS2CP. The day before transfection, cells were seeded in 96-well plates (30,000 cells per well). Co-transfections were performed using 300ng and 600ng of the protein and RNA plasmid respectively. Two days after transfection, cells were washed twice with ice-cold 1x PBS and lysed in ice-cold RQ1-HENG buffer (20 mM HEPES-KOH pH 7.9, 150 mM NaCl, 10 mM MgCl2, 1 mM CaCl2, 5% glycerol, 1% Triton X-100), supplemented with protease inhibitors

(aprotinin, leupeptin and pepstatin, 1 µg/ml each, PMSF 0.5 mM), 40 U/ml of RNaseOUT Ribonuclease Inhibitor (Invitrogen, #10777-019) and 30 U/ml of RQ1 DNase (Promega, #M6101). After lysis (10 min, 4 °C) and DNase incubation (30 min, 37 °C), the lysates were transferred to 384-well plates previously coated. Following a three-hour incubation at 4 °C, plates were washed seven times with RQ1-HENG buffer. A 1:200 dilution of furimazine substrate (Promega, #N1110) was added to the plates and luminescence in each well was measured with a plate reader (TECAN Spark). Following luminescence measurement, HRP-conjugated anti-FLAG antibody (Abcam, #ab1238, 1:10,000 dilution) in ELISA buffer (1× PBS, 1% goat serum, 1% Tween-20) was added to each well. After 30 min of incubation at room temperature, plates were washed with 1× PBS, 0.05% Tween-20, SuperSignal™ ELISA Pico Chemiluminescent Substrate (Thermo Fisher Scientific, #37069) were added and ELISA signals were detected with the microplate reader. All 96- and 384-well plate washes were performed with a microplate washer (Biotek 405LSUVS). Data was normalized as previously described.

HTS of small molecule inhibitors and cell cytotoxicity assay. The Prestwick Chemical Library and Prestwick Drug-Fragment Library Core Set (Prestwick Chemical Inc.) were used for the drug screening, with compounds coming dissolved in dimethyl sulfoxide (DMSO) at a concentration of 10 mM. Two hours prior cell lysis, compounds were solved in DMEM at the appropriate concentration and added to the cells. DMSO treatment was used as a negative control and each experiment was performed in 4 replicates. Pirarubicin (#HY-13725A), Valrubicin (#HY-13772), Idarubicin (#HY-17381), Isomangiferin (#HY-N0772), Mangiferin (#HY-N0290), Aloin (#HY-N0123), Aloin B (#HY-N0886), Licogliflozin (#HY-109092), Puerarin (#HY-N0145), Forskolin (#HY-15371), Tofogliflozin (#HY-13413). Ipragliflozin (#HY-14894), Aloesin (#HY-N2460) and Lancerin (#HY-N2159) were purchased in Clinisciences. CAS nº 14049-11-7 (#Amb10845051), 99907-84-3 (#Amb18667086), 16144-91-5 (#Amb6357187), 17648-03-2 (#Amb533070), 60660-75-5 (#Amb33871605), 81991-99-3 (#Amb34936865) and 108032-11-7 (#Amb33780108) were purchased in GreenPharma. CAS nº 1225897-51-7 (#NS-01845) was purchased in KeyOrganics. To compare the interaction intensities among different drug treatments, an interaction score was calculated: the ratio between treatment and DMSO luminescence intensities was used as normalization and was adjusted using the ratio between each treatment and the control protein expression values.

Cell death assay. To control for cell toxicity among different drug treatments, LDH-Glo™ Cytotoxicity Assay (Promega, #J2380) was used according to the manufacturer's protocol. Data was normalized with 10% SDS treated cells as 100% cell death.

RNA immunoprecipitation. RIP experiments were performed using corresponding HA-tagged DDX50 and DUSP12 501mel cell lines. Approximately 9 million cells seeded in a 15 cm dish were washed once with ice-cold 1× PBS and UV-crosslinked using 800 mJ/cm^2, at 254 nm (Stratalinker,

Stratagene). Cells were then collected and lysed for 10 min on ice in 200 µl of RIPA buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS) with freshly added 1× cOmplete EDTA-free protease inhibitor cocktail (Roche, #11873580001) and 100 U/ml RNAseOUT (Invitrogen, #10777019). Cell lysates were sonicated using a Bioruptor (Diagenode) (6 cycles; 15 s on, 30 s off, M) and centrifuged for 10 min at 15,000 × g. Supernatant were collected and diluted in RQ1- HENG buffer to 1 ml, and 1% was saved as total cell lysate input for western blot analysis. Fifty microliters of anti-HA bead slurry (Pierce, #88836) were washed twice with 1ml of RQ1-HENG buffer and incubated with the cell lysate for 2 h at 4 °C. Following the incubation, beads were washed four times with 1 ml of RQ1-HENG. Proteins were then eluted from the beads in elution buffer (10 mM Tris pH 8, 1 mM EDTA, 1% SDS) for 2 min at 90 °C, followed by proteinase K (Roche, #03115828001) treatment for 30 min at 37 °C. 20% of each eluate was collected for further analysis by western blot. RT-qPCR data was normalized to a housekeeping gene and RNA enrichment in the eluate was displayed relative to input.

Western blot. Protein samples collected from RIP were heat-denaturated at 95 ˚C for 5 min, resolved on 4–12% NuPAGE gels, and transferred to a nitrocellulose membrane (Amersham Protran, GE Healthacre Life Sciences, #10600002) using the wet-transfer system. After blocking with 5% nonfat milk in TPBS (PBS with 0.1% Tween-20) for 1 at room temperature, the membranes were incubated with a primary antibody detecting HA-tag (Cell Signaling, #C29F4, 1:1,000 dilution) at 4 °C overnight. Membranes were washed thrice with TPBS followed by incubation with the secondary antibody conjugated with horseradish peroxidase (Promega, #W401B, 1:5,000 dilution) for 1 h at room temperature and three more washes. Signal was detected using (Amersham Protran, GE Healthacre Life Sciences, #RPN2232) on the Chemidoc MP imaging system (BioRad).

Protein domain analysis. A list of the Pfam domains of each protein was downloaded from the Ensembl site and manually revised. Enrichment of each domain in the set of interactors of a particular RNA vs. the protein library was calculated using the Exact Fisher test and adjusted using FDR.

RNA-seq Data Processing and Differential Expression Testing. Following RNA extraction, 1µg of total RNA was used for RNA-seq library preparation using Stranded mRNA prep Ligation-Illumina according to the manufacturer's recommended protocol. The libraries were sequenced on an Illumina NovaSeq-S1-PE100. For analysis, reads were mapped to Hg19. The differentially expressed genes were analyzed using DESeq2 (1.32.0) (Anders & Huber, 2010) with default settings, and lowly expressed transcripts were removed from the analysis (sum of the number of counts <10 in all samples). Significantly differentially expressed genes were defined as genes with log2 fold-change 0.5 with an adjusted p-value < 0.05. Enrichments analyses were performed with ClusterProfiler (v4.0.5 package) (T. Wu et al., 2021) with an adjusted p-value (FDR) at 0.05.

Motif Analysis. Promoter motif enrichment was performed using HOMER (v4.11). The list of genes was obtained from the significant differentially expressed genes, the promoters were downloaded using Hg19.

HiC and CHIP-seq data. HiC data from human NHEK (GSE63525) and RPMI-7951 (GSE106022) was visualized using HiGlass. CHIP-seq data from transcription factors (GSE61967) and histone marks (GSE94488) was visualised in UCSC Genome Browser.

SUPPLEMENTARY TABLE

Primers for 501mel transcript quantification by RT-qPCR:

GAPDH forward	GAAGGTGAAGGTCGGAGTCAAC
GAPDH reverse	TGATTTTGGAGGGATCTCGCTC
ß-actin forward	TCCTTCCTGGGCATGGAGTC
ß-actin reverse	CGGCAATGCCAGGGTACAT
SOX4 forward	ACTCCTCCTCTTCCTCCTCC
SOX4 reverse	TAAAATCCAGGTCCCGGTCG
h.CASC15 e9b-e11 forward	TGCTTAACATGCATCTTACTGGT
h.CASC15 e9b-e11 reverse	TTTTGATTTTGGCCTTCTGGTCTC
h.CASC15 e11 forward	CCAAAATCAAAAGTATGGGCAGGC
h.CASC15 e12 reverse	CAGTTTTGTGGCAGGTAGGGG
h.CASC15 e12b reverse	AAGAGGAGAGGGCTGAGCAC
DDX50 forward	GGAACTGGCAAACCAAGTAGCC
DDX50 reverse	GATACGACCAGGTGTTCCAACC
DUSP12 forward	AGACGGACCTACTCAGCCATCT
DUSP12 reverse	AGACGGACCTACTCAGCCATCT
CCDC86 forward	AGAATGAGCGGAAGGCAGAGGT
CCDC86 reverse	GTGTCCCGCTTCTCAATGGAGC

Primers for transcript quantification in RIP experiments:

NEAT1 forward	AGTTAGCGACAGGGAGGGATGC
NEAT1 reverse	TGTCCCCTGAAGCCCTGAGCTA
NORAD forward	GATTCATCTTGCCTCGCTGT
NORAD reverse	TGTTTGTGCAGTGGTTCAGG
PVT1 forward	TTACAGGCGTGTGCCACAAAGC
PVT1 reverse	GCCTGTAATCCCAGCACGTTGA

Primers for RNA-MS2 transcript quantification by RT-qPCR:

MS2 forward	TTCTGCAGATATCCAGCACAGT
MS2 reverse	CAACAGATGGCTGGCAACTAGAAG

Generation of *h.CASC15* exon 12 deletion in 501mel cells:

gRNA 5' forward	TATTGCACCATGAGTAAGTG
gRNA 5' reverse	CACTTACTCATGGTGCAATA
gRNA 3' forward	TTTGTTACATACTTCTCCGT
gRNA 3' reverse	ACGGAGAAGTATGTAACAAA
genotyping inside forward	GCCACTGTTTCAGCAACCTCAGCCA
genotyping inside reverse	TGAGACAGCCAGGGAGTTAGCCAGG
genotyping outside forward	GGCTGAATTGTCCCAACACTGAAGAGTTGG
genotyping outside reverse	CCCATTTCTGCTGAGCCATACCACAACC

Generation of HA-tagged RBPs in 501mel cells:

HA-Precission-HIS tag N-term	TACCCATACGATGTTCCAGATTACGCTCTGGAGGTGCTGT TCCAGGGCCCCCATCACCATCACCACCAC
HA-Precission-HIS tag C-term	CATCACCATCACCACCACCTGGAGGTGCTGTTCCAGGGCC CCTACCCATACGATGTTCCAGATTACGCT
gRNA DDX50 N-term forward	TAATGCCTGGGAAACTCCTC
gRNA DDX50 N-term reverse	GAGGAGTTTCCCAGGCATTA
genotyping DDX50 forward	CTTCCTTTCACGCTGTCGCT
genotyping DDX50 reverse	CTAAGAGGGCCAATGAAAACGAA
gRNA DUSP12 N-term forward	TCTAGCCGGAGTCTACTCGA
gRNA DUSP12 N-term reverse	TCGAGTAGACTCCGGCTAGA
genotyping DUSP12 N-t forward	AAGGATGTTCTAGCCGGAGTCTA
genotyping DUSP12 N-t reverse	CTCTCTTGGAAACACGCCTCC
gRNA DUSP12 C-term forward	CCTGTTTTGGGATCACAAAC
gRNA DUSP12 C-term reverse	GTTTGTGATCCCAAAACAGG
genotyping DUSP12 C-t forward	TATGCCACATCCATGAAACCTTC
genotyping DUSP12 C-t reverse	TATCATCTGCAAGTTTCTTCCCA

Primers for sequence amplification and subsequent cloning:

RRE forward	GGGTCTCTCTGGTTAGAC
RRE reverse	ATCAGCCAAAGTGGATCTC
h.CASC15-short forward	GTGAGAATCAAATGATTCAAC
h.CASC15-short reverse	CTCTGACTTACTCTCTGTTTC

Résumé détaillé

CHAPITRE 1

Il y a vingt ans, le programme ENCODE a révélé qu'une fraction substantielle du génome humain est transcrite, bien que moins de 2% de ces ARNs soient codant pour des protéines. Parmi ces transcrits non codants majoritaires, les longs ARN non codants (lncRNAs) représentent l'un des groupes les plus hétérogènes. Les projets de tri les plus récents, tels que FANTOM (v5) et GENCODE (v40), estiment respectivement que le génome humain contient 27 919 et 18 805 gènes lncRNA. Jusqu'à présent, parmi ces lncRNA, seule une petite proportion a été caractérisée fonctionnellement. Dans la plupart des cas, ils ne sont pas étudiés sous l'angle de l'évolution.

L'un de ces lncRNA, *CASC15*, est un transcrit connu pour être impliqué dans le mélanome et le neuroblastome. Le *CASC15* humain (*h.CASC15*) a des lncRNA syntologues chez de nombreuses espèces de vertébrés. Malgré une faible conservation de la séquence au niveau des nucléotides, le laboratoire Shkumatava a décrit que les orthologues du *CASC15* du poisson zèbre (*zf.casc15*) et de l'humain sont conservés fonctionnellement, car tous deux inhibent l'initiation et la progression du mélanome *in vivo*. Mes travaux se concentrent sur l'origine de cette équivalence fonctionnelle, en émettant l'hypothèse que les deux transcrits remplissent la même fonction par le biais d'un mécanisme d'action moléculaire conservé. Puisque les protéines de liaison à l'ARN (RBP) définissent et transmettent la fonction des lncRNA, j'ai supposé que les deux transcrits interagissent avec un ensemble similaire de protéines qui instruisent un rôle biologique compatible dans le contexte du mélanome.

Pour examiner mécaniquement *h.CASC15*, j'ai d'abord cherché à générer un allèle de perte de fonction dans les cellules de mélanome afin d'obtenir des informations précises sur sa fonction dans un contexte cellulaire. Pour générer des clones avec une perte homozygote de l'exon 12 de *h.CASC15*, j'ai utilisé la technologie CRISPR/Cas9. Un examen détaillé du locus, à l'aide des données transcriptomiques des cellules 501mel obtenues ultérieurement, a révélé que l'isoforme principale dans cette lignée cellulaire humaine utilise un site de départ de la transcription (TSS) en aval du TSS principal, ressemblant à l'isoforme *zf.casc15* spécifique du cerveau. Le transcrit qui en résulte, plus court, ne contient que les trois derniers exons et est similaire à ceux précédemment décrits comme *CASC15-S* ou *CASC15-004*, désignés ici comme *h.CASC15-short*. En ce qui concerne les effets de la délétion de l'exon 12 sur la stabilité du transcrit, j'ai observé une abrogation incomplète de l'expression de *h.CASC15*. Un transcrit tronqué résiduel est exprimé à ~30%, générant un mutant hypomorphe *h.CASC15*, ci-après dénommé *h.CASC15-/-*. L'expression n'est pas entièrement abolie, probablement en raison de l'utilisation de l'exon 12b, épissé de manière inefficace dans des conditions non modifiées, ce qui stabilise finalement le transcrit après l'élimination de l'exon 12.

Les lncRNA et les facteurs de transcription situés dans le voisinage génomique ont tendance à se réguler mutuellement au niveau transcriptionnel. Par conséquent, j'ai étudié une fonction potentielle de régulation cis locale du transcrit *h.CASC15*. Lors de la délétion de l'exon 12 de *h.CASC15*, aucun changement d'expression n'a été détecté dans le gène codant pour la protéine SOX4, récapitulant ce qui a été observé chez le poisson zèbre *zf.casc15*-/-. L'insertion aléatoire dépendante de la transposase Tol2 de *h.CASC15* pendant les expériences de sauvetage in vivo soutient également la fonction trans-active de ce lncRNA. Ensemble, ces données montrent que la délétion de l'exon 12 diminue largement l'expression de *h.CASC15* dans les cellules de mélanome.

Pour caractériser davantage le rôle de *h.CASC15* dans les cellules 501mel, j'ai examiné l'impact de la délétion de l'exon 12 sur les propriétés oncogènes de l'allèle cellulaire nouvellement généré. Pour ce faire, j'ai soumis les cellules 501mel à une série de tests cellulaires standard. J'ai d'abord testé l'effet de la délétion sur le comportement prolifératif, invasif et clonogène des cellules 501mel. Cependant, aucune différence n'a été observée en comparant les cellules *h.CASC15*-/- et *h.CASC15*+/+. Enfin, j'ai utilisé le test de cicatrisation pour examiner la migration cellulaire. Dans ce test, deux populations de cellules sont ensemencées à 500 µm de distance et enregistrées afin de mesurer le temps de fermeture de la plaie, ce qui permet de définir leurs capacités de migration respectives. Les cellules *h.CASC15*-/- ont montré un comportement migratoire accru par rapport aux cellules *h.CASC15*+/+, ce qui rappelle l'effet de l'inactivation de *zf.casc15* sur la progression du mélanome chez le poisson zèbre. Pris ensemble, ces résultats indiquent que *h.CASC15* inhibe la migration cellulaire dans le contexte du mélanome.

Pour étudier l'effet global de la déplétion de *h.CASC15* sur l'expression des gènes des cellules de 501mel, j'ai effectué un RNA-seq sur trois clones de contrôle et trois clones knockout, qui se sont regroupés. Avec Florian Constanty, doctorant de l'équipe, nous avons identifié 873 gènes exprimés de manière différentielle. L'analyse des processus biologiques des gènes régulés a révélé que les cellules de la crête neurale et les processus de migration sont des termes sur-représentés. Par exemple, BMP4, EMC10 et ENPP2 ont été impliqués dans la migration et la motilité des cellules dans différents modèles de développement et de cancer. En conclusion, l'analyse transcriptomique de nos cellules mutantes est cohérente avec les changements comportementaux observés dans les cellules mutantes *h.CASC15*, révélant une réactivation des signatures génétiques de la crête neurale et de la migration.

J'ai ensuite voulu aborder le mécanisme moléculaire à l'origine de la conservation fonctionnelle de *h.CASC15* et *zf.casc15*. Le transcrit n'a pas de sites de liaison miRNA prédits et des expériences précédentes constituent des preuves in vivo et in cellulo que les orthologues de *CASC15* régulent la progression du mélanome en trans. J'ai donc émis l'hypothèse que les deux lncRNA se lieraient à des RBP équivalents qui piloteraient leur fonction. Pour caractériser fonctionnellement les orthologues de *CASC15*, j'ai cherché à identifier de manière impartiale l'interactome protéique complet des deux transcrits synténiques. Pour ce faire, j'ai utilisé la technique développée au laboratoire nommée in cell

protein-RNA interaction ou incPRINT. Elle est basée sur une immunoprécipitation à haut débit des RBP suivie d'une détection et d'une quantification par luciférase de leurs interactions avec l'ARN d'intérêt. Cette technique permet de cribler la liaison intracellulaire de 3 000 protéines humaines, dont des RBP, des facteurs de transcription et des modificateurs de la chromatine.

J'ai testé *h.CASC15* et *zf.casc15* avec la bibliothèque de RBP humaines en utilisant incPRINT. En comparant les deux ensembles de données, 15 protéines ont été détectées pour interagir spécifiquement avec *h.CASC15* et *zf.casc15*. Ensuite, j'ai cherché à déterminer si l'un des 15 interacteurs protéiques communs était essentiel à la fonction de l'ARNmc. Pour ce faire, j'ai délété individuellement chacune des RBP dans les cellules 501mel *h.CASC15*+/+ et j'ai testé leur comportement migratoire à l'aide du test de cicatrisation. La déplétion de DDX50 et de DUSP12, mais pas des autres interacteurs, a récapitulé la migration cellulaire accrue des cellules 501mel *h.CASC15*-/-. Cela suggère que ces protéines sont impliquée dans la fonction d'inhibiteur du mélanome exercer par h.*CASC15*. Ensemble, ces résultats démontrent que DDX50 et DUSP12 sont des partenaires fonctionnels de *h.CAS15* et *zf.casc15*. Dans le but de savoir si la liaison se produit également dans les conditions natives du mélanome, j'ai commencé par une validation orthogonale de l'interaction dans des conditions endogènes. Pour ce faire, je réalise actuellement une immunoprécipitation d'ARN (RIP) suivie d'analyses RT-qPCR.

L'interaction des RBP communes avec les syntologues de *CASC15* se produit malgré un manque de conservation des séquences nucléotidiques alignables. Ainsi, j'ai émis l'hypothèse que de courtes séquences conservées représentant des motifs d'ARN fonctionnels dans les deux transcrits lncRNA orthologues sont responsables de la liaison des RBP, permettant leur fonctionnalité équivalente. Afin d'identifier les éléments d'ARN qui sont conservés de *zf.casc15* à *h.CASC15*, avec Caroline Ross et Igor Ulitsky (Weizmann Institute of Science, Rehovot, Israël), nous avons effectué une analyse des motifs d'ARN par LncLOOM (lncRNA linear ordered conserved motifs). Cet algorithme récemment développé utilise les séquences des lncRNA synténiques de différentes espèces pour rechercher des k-mers courts, ordonnés et conservés au cours de l'évolution.

L'identification et le tri des séquences orthologues de *CASC15* provenant de huit espèces de vertébrés, localisées à l'aide de données RNA-seq, et l'exploration ultérieure à l'aide de LncLOOM ont révélé quatre motifs conservés de séquences d'ARN. Le tronçon super-conservé situé au niveau du premier exon est susceptible d'interférer avec la découverte d'éléments pertinents distribués le long du transcrit. Il convient de noter que le transcrit *h.CASC15-MS2* utilisé pour l'expérience de sauvetage du poisson zèbre et dans incPRINT ne contenait pas le premier exon. Nous avons ensuite effectué une analyse LncLOOM des orthologues de *CASC15* avant la suppression du premier exon dans toutes les espèces, ce qui n'a donné aucun motif conservé, excepté pour le poisson zèbre. Dans le but d'affiner la recherche, nous avons cherché à utiliser *h.CASC15*-short ainsi que les isoformes équivalents plus courts spécifiques de l'ectoderme chez d'autres espèces. Cependant, le fait que certains isoformes aient été

épissés de manière inefficace, comme le *Casc15* de 20 kb de la souris, a rendu impossible la détection par l'algorithme LncLOOM de motifs de séquence d'ARN conservés. Ensemble, les résultats ci-dessus indiquent l'impossibilité actuelle d'identifier les motifs d'ARN conservés entre *h.CASC15* et *zf.casc15* en raison des limites de l'algorithme.

Pour résumer, afin d'aborder le mécanisme d'action moléculaire des orthologues de *CASC15* dans le cadre du mélanome, j'ai cherché à identifier les RBP interagissant avec *zf.casc15* et *h.CASC15* qui pourraient finalement sous-tendre leur fonctionnalité conservée. Grâce à la caractérisation des protéomes d'ARN interagissant avec le poisson zèbre et l'homme et à une comparaison entre espèces, j'ai identifié un ensemble de quinze protéines comme interacteurs commun des orthologues de *CASC15*. Sur la base de ces résultats, j'ai cherché à savoir si la liaison de l'une de ces protéines était suffisante pour assurer la fonction de *CASC15*. Pour ce faire, j'ai effectué un knockdown de chaque RBP respectivement dans des cellules de mélanome et caractérisé leur migration. J'ai observé que le knockdown de DDX50 et de DUSP12 était suffisant pour provoquer une augmentation de la migration cellulaire et qu'elles étaient donc considérées comme des interacteurs fonctionnels de *CASC15*. Alors que DDX50 et DUSP12 interagissent avec notre transcrit d'intérêt, le mécanisme précis par lequel ces RBP médient la fonction inhibitrice du mélanome *CASC15* reste à explorer. Au cours des prochains mois, nos efforts seront dirigés vers une compréhension précise de ce mécanisme.

Pour affirmer avec certitude que le mécanisme d'action moléculaire est conservé entre les orthologues *CASC15* du poisson zèbre et de l'homme, d'autres expériences sont nécessaires pour caractériser complètement *zf.casc15*. Il reste à décrire si la localisation subcellulaire du lncRNA du poisson zèbre diffère de celle de l'humain. En ce qui concerne les interacteurs protéiques, l'une des limites de cette étude est l'utilisation des RBP humaines pour l'identification de l'interactome des lncRNA d'autres espèces. Comme le poisson zèbre ne possède pas d'orthologue de DDX50, l'interaction in vivo entre *zf.casc15* et le paralogue de Ddx21 du poisson zèbre doit encore être étudiée. Il en va de même pour le zebrafish dusp12. De plus, l'évaluation du comportement migratoire des cellules 501mel lors de la déplétion de DDX21, qui est également un des principaux interacteurs de *h.CASC15*, pourrait permettre d'aborder les différences entre les deux paralogues dans les cellules humaines.

Les LncRNAs sont des cibles thérapeutiques inexploitées et pourraient offrir des opportunités pour de nouvelles stratégies médicamenteuses. Ceci est particulièrement pertinent pour les transcrits oncogènes ou favorisant le cancer. Il a été démontré que la structure en triple hélice de *MALAT1* peut être ciblée, ce qui entraîne la déstabilisation et la régulation négative du lncRNA. Il a également été démontré que le ciblage du lncRNA *SAMMSON* supprime la croissance et la prolifération d'un modèle de mélanome dérivé de patients sous forme de xénogreffe. Mais bien qu'ils soient considérés depuis longtemps comme des cibles médicamenteuses, aucun médicament ciblant les lncRNA n'est encore entré dans les essais cliniques.

Dans la nouvelle ère de la médecine par ARN, le potentiel de l'utilisation du transcrit *CASC15* comme ARN thérapeutique pour ralentir la progression du mélanome ne peut être sous-estimé. Un exemple remarquable de thérapie basée sur les lncRNA est le *HULC*. On a découvert que ce transcrit facilite l'activité des enzymes du métabolisme de la phénylalanine, et le traitement avec un agent thérapeutique ARN basé sur *HULC* a réduit l'excès de phénylalanine dans un modèle de phénylcétonurie chez la souris. Dans ce sens, étant donné la fonction inhibitrice du mélanome de *CASC15*, on pourrait envisager une thérapie à base d'ARN *CASC15* administrée topiquement en attendant la résection de la tumeur, qui pourrait être utilisée aussi bien comme traitement préventif. Certes, cette stratégie devrait être évaluée dans des modèles murins de mélanome.

Une autre approche possible pour ralentir la progression du mélanome serait de cibler les interactions *CASC15* - RBP. Ceci est bien illustré par le lncRNA *GAS5* et la protéine UPF1, dont l'interaction induit la dégradation du transcrit. Le blocage de la liaison de cette RBP a augmenté la demi-vie de *GAS5*, stimulant finalement le récepteur de l'insuline et induisant l'absorption du glucose. À la lumière de nos résultats, un composé pourrait théoriquement inhiber la progression du mélanome en stimulant la liaison entre *CASC15* et ses partenaires protéiques DDX50 ou DUS12. De tels stabilisateurs de l'interaction ARN-RBP ont déjà été décrits pour d'autres ARN hélicases et facteurs d'épissage. De plus, l'absence d'expression de *CASC15* dans les mélanocytes normaux assurerait une spécificité de cible de cette stratégie. Comme nous le verrons en détail dans le chapitre suivant, le ciblage des interactions entre l'ARN et les protéines apparaît comme un scénario thérapeutique prometteur.

CHAPITRE 2

La leucémie myéloïde aiguë (LMA) est un cancer du sang agressif et très répandu chez l'adulte. Les cellules souches leucémiques (CSL) sont à l'origine de la maladie et la propagent, mais elles sont souvent résistantes aux traitements : même si la plupart des patients répondent initialement à la chimiothérapie, les rechutes sont fréquentes et environ 75 % d'entre eux meurent dans les cinq ans suivant le diagnostic. Par conséquent, l'identification de nouvelles cibles thérapeutiques spécifiques pour l'élimination des CSL est un besoin clinique non satisfait.

Il a été récemment démontré que YTHDF2, l'une des protéines de lecture qui interagit avec les ARN modifiés m^6A, est surexprimée dans les LMA et est requise pour l'initiation et la propagation de la maladie. L'inactivation génétique de la protéine de liaison à l'ARN YTHDF2 compromet sélectivement la LSC in vivo tout en préservant l'hématopoïèse normale, faisant de cette protéine une cible thérapeutique unique. YTHDF2 est l'un des "lecteurs" de la N6-méthyl-adénosine (m^6A), la modification post-transcriptionnelle de l'ARN la plus abondante dans les ARNm et les lncRNA chez les eucaryotes. Notamment, YTHDF2 a plus de 3 000 cibles ARN cellulaires qui contiennent la séquence consensus [G/A/U] [G>A] m^6A [C] [U>A>C], appelée motif DRACH. Malgré la fonction clé

de YTHDF2 dans la LMA, aucun médicament n'a encore été identifié pour cibler YTHDF2. Compte tenu de son rôle biologique crucial dans la LMA, nous avons cherché à trouver des inhibiteurs des interactions entre YTHDF2 et l'ARN modifié par m^6A en utilisant notre technologie quantitative incPRINT.

La nature quantitative du test de luciférase incPRINT permet, en théorie, de mesurer une perturbation de l'interaction ARN-protéine. La perturbation de la liaison, soit par une mutation dans l'ARN, soit dans la protéine, soit en raison de son inhibition par de petites molécules, serait mesurée par une diminution de l'intensité de l'activité luciférase. Cependant, la question de savoir si la technologie incPRINT possède une résolution suffisante pour détecter l'impact de telles perturbations n'avait jamais été testée jusqu'à présent. Pour adapter l'incPRINT à une méthode de criblage de médicaments, ci-après dénommée incPRINT-Drug, j'ai cherché à trouver des interactions ARN-protéines de preuve de concept qui soient (i) ciblables et (ii) cliniquement pertinentes. Étant donné le rôle biologique crucial de YTHDF2 dans le développement de la LMA, j'ai évalué la liaison entre YTHDF2 et les ARN contenant du m^6A.

Je me suis concentré sur YTHDF2 et sur l'interaction avec l'une de ses cibles, le récepteur du facteur de nécrose tumorale de la superfamille 2 (*Tnfrsf2*). Leur liaison a pu être détectée de manière efficace et reproductible par incPRINT, montrant une différence neuf fois supérieure à la GFP. Afin de déterminer la gamme dynamique d'une inhibition de liaison détectable par incPRINT, j'ai suivi une approche de mutagenèse. Il a été démontré que la mutation ponctuelle W491A dans la poche aromatique à trois tryptophanes de YTHDF2, qui constitue le site de reconnaissance de la m^6A, diminue significativement de dix fois son affinité de liaison in vitro. Ainsi, j'ai comparé la capacité de YTHDF2 de type sauvage (YTHDF2WT) ou de YTHDF2W491A à interagir avec *Tnfrsf2*, pour tester si une diminution significative de la liaison peut être détectée par incPRINT. L'intensité d'interaction de Tnfrsf2 et de la protéine mutée a diminué de plus de deux fois par rapport au type sauvage, alors que les niveaux d'expression de la protéine sont restés inchangés. La liaison restante détectée de la protéine mutante résulte probablement d'une liaison non spécifique aux boucles 10xMS2. Ces données démontrent que *Tnfrsf2* - YTHDF2 constitue une bonne interaction de preuve de concept pour évaluer l'application de incPRINT-Drug.

Plus tard, j'ai établi les conditions optimales pour le criblage de incPRINT-Drug : j'ai ajouté une étape de traitement médicamenteux de 2 h, 10 µM, avant la lyse des cellules. Pour cribler les inhibiteurs de la liaison *Tnfrsf2* - YTHDF2, j'ai utilisé la bibliothèque chimique de Prestwick. Cette collection de ~1200 médicaments hors brevet approuvés par la FDA présente une grande diversité chimique et pharmacologique et garantit une biodisponibilité et une faible toxicité cellulaire. Par ailleurs, la dimension réduite de la bibliothèque m'a permis de réaliser le crible à trous en une seule semaine.

En utilisant cette approche, j'ai identifié sept composés qui ont montré une diminution de l'interaction *Tnfrsf2* - YTHDF2WT comparable au niveau de l'interaction avec YTHDF2W491A. Pour comparer les données entre les différents composés, les scores d'interaction ont été calculés en tenant compte à la fois des changements de luminescence - représentant les intensités d'interaction - ainsi que des différences dans les valeurs ELISA - contrôlant les effets des tests sur les niveaux d'expression des protéines. Après l'identification à haut débit des inhibiteurs de Tnfrsf2 - YTHDF2, j'ai validé les hits identifiés par incPRINT-Drug à petite échelle. Cette expérience a permis de confirmer la spécificité de l'inhibition et d'écarter tout artefact potentiel de normalisation des données. Celles-ci soulignent également la spécificité de l'inhibition puisque la plupart des composés n'ont pas inhibé la liaison.

Pour valider davantage les hits identifiés et s'assurer qu'ils n'agissent pas de façon promiscue, ainsi que pour comprendre la nature de leurs effets inhibiteurs, j'ai établi une série de contrôles rigoureux dans chacune des étapes d'incPRINT-Drug. À cette fin, j'ai d'abord effectué des tests de dépendance à la concentration. Une série de concentrations différentes de chaque composé a été testée, ce qui a confirmé une dépendance de l'effet de cinq des sept composés à la dose. L'aminacrine et le topotécan ont été écartés comme autres candidats car leur effet inhibiteur sur l'interaction *Tnfrsf2* - YTHDF2 est resté le même quelle que soit la concentration utilisée, ce qui suggère qu'ils ne sont probablement pas des inhibiteurs spécifiques de l'interaction *Tnfrsf2* - YTHDF2WT. D'autre part, la mitoxantrone, la doxorubicine, la daunorubicine, l'épirubicine et l'haloprogin ont présenté une inhibition dose-dépendante, confirmant leur spécificité.

Bien qu'elles soient approuvées par la FDA, ces petites molécules pourraient affecter l'intégrité cellulaire. Pour exclure cette interférence potentielle sur le test, j'ai testé la cytotoxicité des composés. En observant la mort cellulaire après différents traitements, je n'ai identifié aucun changement entre les cinq composés et le contrôle DMSO. Ce résultat corrobore un effet indépendant de la mort cellulaire de la diminution de l'interaction par les petites molécules identifiées.

Notamment, incPRINT-Drug peut détecter des changements dans les niveaux d'expression des protéines qui sont corrigés a posteriori par le calcul du score d'interaction. Cependant, bien que le traitement médicamenteux dure 2h, les composés pourraient compromettre la transcription de l'ARN ou la stabilité du transcrit et entraîner un résultat artéfactuel. Une diminution de la quantité de transcription marquée MS2 aurait un impact sur le signal de luminescence et serait interprétée comme une inhibition de l'interaction, ce qui constituerait un faux positif. Comme le score d'interaction ne tient pas compte des changements au niveau des transcriptions, j'ai analysé les niveaux de *Tnfrsf2* par RT-qPCR et exclu l'effet des médicaments sur les niveaux d'ARN. Ensemble, ces résultats ont définis comme meilleurs candidats cinq petites molécules.

J'ai ensuite profité de la polyvalence de incPRINT-Drug pour établir des contrôles de spécificité de l'ARN et des protéines, ce qui m'a permis de mieux comprendre la cible cellulaire avec laquelle ces

composés s'engagent. J'ai d'abord évalué la spécificité de l'ARN en testant d'autres transcrits cibles et non cibles connus de YTHDF2. Comme contrôle négatif, j'ai sélectionné l'ARN *Ly6A* faiblement modifié par m^{6}A. Quatre des petites molécules n'ont montré aucune diminution de la liaison *Ly6A* - YTHDF2WT, ce qui suggère un effet inhibiteur dépendant de m^{6}A -. En revanche, l'haloprogine a montré une diminution de l'intensité de l'interaction, indiquant un effet non spécifique potentiel. Ensuite, la transcription hautement méthylée Gadd45G a été utilisée comme alternative pour l'interaction avec Tnfrsf2 et on s'attendait à ce qu'elle se comporte comme telle lors du traitement médicamenteux. Les cinq composés principaux ont inhibé *Gadd45G* - YTHDF2WT de manière similaire par rapport à *Tnfrsf2*.

Ensuite, j'ai testé les médicaments en utilisant les interactions *Tnfrsf2* - YTHDF2W491A ainsi que les interactions *Tnfrsf2* - MOV10 non liées au m^{6}A. L'intensité de l'interaction de fond YTHDF2W491A n'a pas réagi au traitement médicamenteux de quatre médicaments spécifiques, bien que l'haloprogin ait montré à nouveau une diminution artéfactuelle de l'intensité de l'interaction. De même, la liaison de MOV10 était peu ou pas affectée, sauf pour l'haloprogine. Ensemble, ces résultats indiquent que quatre composés principaux empêche spécifiquement l'inhibition YTHDF2 - m^{6}A, et que les tests de spécificité incPRINT-Drug peuvent être utilisés pour distinguer les vrais hits des faux positifs.

Il est important de noter que le domaine YTH des trois composants de la famille YTHDF est très similaire dans la composition de la séquence d'acides aminés. Des études structurelles ont identifié la cage aromatique conservée quasi-identique qui reconnaît la fraction méthyle de la modification m^{6}A, qui est composée de trois tryptophanes. De plus, comme discuté ci-dessus, les paralogues de YTHDF se lient aux transcrits m^{6}A de manière équivalente. Ainsi, en faisant l'hypothèse que les composés ne reconnaîtraient pas seulement la poche de toutes les protéines à domaine YTH, j'ai testé si l'interaction de *Tnfrsf2* - YTHDF1WT et *Tnfrsf2* - YTHDF3WT pouvait être ciblée par incPRINT-Drug. YTHDF2WT

J'ai commencé à tester l'interaction des protéines de type sauvage et muté avec *Tnfrsf2*, de la même manière que YTHDF2 a été testé en premier. Les intensités d'interaction se sont comportées de manière similaire entre les trois cas, suggérant qu'une baisse de l'interaction due aux traitements médicamenteux serait effectivement détectée. Cependant, une comparaison des niveaux d'expression de la protéine YTHDF2 a montré une chute de quatre à sept fois de l'expression des autres membres de la famille dans la configuration incPRINT. Une différence dans l'expression des protéines mutées par rapport à leur type sauvage respectif a également pu être observée.

Bien que ces conditions soient sous-optimales, j'ai testé si les quatre composés pouvaient inhiber l'interaction avec d'autres domaines YTH. Le traitement de *Tnfrsf2* - YTHDF1WT avec trois des quatre médicaments a récapitulé la diminution de l'interaction de près de deux fois. Pour *Tnfrsf2* - YTHDF3WT, la faible expression de la protéine a rencontré la résolution d'incPRINT-Drug, empêchant

la détection de l'interaction dans le cas de la mitoxantrone. Ces résultats indiquent, dans une certaine mesure, que les quatre composés principaux inhibent l'interaction du domaine YTH avec les transcrits modifiés par m^6A.

La doxorubicine, la daunorubicine et l'épirubicine appartiennent à la famille des rubicines de composés naturels, également appelés anthracyclines, et la mitoxantrone est un médicament synthétisé artificiellement qui ressemble aux rubicines. Pour comprendre si la structure centrale des rubicines est responsable de l'inhibition de la m^6A - YTHD décrite ci-dessus, j'ai effectué une analyse structure-fonction indirecte. En testant d'autres membres de cette famille de composés, j'ai voulu déterminer si l'effet inhibiteur était universel pour tous les membres, et si des modifications chimiques des rubicines confèrent aux petites molécules un effet inhibiteur plus élevé. J'ai effectué incPRINT-Drug avec trois congénères supplémentaires des rubicines. J'ai identifié l'inhibition de l'idarubicine comme étant comparable à celle des composés principaux. Bien qu'une diminution de l'interaction ait pu être observée pour le traitement par la pirarubicine, l'effet était plutôt faible. Ce résultat exclut le sous-groupe structurel central des rubicines comme suffisant pour l'inhibition de Tnfrsf2 - YTHDF2WT et confirme que incPRINT-Drug est sensible aux petites modifications des composés.

Il est notable que les rubicines soient actuellement utilisées comme traitement standard de la leucémie, ce qui n'a pas changé de manière substantielle ces dernières années, bien que leur mécanisme n'ait jamais été lié à YTHDF2. Quoique controversée, la cible principale proposée des rubicines est l'empoisonnement de la topoisomérase 2. Afin de découpler le mécanisme alternatif dépendant de YTHDF2 que nous proposons de la cible décrite ci-dessus, j'ai suivi une approche de sélection de médicaments basée sur les fragments. J'ai émis l'hypothèse que des molécules plus simples, ou fragments, définis comme des médicaments contenant moins de 20 atomes non hydrogène, maintiendraient l'inhibition de l'interaction *Tnfrsf2*-YTHDF2 ou même s'orienteraient mieux dans le site de liaison, tout en évitant les interactions hors cible.

Pour identifier la partie essentielle des rubicines qui agit comme inhibiteur du YTHDF2, j'ai testé environ 500 chémotypes, ou motifs de structure chimique. Cette bibliothèque de fragments Core-Set comprend divers composés à sous-structure primaire dérivés de petites molécules présentes dans la bibliothèque chimique de Prestwick. Ces fragments sont plus petits et plus simples que ceux que j'ai testés précédemment. Cette chimiothèque comprenait un fragment dérivé de nos quatre molécules principales (n° 1884). En outre, j'ai testé cinq fragments personnalisés dérivés de la rubicine. Aucun des fragments de la bibliothèque ou des fragments supplémentaires n'a montré un comportement inhibiteur comparable à celui de la *Tnfrsf2* - YTHDF2W491A.

J'ai ensuite adopté une approche complémentaire de sélection de médicaments basée sur les ligands en effectuant une recherche computationnelle de sous-structures. Les composés actifs de la rubicine contiennent un sous-graphe commun maximal qui peut être tronqué. Afin d'identifier les portions de

liaison possibles de la rubicine et de voir si un découpage est possible, Steven Shave et Manfred Auer (Université d'Édimbourg) ont effectué une recherche de sous-structure sur le système cyclique diol conservé et la portion tétrahydropyrane. Grâce à cette approche in silico, 19 composés ont été identifiés comme contenant un sous-graphe commun maximal de rubicine. Quatre d'entre eux ont inhibé la liaison YTHDF2 - m^6A dans l'incPRINT-Drug, et seules la mangiférine et la licogliflozine l'ont fait de manière significative. À noter que les deux inhibitions commencent à être significatives à 33µM, par rapport à la concentration de 4 µM trouvée sur les composés principaux.

Depuis la découverte de la modification transcriptomique globale m^6A, des efforts importants ont été déployés pour comprendre son mécanisme précis. Il est maintenant bien établi qu'un ensemble raffiné de RBP régule le dépôt, l'élimination et la lecture de m^6A. L'interactome m^6A est resté inexploité en tant que cible thérapeutique, principalement en raison d'un manque de technologie pour l'identification des médicaments, mais aussi parce qu'il est apparu comme cliniquement pertinent il n'y a pas longtemps. Dès lors, plusieurs sociétés de biotechnologie et laboratoires universitaires se sont lancés dans une course à la découverte de petites molécules permettant de traiter les protéines modifiant l'ARN. Récemment, l'interaction *Tnfrsf2* - YTHDF2 a été identifiée comme une cible thérapeutique clé pour le traitement de la LMA. Mon travail présente la personnalisation d'une méthodologie de criblage in cellulo pour la découverte de médicaments ciblant des interactions ARN méthylés - YTHDF2, qui s'est concrétisée par l'identification de petites molécules inhibitrices de cette interaction ARN-protéine.

Les résultats ci-dessus démontrent la capacité rapide et robuste de la technique incPRINT à quantifier les interactions ARN-RBP dans les cellules. Le système de détection de la luciférase est présenté ici comme une approche sensible pour la détection des changements dans la liaison ARN-protéine. Nos résultats démontrent également le grand potentiel de incPRINT-Drug comme méthode rapide in cellulo pour la découverte de petites molécules inhibant les interactions ARN-protéines. Le même pipeline pourrait être utilisé pour identifier des composés stimulant la liaison ARN des RBP, comme l'interaction anti-tumorigène entre *CASC15* et les protéines DDX50 et DUSP12.

En réalité, ces applications ne sont que le sommet de l'iceberg d'une vague entière d'implémentations alternatives pour lesquelles incPRINT-Drug pourrait servir de modèle. Il est intéressant de spéculer que cette méthode pourrait être utilisée pour cribler des agents à base de nucléotides ciblant des interactions ARN-protéines spécifiques causant des maladies. En théorie, incPRINT-Drug pourrait également être utilisé pour étudier et cibler un spectre plus large d'interactions, comme les protéines de détection de l'ARN localisées dans la membrane, également responsables de nombreux processus pathologiques. En ce qui concerne le travail présenté ici, une fois que l'évolutivité et la rentabilité d'incPRINT seront résolues, l'identification de nouveaux composés ciblant l'interaction *Tnfrsf2* - YTHDF2 et l'évaluation ultérieure de l'inhibition pharmacologique in vivo deviendront certainement une réalité.

BIBLIOGRAPHY

Achsel, T., & Bagni, C. (2016). Cooperativity in RNA-protein interactions: The complex is more than the sum of its partners. *Current Opinion in Neurobiology*, *39*, 146–151. https://doi.org/10.1016/j.conb.2016.06.007

Adams, D., Gonzalez-Duarte, A., O'Riordan, W. D., Yang, C.-C., Ueda, M., Kristen, A. v., Tournev, I., Schmidt, H. H., Coelho, T., Berk, J. L., Lin, K.-P., Vita, G., Attarian, S., Planté-Bordeneuve, V., Mezei, M. M., Campistol, J. M., Buades, J., Brannagan, T. H., Kim, B. J., … Suhr, O. B. (2018). Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis. *New England Journal of Medicine*, *379*(1), 11–21. https://doi.org/10.1056/nejmoa1716153

Aguilar, R., Spencer, K. B., Kesner, B., Rizvi, N. F., Badmalia, M. D., Mrozowich, T., Mortison, J. D., Rivera, C., Smith, G. F., Burchard, J., Dandliker, P. J., Patel, T. R., Nickbarg, E. B., & Lee, J. T. (2022). Targeting Xist with compounds that disrupt RNA structure and X inactivation. *Nature*, *604*(7904), 160–166. https://doi.org/10.1038/s41586-022-04537-z

Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L. A., & Rubin, E. M. (2007). Deletion of ultraconserved elements yields viable mice. *PLoS Biology*, *5*(9), 1906–1911. https://doi.org/10.1371/journal.pbio.0050234

Aillaud, M., & Schulte, L. N. (2020). Emerging roles of long noncoding rnas in the cytoplasmic milieu. *Non-Coding RNA*, *6*(4), 1–14. https://doi.org/10.3390/ncrna6040044

Akay, A., Jordan, D., Navarro, I. C., Wrzesinski, T., Ponting, C. P., Miska, E. A., & Haerty, W. (2019). Identification of functional long non-coding RNAs in C. elegans. *BMC Biology*, *17*(1). https://doi.org/10.1186/s12915-019-0635-7

Alarcón, C. R., Lee, H., Goodarzi, H., Halberg, N., & Tavazoie, S. F. (2015). N6-methyladenosine marks primary microRNAs for processing. *Nature*, *519*(7544), 482–485. https://doi.org/10.1038/nature14281

Alkallas, R., Lajoie, M., Moldoveanu, D., Hoang, K. V., Lefrançois, P., Lingrand, M., Ahanfeshar-Adams, M., Watters, K., Spatz, A., Zippin, J. H., Najafabadi, H. S., & Watson, I. R. (2020). Multi-omic analysis reveals significantly mutated genes and DDX3X as a sex-specific tumor suppressor in cutaneous melanoma. *Nature Cancer*, *1*(6), 635–652. https://doi.org/10.1038/s43018-020-0077-8

Allou, L., Balzano, S., Magg, A., Quinodoz, M., Royer-Bertrand, B., Schöpflin, R., Chan, W. L., Speck-Martins, C. E., Carvalho, D. R., Farage, L., Lourenço, C. M., Albuquerque, R., Rajagopal, S., Nampoothiri, S., Campos-Xavier, B., Chiesa, C., Niel-Bütschi, F., Wittler, L., Timmermann, B., … Superti-Furga, A. (2021). Non-coding deletions identify Maenli lncRNA as a limb-specific En1 regulator. *Nature*, *592*(7852), 93–98. https://doi.org/10.1038/s41586-021-03208-9

Alonso, A., Rahmouni, S., Williams, S., van Stipdonk, M., Jaroszewski, L., Godzik, A., Abraham, R. T., Schoenberger, S. P., & Mustelin, T. (2003). Tyrosine phosphorylation of VHR phosphatase by ZAP-70. *Nature Immunology*, *4*(1), 44–48. https://doi.org/10.1038/ni856

Andergassen, D., & Rinn, J. L. (2021). From genotype to phenotype: genetics of mammalian long non-coding RNAs in vivo. *Nature Reviews Genetics*, *23*, 229–243. https://doi.org/10.1038/s41576-021-00427-8

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

Andersen, R. E., Hong, S. J., Lim, J. J., Cui, M., Harpur, B. A., Hwang, E., Delgado, R. N., Ramos, A. D., Liu, S. J., Blencowe, B. J., & Lim, D. A. (2019). The Long Noncoding RNA Pnky Is a Trans-acting Regulator of Cortical Development In Vivo. *Developmental Cell*, *49*(4), 632-642.e7. https://doi.org/10.1016/j.devcel.2019.04.032

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

Arcamone, F., Gaetani, M., Scotti, T., & Dimarco, A. (1961). Isolamento ed attivita antitumorale di un antibiotico da Streptomyces sp. *Giornale Di Microbiologia*, *9*(2), 83.

Arun, G., Diermeier, S., Akerman, M., Chang, K. C., Wilkinson, J. E., Hearn, S., Kim, Y., MacLeod, A. R., Krainer, A. R., Norton, L., Brogi, E., Egeblad, M., & Spector, D. L. (2016). Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes and Development*, *30*(1), 34–51. https://doi.org/10.1101/gad.270959.115

Ashburn, T. T., & Thor, K. B. (2004). Drug repositioning: Identifying and developing new uses for existing drugs. *Nature Reviews Drug Discovery*, *3*(8), 673–683. https://doi.org/10.1038/nrd1468

Aspengren, S., Hedberg, D., Sköld, H. N., & Wallin, M. (2008). New Insights into Melanosome Transport in Vertebrate Pigment Cells. In *International Review of Cell and Molecular Biology* (Vol. 272, pp. 245–302). https://doi.org/10.1016/S1937-6448(08)01606-7

Athie, A., Marchese, F. P., González, J., Lozano, T., Raimondi, I., Juvvuna, P. K., Abad, A., Marin-Bejar, O., Serizay, J., Martínez, D., Ajona, D., Pajares, M. J., Sandoval, J., Montuenga, L. M., Kanduri, C., Lasarte, J. J., & Huarte, M. (2020). Analysis of copy number alterations reveals the lncRNA ALAL-1 as a regulator of lung cancer immune evasion. *Journal of Cell Biology*, *219*(9). https://doi.org/10.1083/JCB.201908078

Auweter, S. D., Oberstrass, F. C., & Allain, F. H. T. (2006). Sequence-specific binding of single-stranded RNA: Is there a code for recognition? *Nucleic Acids Research*, *34*(17), 4943–4959. https://doi.org/10.1093/nar/gkl620

Baggiolini, A., Callahan, S. J., Montal, E., Weiss, J. M., Trieu, T., Tagore, M. M., Tischfield, S. E., Walsh, R. M., Suresh, S., Fan, Y., Campbell, N. R., Perlee, S. C., Saurat, N., Hunter, M. v., Simon-Vermot, T., Huang, T. H., Ma, Y., Hollmann, T., Tickoo, S. K., … White, R. M. (2021). Developmental chromatin programs determine oncogenic competence in melanoma. *Science*, *373*(6559), eabc1048. https://doi.org/10.1126/science.abc1048

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, *37*(Web Server issue), W202–W208. https://doi.org/10.1093/nar/gkp335

Ballarino, M., Cipriano, A., Tita, R., Santini, T., Desideri, F., Morlando, M., Colantoni, A., Carrieri, C., Nicoletti, C., Musarò, A., Carroll, D. O., & Bozzoni, I. (2018). Deficiency in the nuclear long noncoding RNA Charme causes myogenic defects and heart remodeling in mice. *The EMBO Journal*, *37*(18), e99697. https://doi.org/10.15252/embj.201899697

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

Banerjee, D., Gryder, B., Bagchi, S., Liu, Z., Chen, H.-C., Xu, M., Sun, M., Vaksman, Z., Diskin, S. J., Khan, J., & Thiele, C. J. (2020). Lineage specific transcription factor waves reprogram neuroblastoma from self-renewal to differentiation. *Preprint at BioRxiv*. https://doi.org/10.1101/2020.07.23.218503

Bansal, H., Yihua, Q., Iyer, S. P., Ganapathy, S., Proia, D., Penalva, L. O., Uren, P. J., Suresh, U., Carew, J. S., Karnad, A. B., Weitman, S., Tomlinson, G. E., Rao, M. K., Kornblau, S. M., & Bansal, S. (2014). WTAP is a novel oncogenic protein in acute myeloid leukemia. *Leukemia*, *28*(5), 1171–1174. https://doi.org/10.1038/leu.2014.16

Barbieri, I., Tzelepis, K., Pandolfini, L., Shi, J., Millán-Zambrano, G., Robson, S. C., Aspris, D., Migliori, V., Bannister, A. J., Han, N., de Braekeleer, E., Ponstingl, H., Hendrick, A., Vakoc, C. R., Vassiliou, G. S., & Kouzarides, T. (2017). Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control. *Nature*, *552*(7683), 126–131. https://doi.org/10.1038/nature24678

Barra, J., & Leucci, E. (2017). Probing long non-coding RNA-protein interactions. *Frontiers in Molecular Biosciences*, *4*(JUL). https://doi.org/10.3389/fmolb.2017.00045

Batista, P. J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D. M., Lujan, E., Haddad, B., Daneshvar, K., Carter, A. C., Flynn, R. A., Zhou, C., Lim, K. S., Dedon, P., Wernig, M., Mullen, A. C., Xing, Y., Giallourakis, C. C., & Chang, H. Y. (2014). M6A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell*, *15*(6), 707–719. https://doi.org/10.1016/j.stem.2014.09.019

Beckmann, B. M., Castello, A., & Medenbach, J. (2016). The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *Pflugers Archiv European Journal of Physiology*, *468*(6), 1029–1040. https://doi.org/10.1007/s00424-016-1819-4

Bedell, V. M., Wang, Y., Campbell, J. M., Poshusta, T. L., Starker, C. G., Krug, R. G., Tan, W., Penheiter, S. G., Ma, A. C., Leung, A. Y. H., Fahrenkrug, S. C., Carlson, D. F., Voytas, D. F., Clark, K. J., Essner, J. J., & Ekker, S. C. (2012). In vivo genome editing using a high-efficiency TALEN system. *Nature*, *491*(7422), 114–118. https://doi.org/10.1038/nature11537

Bedi, R. K., Huang, D., Wiedmer, L., Li, Y., Dolbois, A., Wojdyla, J. A., Sharpe, M. E., Caflisch, A., & Sledz, P. (2020). Selectively Disrupting m6A-Dependent Protein-RNA Interactions with Fragments. *ACS Chemical Biology*, *15*(3), 618–625. https://doi.org/10.1021/acschembio.9b00894

Berdasco, M., & Esteller, M. (2021). Towards a druggable epitranscriptome: Compounds that target RNA modifications in cancer. *British Journal of Pharmacology*, *179*(12), 2868–2889. https://doi.org/10.1111/bph.15604

Bevilacqua, P. C., & Blose, J. M. (2008). Structures, kinetics, thermodynamics, and biological functions of RNA hairpins. *Annual Review of Physical Chemistry*, *59*, 79–103. https://doi.org/10.1146/annurev.physchem.59.032607.093743

Birling, M. C., Yoshiki, A., Adams, D. J., Ayabe, S., Beaudet, A. L., Bottomley, J., Bradley, A., Brown, S. D. M., Bürger, A., Bushell, W., Chiani, F., Chin, H. J. G., Christou, S., Codner, G. F., DeMayo, F. J., Dickinson, M. E., Doe, B., Donahue, L. R., Fray, M. D., … Murray, S. A. (2021). A resource of targeted mutant mouse lines for 5,061 genes. *Nature Genetics*, *53*(4), 416–419. https://doi.org/10.1038/s41588-021-00825-y

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

Blobel, G. (1973). A Protein of Molecular Weight 78,000 Bound to the Polyadenylate Region of Eukaryotic Messenger RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(3), 924–928. https://doi.org/10.1073/pnas.70.3.924

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

Bordeleau, M. E., Cencic, R., Lindqvist, L., Oberer, M., Northcote, P., Wagner, G., & Pelletier, J. (2006). RNA-Mediated Sequestration of the RNA Helicase eIF4A by Pateamine A Inhibits Translation Initiation. *Chemistry and Biology*, *13*(12), 1287–1295. https://doi.org/10.1016/j.chembiol.2006.10.005

Bordeleau, M.-E., Matthews, J., Wojnar, J. M., Lindqvist, L., Novac, O., Jankowsky, E., Sonenberg, N., Northcote, P., Teesdale-Spittle, P., & Pelletier, J. (2005). Stimulation of mammalian translation initiation factor eIF4A activity by a small molecule inhibitor of eukaryotic translation. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(30), 10460–10465. https://doi.org/10.1073/pnas.0504249102

Boriack-Sjodin, P. A., Ribich, S., & Copeland, R. A. (2018). RNA-modifying proteins as anticancer drug targets. *Nature Reviews Drug Discovery*, *17*(6), 435–453. https://doi.org/10.1038/nrd.2018.71

Brown, G. J., Cañete, P. F., Wang, H., Medhavy, A., Bones, J., Roco, J. A., He, Y., Qin, Y., Cappello, J., Ellyard, J. I., Bassett, K., Shen, Q., Burgio, G., Zhang, Y., Turnbull, C., Meng, X., Wu, P., Cho, E., Miosge, L. A., … Vinuesa, C. G. (2022). TLR7 gain-of-function genetic variation causes human lupus. *Nature*, *605*(7909), 349–356. https://doi.org/10.1038/s41586-022-04642-z

Brown, R. H., & Al-Chalabi, A. (2017). Amyotrophic Lateral Sclerosis. *New England Journal of Medicine*, *377*(2), 162–172. https://doi.org/10.1056/NEJMra1603471

Brunton, H., Caligiuri, G., Cunningham, R., Upstill-Goddard, R., Bailey, U. M., Garner, I. M., Nourse, C., Dreyer, S., Jones, M., Moran-Jones, K., Wright, D. W., Paulus-Hock, V., Nixon, C., Thomson, G., Jamieson, N. B., McGregor, G. A., Evers, L., McKay, C. J., Gulati, A., … Bailey, P. J. (2020). HNF4A and GATA6 Loss Reveals Therapeutically Actionable Subtypes in Pancreatic Cancer. *Cell Reports*, *31*(6), 107625. https://doi.org/10.1016/j.celrep.2020.107625

Burd, C. G., & Dreyfuss, G. (1994). Conserved Structures and Diversity of Functions of RNA-Binding Proteins. *Science*, *265*(5172), 615–621. https://doi.org/10.1126/science.8036511.

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

Calo, E., Flynn, R. A., Martin, L., Spitale, R. C., Chang, H. Y., & Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature*, *518*(7538), 249–253. https://doi.org/10.1038/nature13923

Calo, E., Gu, B., Bowen, M. E., Aryan, F., Zalc, A., Liang, J., Flynn, R. A., Swigut, T., Chang, H. Y., Attardi, L. D., & Wysocka, J. (2018). Tissue-selective effects of nucleolar stress and rDNA damage in developmental disorders. *Nature*, *554*(7690), 112–117. https://doi.org/10.1038/nature25449

Carlevaro-Fita, J., Polidori, T., Das, M., Navarro, C., Zoller, T. I., & Johnson, R. (2018). Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. *Genome Research*, *29*(2), 208–222. https://doi.org/10.1101/gr.229922.117

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A. M., Taylor, M. S., Engström, P. G., Frith, M. C., Forrest, A. R. R., Alkema, W. B., Tan, S. L., Plessy, C., Kodzius, R., Ravasi, T.,

Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., … Hayashizaki, Y. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*, *38*(6), 626–635. https://doi.org/10.1038/ng1789

Caruthers, J. M., Johnson, E. R., & Mckay, D. B. (2000). Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(24), 13080–13085. https://doi.org/10.1073/pnas.97.24.13080

Casar, B., Badrock, A. P., Jiménez, I., Arozarena, I., Colón-Bolea, P., Lorenzo-Martín, L. F., Barinaga-Rementería, I., Barriuso, J., Cappitelli, V., Donoghue, D. J., Bustelo, X. R., Hurlstone, A., & Crespo, P. (2018). RAS at the Golgi antagonizes malignant transformation through PTPRϰ-mediated inhibition of ERK activation. *Nature Communications*, *9*(1), 3595. https://doi.org/10.1038/s41467-018-05941-8

Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A. M., Foehr, S., Curk, T., Krijgsveld, J., & Hentze, M. W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular Cell*, *63*(4), 696–710. https://doi.org/10.1016/j.molcel.2016.06.029

Chen, J. Y., Shen, Q. S., Zhou, W. Z., Peng, J., He, B. Z., Li, Y., Liu, C. J., Luan, X., Ding, W., Li, S., Chen, C., Tan, B. C. M., Zhang, Y. E., He, A., & Li, C. Y. (2015). Emergence, Retention and Selection: A Trilogy of Origination for Functional De Novo Proteins from Ancestral LncRNAs in Primates. *PLoS Genetics*, *11*(7), e1005391. https://doi.org/10.1371/journal.pgen.1005391

Chen, Y. G., Satpathy, A. T., & Chang, H. Y. (2017). Gene regulation in the immune system by long noncoding RNAs. *Nature Immunology*, *18*(9), 962–972. https://doi.org/10.1038/ni.3771

Cheng, K., Wang, X., & Yin, H. (2011). Small-molecule inhibitors of the TLR3/dsRNA complex. *Journal of the American Chemical Society*, *133*(11), 3764–3767. https://doi.org/10.1021/ja111312h

Cheng, Y., Luo, H., Izzo, F., Pickering, B. F., Nguyen, D., Myers, R., Schurer, A., Gourkanti, S., Brüning, J. C., Vu, L. P., Jaffrey, S. R., Landau, D. A., & Kharas, M. G. (2019). m6A RNA Methylation Maintains Hematopoietic Stem Cell Identity and Symmetric Commitment. *Cell Reports*, *28*(7), 1703–1716. https://doi.org/10.1016/j.celrep.2019.07.032

Cheng, Y., Xie, W., Pickering, B. F., Chu, K. L., Savino, A. M., Yang, X., Luo, H., Nguyen, D. T., Mo, S., Barin, E., Velleca, A., Rohwetter, T. M., Patel, D. J., Jaffrey, S. R., & Kharas, M. G. (2021). N6-Methyladenosine on mRNA facilitates a phase-separated nuclear body that suppresses myeloid leukemic differentiation. *Cancer Cell*, *39*(7), 958–972. https://doi.org/10.1016/j.ccell.2021.04.017

Cheng, Y.-C., & Prusoff, W. H. (1973). Relationship between the inhibition constant (Ki) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochemical Pharmacology*, *22*(23), 3099–3108. https://doi.org/10.1016/0006-2952(73)90196-2.

Childs-Disney, J. L., & Disney, M. D. (2016). Approaches to Validate and Manipulate RNA Targets with Small Molecules in Cells. *Annual Review of Pharmacology and Toxicology*, *56*, 123–140. https://doi.org/10.1146/annurev-pharmtox-010715-103910

Cho, S. S. L., Han, J., James, S. J., Png, C. W., Weerasooriya, M., Alonso, S., & Zhang, Y. (2017). Dual-specificity phosphatase 12 targets p38 MAP kinase to regulate macrophage response to intracellular bacterial infection. *Frontiers in Immunology*, *8*, 1259. https://doi.org/10.3389/fimmu.2017.01259

Choder, M. (2011). mRNA imprinting. Additional level in the regulation of gene expression. *Cellular Logistics*, *1*(1), 37–40. https://doi.org/10.4161/cl.1.1

Chu, C., Spitale, R. C., & Chang, H. Y. (2015). Technologies to probe functions and mechanisms of long noncoding RNAs. *Nature Structural and Molecular Biology*, *22*(1), 29–35. https://doi.org/10.1038/nsmb.2921

Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., Magnuson, T., Heard, E., & Chang, H. Y. (2015). Systematic discovery of Xist RNA binding proteins. *Cell*, *161*(2), 404–416. https://doi.org/10.1016/j.cell.2015.03.025

Cid-Samper, F., Gelabert-Baldrich, M., Lang, B., Lorenzo-Gotor, N., Delli Ponti, R., Severijnen, L. A. W. F. M., Bolognesi, B., Gelpi, E., Hukema, R. K., Botta-Orfila, T., & Tartaglia, G. G. (2018). An Integrative Study of Protein-RNA Condensates Identifies Scaffolding RNAs and Reveals Players in Fragile X-Associated Tremor/Ataxia Syndrome. *Cell Reports*, *25*(12), 3422–3434. https://doi.org/10.1016/j.celrep.2018.11.076

Cifuentes-Rojas, C., Hernandez, A. J., Sarma, K., & Lee, J. T. (2014). Regulatory Interactions between RNA and Polycomb Repressive Complex 2. *Molecular Cell*, *55*(2), 171–185. https://doi.org/10.1016/j.molcel.2014.05.009

Clark, M. B., Johnston, R. L., Inostroza-Ponta, M., Fox, A. H., Fortini, E., Moscato, P., Dinger, M. E., & Mattick, J. S. (2012). Genome-wide analysis of long noncoding RNA stability. *Genome Research*, *22*(5), 885–898. https://doi.org/10.1101/gr.131037.111

Cléry, A., & Allain, F. H.-T. (2012). *From Structure to Function of RNA Binding Domains* (Z. J. Lorković, Ed.; RNA Binding Proteins). Landes Bioscience. https://doi.org/10.1201/9781498713368

Clingman, C. C., Deveau, L. M., Hay, S. A., Genga, R. M., Shandilya, S. M. D., Massi, F., & Ryder, S. P. (2014). Allosteric inhibition of a stem cell RNA-binding protein by an intermediary metabolite. *ELife*, *3*, e02848. https://doi.org/10.7554/eLife.02848

Coe, E. A., Tan, J. Y., Shapiro, M., Louphrasitthiphol, P., Bassett, A. R., Marques, A. C., Goding, C. R., & Vance, K. W. (2019). The MITF-SOX10 regulated long non-coding RNA DIRC3 is a melanoma tumour suppressor. *PLoS Genetics*, *15*(12), e1008501. https://doi.org/10.1371/journal.pgen.1008501

Coelho, T., Adams, D., Silva, A., Lozeron, P., Hawkins, P. N., Mant, T., Perez, J., Chiesa, J., Warrington, S., Tranter, E., Munisamy, M., Falzone, R., Harrop, J., Cehelsky, J., Bettencourt, B. R., Geissler, M., Butler, J. S., Sehgal, A., Meyers, R. E., … Suhr, O. B. (2013). Safety and Efficacy of RNAi Therapy for Transthyretin Amyloidosis. *New England Journal of Medicine*, *369*(9), 819–829. https://doi.org/10.1056/nejmoa1208760

Conrad, T., Albrecht, A. S., de Melo Costa, V. R., Sauer, S., Meierhofer, D., & Ørom, U. A. (2016). Serial interactome capture of the human cell nucleus. *Nature Communications*, *7*(1), 11212. https://doi.org/10.1038/ncomms11212

Constanty, F., & Shkumatava, A. (2021). lncRNAs in development and differentiation: From sequence motifs to functional characterization. *Development (Cambridge)*, *148*(1), dev182741. https://doi.org/10.1242/dev.182741

Cook, K. B., Kazan, H., Zuberi, K., Morris, Q., & Hughes, T. R. (2011). RBPDB: A database of RNA-binding specificities. *Nucleic Acids Research*, *39*(Database issue), D301–D308. https://doi.org/10.1093/nar/gkq1069

Corley, M., Burns, M. C., & Yeo, G. W. (2020). How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Molecular Cell*, *78*(1), 9–29. https://doi.org/10.1016/j.molcel.2020.03.011

Cornelis, G., Souquere, S., Vernochet, C., Heidmann, T., & Pierron, G. (2016). Functional conservation of the lncRNA NEAT1 in the ancestrally diverged marsupial lineage: Evidence for NEAT1 expression and associated paraspeckle assembly during late gestation in the opossum Monodelphis domestica. *RNA Biology*, *13*(9), 826–836. https://doi.org/10.1080/15476286.2016.1197482

Cortez Vollmers, A., Covarrubias, S., Kuang, D., Shulkin, A., Iwuagwu, J., Katzman, S., Viswanathan, K., Vollmers, C., Wakeland, E., & Carpenter, S. (2021). A conserved long noncoding RNA, GAPLINC, modulates the immune response during endotoxic shock. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(7). https://doi.org/10.1073/pnas.2016648118

Cox, D. B. T., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., & Zhang, F. (2017). RNA editing with CRISPR-Cas13. *Science*, *358*(6366), 1019–1027.

Crabbe, J. C., Wahlsten, D., & Dudek, B. (1999). Genetics of Mouse Behavior: Interactions with Laboratory Environment. *Science*, *284*(5420), 1670–1672. https://doi.org/10.1126/science.284.5420.1670

Cully, M. (2019). Chemical inhibitors make their RNA epigenetic mark. In *Nature Reviews Drug discovery* (Vol. 18, Issue 12, pp. 892–894). NLM (Medline). https://doi.org/10.1038/d41573-019-00179-5

D'Agostino, V. G., Sighel, D., Zucal, C., Bonomo, I., Micaelli, M., Lolli, G., Provenzani, A., Quattrone, A., & Adami, V. (2019). Screening Approaches for Targeting Ribonucleoprotein Complexes: A New Dimension for Drug Discovery. *SLAS Discovery*, *24*(3), 314–331. https://doi.org/10.1177/2472555218818065

Dang, C. v., Reddy, E. P., Shokat, K. M., & Soucek, L. (2017). Drugging the "undruggable" cancer targets. *Nature Reviews Cancer*, *17*(8), 502–508. https://doi.org/10.1038/nrc.2017.36

Dankort, D., Curley, D. P., Cartlidge, R. A., Nelson, B., Karnezis, A. N., Damsky, W. E., You, M. J., DePinho, R. A., McMahon, M., & Bosenberg, M. (2009). BrafV600E cooperates with Pten loss to induce metastatic melanoma. *Nature Genetics*, *41*(5), 544–552. https://doi.org/10.1038/ng.356

Davidovich, C., & Cech, T. R. (2015). The recruitment of chromatin modifiers by long noncoding RNAs: Lessons from PRC2. In *RNA* (Vol. 21, Issue 12, pp. 2007–2022). Cold Spring Harbor Laboratory Press. https://doi.org/10.1261/rna.053918.115

Davidson, A., Leeper, T. C., Athanassiou, Z., Patora-Komisarska, K., Karn, J., Robinson, J. A., & Varani, G. (2009). Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of Tat protein. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(29), 11931–11936. https://doi.org/10.1073/pnas.0900629106

Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M. J., Bottomley, W., Davis, N., Dicks, E., Ewing, R., Floyd, Y., Gray, K., Hall, S., Hawes, R., Hughes, J., Kosmidou, V., … Futreal, & P. A. (2002). Mutations of the BRAF gene in human cancer. *Nature*, *417*(6892), 949–954. https://doi.org/10.1038/nature00766

Davis, M. P., Carrieri, C., Saini, H. K., Dongen, S., Leonardi, T., Bussotti, G., Monahan, J. M., Auchynnikava, T., Bitetti, A., Rappsilber, J., Allshire, R. C., Shkumatava, A., O'Carroll, D., & Enright, A. J. (2017). Transposon-driven transcription is a conserved feature of vertebrate spermatogenesis and transcript evolution. *EMBO Reports*, *18*(7), 1231–1247. https://doi.org/10.15252/embr.201744059

Delaunay, S., & Frye, M. (2019). RNA modifications regulating cell fate in cancer. *Nature Cell Biology*, *21*(5), 552–559. https://doi.org/10.1038/s41556-019-0319-0

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

Desrosiers, R., Friderici, K., & Rottman, F. (1974). Identification of Methylated Nucleosides in Messenger RNA from Novikoff Hepatoma Cells. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(10), 3971–3975. https://doi.org/10.1073/pnas.71.10.3971

Dickinson, M. E., Flenniken, A. M., Ji, X., Teboul, L., Wong, M. D., White, J. K., Meehan, T. F., Weninger, W. J., Westerberg, H., Adissu, H., Baker, C. N., Bower, L., Brown, J. M., Brianna Caddle, L., Chiani, F., Clary, D., Cleak, J., Daly, M. J., Denegre, J. M., … Murakami, A. (2016). High-throughput discovery of novel developmental phenotypes. *Nature*, *537*(7621), 508–514. https://doi.org/10.1038/nature19356

Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., Skinner, M. A., & Valeriot, R. (1989). Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(18), 6925–6929. https://doi.org/10.1073/pnas.86.18.6925

Disney, M. D. (2019). Targeting RNA with Small Molecules To Capture Opportunities at the Intersection of Chemistry, Biology, and Medicine. *Journal of the American Chemical Society*, *141*(17), 6776–6790. https://doi.org/10.1021/jacs.8b13419

Dixon-McDougall, T., & Brown, C. J. (2021). Independent domains for recruitment of PRC1 and PRC2 by human XIST. *PLoS Genetics*, *17*(3), e1009123. https://doi.org/10.1371/journal.pgen.1009123

Döhner, H., Estey, E., Grimwade, D., Amadori, S., Appelbaum, F. R., Dombret, H., Ebert, B. L., Fenaux, P., Larson, R. A., Levine, R. L., Lo-Coco, F., Naoe, T., Niederwieser, D., Ossenkoppele, G. J., Sanz, M., Sierra, J., Tallman, M. S., Tien, H.-F., Wei, A. H., & Bloomfield, C. D. (2017). Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*, *129*(4), 424–447. https://doi.org/10.1182/blood-2016-08

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

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., Sorek, R., & Rechavi, G. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, *485*(7397), 201–206. https://doi.org/10.1038/nature11112

Donlic, A., Morgan, B. S., Xu, J. L., Liu, A., Roble, C., & Hargrove, A. E. (2018). Discovery of Small Molecule Ligands for MALAT1 by Tuning an RNA-Binding Scaffold. *Angewandte Chemie - International Edition*, *57*(40), 13242–13247. https://doi.org/10.1002/anie.201808823

Duret, L., Chureau, C., Samain, S., Weissenbach, J., & PAvner, P. (2006). The Xist RNA Gene Evolved in Eutherians by Pseudogenization of a Protein-Coding Gene. *Science*, *312*(5780), 1653–1655. https://doi.org/10.1126/science.1126863

Duszczyk, M. M., Zanier, K., & Sattler, M. (2008). A NMR strategy to unambiguously distinguish nucleic acid hairpin and duplex conformations applied to a Xist RNA A-repeat. *Nucleic Acids Research*, *36*(22), 7068–7077. https://doi.org/10.1093/nar/gkn776

Eißmann, M., Gutschner, T., Hämmerle, M., Günther, S., Caudron-Herger, M., Groß, M., Schirmacher, P., Rippe, K., Braun, T., Zörnig, M., & Diederichs, S. (2012). Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biology*, *9*(8), 1076–1087. https://doi.org/10.4161/rna.21089

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

Elcheva, I. A., Wood, T., Chiarolanzio, K., Chim, B., Wong, M., Singh, V., Gowda, C. P., Lu, Q., Hafner, M., Dovat, S., Liu, Z., Muljo, S. A., & Spiegelman, V. S. (2020). RNA-binding protein IGF2BP1 maintains leukemia stem cell properties by regulating HOXB4, MYB, and ALDH1A1. *Leukemia*, *34*(5), 1354–1363. https://doi.org/10.1038/s41375-019-0656-9

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

Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., McDonel, P. E., Guttman, M., & Lander, E. S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, *539*(7629), 452–455. https://doi.org/10.1038/nature20149

Engreitz, J. M., Ollikainen, N., & Guttman, M. (2016). Long non-coding RNAs: Spatial amplifiers that control nuclear structure and gene expression. *Nature Reviews Molecular Cell Biology*, *17*(12), 756–770. https://doi.org/10.1038/nrm.2016.126

Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W., & Jhoti, H. (2016). Twenty years on: The impact of fragments on drug discovery. *Nature Reviews Drug Discovery*, *15*(9), 605–619. https://doi.org/10.1038/nrd.2016.109

Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjostedt, E., Lundberg, E., Szigyarto, C. A. K., Skogs, M., Ottosson Takanen, J., Berling, H., Tegel, H., Mulder, J., … Uhlen, M. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular and Cellular Proteomics*, *13*(2), 397–406. https://doi.org/10.1074/mcp.M113.035600

Falcão Monteiro, L., & Forti, F. L. (2019). Network analysis of DUSP12 partners in the nucleus under genotoxic stress. *Journal of Proteomics*, *197*, 42–52. https://doi.org/10.1016/j.jprot.2019.02.008

Fang, H., Bonora, G., Lewandowski, J. P., Thakur, J., Filippova, G. N., Henikoff, S., Shendure, J., Duan, Z., Rinn, J. L., Deng, X., Noble, W. S., & Disteche, C. M. (2020). Trans- and cis-acting effects of Firre on epigenetic features of the inactive X chromosome. *Nature Communications*, *11*(1), 6053. https://doi.org/10.1038/s41467-020-19879-3

Farnsworth, D. R., Saunders, L. M., & Miller, A. C. (2020). A single-cell transcriptome atlas for zebrafish development. *Developmental Biology*, *459*(2), 100–108. https://doi.org/10.1016/j.ydbio.2019.11.008

Fatica, A., & Bozzoni, I. (2014). Long non-coding RNAs: New players in cell differentiation and development. *Nature Reviews Genetics*, *15*(1), 7–21. https://doi.org/10.1038/nrg3606

Feretzaki, M., Pospisilova, M., Valador Fernandes, R., Lunardi, T., Krejci, L., & Lingner, J. (2020). RAD51-dependent recruitment of TERRA lncRNA to telomeres through R-loops. *Nature*, *587*(7833), 303–308. https://doi.org/10.1038/s41586-020-2815-6

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., Aros, C. J., Waters, E. v., Casero, D., Basso, G., Pigazzi, M., & Rao, D. S. (2017). The lncRNA CASC15 regulates SOX4 expression in RUNX1-rearranged acute leukemia. *Molecular Cancer*, *16*(1). https://doi.org/10.1186/s12943-017-0692-x

Feuerhahn, S., Iglesias, N., Panza, A., Porro, A., & Lingner, J. (2010). TERRA biogenesis, turnover and implications for function. *FEBS Letters*, *584*(17), 3812–3818. https://doi.org/10.1016/j.febslet.2010.07.032

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specificgenetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, *391*, 806–811. https://doi.org/10.1038/35888

Fitzgerald, K. A., & Kagan, J. C. (2020). Toll-like Receptors and the Control of Immunity. *Cell*, *180*(6), 1044–1066. https://doi.org/10.1016/j.cell.2020.02.041

Fitzpatrick, G. v., Soloway, P. D., & Higgins, M. J. (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nature Genetics*, *32*(3), 426–431. https://doi.org/10.1038/ng988

Flamholz, A., Phillips, R., & Milo, R. (2014). The quantified cell. *Molecular Biology of the Cell*, *25*(22), 3497–3500. https://doi.org/10.1091/mbc.E14-09-1347

Flynn, R. A., Belk, J. A., Qi, Y., Yasumoto, Y., Wei, J., Alfajaro, M. M., Shi, Q., Mumbach, M. R., Limaye, A., DeWeirdt, P. C., Schmitz, C. O., Parker, K. R., Woo, E., Chang, H. Y., Horvath, T. L., Carette, J. E., Bertozzi, C. R., Wilen, C. B., & Satpathy, A. T. (2021). Discovery and functional interrogation of SARS-CoV-2 RNA-host protein interactions. *Cell*, *184*(9), 2394–2411. https://doi.org/10.1016/j.cell.2021.03.012

Fontanals-Cirera, B., Hasson, D., Vardabasso, C., di Micco, R., Agrawal, P., Chowdhury, A., Gantz, M., de Pablos-Aragoneses, A., Morgenstern, A., Wu, P., Filipescu, D., Valle-Garcia, D., Darvishian, F., Roe, J. S., Davies, M. A., Vakoc, C. R., Hernando, E., & Bernstein, E. (2017). Harnessing BET Inhibitor Sensitivity Reveals AMIGO2 as a Melanoma Survival Gene. *Molecular Cell*, *68*(4), 731-744.e9. https://doi.org/10.1016/j.molcel.2017.11.004

Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C. A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A., Noro, Y., Wong, C. H., de Hoon, M., Andersson, R., Sandelin, A., Suzuki, H., Wei, C. L., Koseki, H., Hasegawa, Y., … Carninci, P. (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nature Genetics*, *46*(6), 558–566. https://doi.org/10.1038/ng.2965

Frankish, A., Diekhans, M., Ferreira, A. M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., di Domenico, T., Donaldson, S., Fiddes, I. T., … Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research*, *47*(D1), D766–D773. https://doi.org/10.1093/nar/gky955

Frantz, W. T., & Ceol, C. J. (2020). From Tank to Treatment: Modeling Melanoma in Zebrafish. *Cells*, *9*(5), 1289. https://doi.org/10.3390/cells9051289

Frisca, F., Colquhoun, D., Goldshmit, Y., Änkö, M. L., Pébay, A., & Kaslin, J. (2016). Role of ectonucleotide pyrophosphatase/phosphodiesterase 2 in the midline axis formation of zebrafish. *Scientific Reports*, *6*. https://doi.org/10.1038/srep37678

Fu, Y., Dominissini, D., Rechavi, G., & He, C. (2014). Gene expression regulation mediated through reversible m 6 A RNA methylation. In *Nature Reviews Genetics* (Vol. 15, Issue 5, pp. 293–306). Nature Publishing Group. https://doi.org/10.1038/nrg3724

Fueyo, R., Judd, J., Feschotte, C., & Wysocka, J. (2022). Roles of transposable elements in the regulation of mammalian transcription. *Nature Reviews Molecular Cell Biology*, *23*(7), 481–497. https://doi.org/10.1038/s41580-022-00457-y

Gabut, M., Chaudhry, S., & Blencowe, B. J. (2008). SnapShot: The Splicing Regulatory Machinery. *Cell*, *133*(1), 192. https://doi.org/10.1016/j.cell.2008.03.010

Gallagher, S. J., Ashby, W. J., Cordeliéres, F. P., & Larue, L. (2019). *Image J Documentation Wiki*. https://imagejdocu.list.lu/plugin/analysis/scratch_wound_assay_automatic_analysis_macro/start

Ganesh, S., & Svoboda, P. (2016). Retrotransposon-associated long non-coding RNAs in mice and men. *European Journal of Physiology*, *468*(6), 1049–1060. https://doi.org/10.1007/s00424-016-1818-5

Ganser, L. R., Kelly, M. L., Herschlag, D., & Al-Hashimi, H. M. (2019). The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology*, *20*(8), 474–489. https://doi.org/10.1038/s41580-019-0136-0

Gashaw, I., Ellinghaus, P., Sommer, A., & Asadullah, K. (2011). What makes a good drug target? *Drug Discovery Today*, *16*(23–24), 1037–1043. https://doi.org/10.1016/j.drudis.2011.09.007

Gebauer, F., Schwarzl, T., Valcárcel, J., & Hentze, M. W. (2021). RNA-binding proteins in human genetic disease. *Nature Reviews Genetics*, *22*(3), 185–198. https://doi.org/10.1038/s41576-020-00302-y

Geisler, S., & Coller, J. (2013). RNA in unexpected places: Long non-coding RNA functions in diverse cellular contexts. *Nature Reviews Molecular Cell Biology*, *14*(11), 699–712. https://doi.org/10.1038/nrm3679

Geng, Q., Xhabija, B., Knuckle, C., Bonham, C. A., & Vacratsis, P. O. (2017). The Atypical Dual Specificity Phosphatase hYVH1 Associates with Multiple Ribonucleoprotein Particles. *Journal of Biological Chemistry*, *292*(2), 539–550. https://doi.org/10.1074/jbc.M116.715607

Gerstberger, S., Hafner, M., & Tuschl, T. (2014). A census of human RNA-binding proteins. *Nature Reviews Genetics*, *15*(12), 829–845. https://doi.org/10.1038/nrg3813

Gewirtz, D. A. (1999). A Critical Evaluation of the Mechanisms of Action Proposed for the Antitumor Effects of the Anthracycline Antibiotics Adriamycin and Daunorubicin. *Biochemical Pharmacology*, *57*, 727–741. https://doi.org/10.1016/s0006-2952(98)00307-4

Gil, N., & Ulitsky, I. (2020). Regulation of gene expression by cis-acting long non-coding RNAs. *Nature Reviews Genetics*, *21*(2), 102–117. https://doi.org/10.1038/s41576-019-0184-5

Goudarzi, M., Berg, K., Pieper, L. M., & Schier, A. F. (2019). Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility. *ELife*, *8*, e40815. https://doi.org/10.7554/eLife.40815.001

Graindorge, A., Pinheiro, I., Nawrocka, A., Mallory, A. C., Tsvetkov, P., Gil, N., Carolis, C., Buchholz, F., Ulitsky, I., Heard, E., Taipale, M., & Shkumatava, A. (2019). In-cell identification and measurement of RNA-protein interactions. *Nature Communications*, *10*(1), 5317. https://doi.org/10.1038/s41467-019-13235-w

Grant, J., Mahadevaiah, S. K., Khil, P., Sangrithi, M. N., Royo, H., Duckworth, J., McCarrey, J. R., Vandeberg, J. L., Renfree, M. B., Taylor, W., Elgar, G., Camerini-Otero, R. D., Gilchrist, M. J., & Turner, J. M. A. (2012). Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature*, *487*(7406), 254–258. https://doi.org/10.1038/nature11171

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

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman, S. (1963). The RNA Moiety of Ribonuclease P Is the Catalytic Subunit of the Enzyme. *Cell*, *35*(2), 649–857. https://doi.org/10.1016/0092-8674(83)90117-4

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

Gutschner, T., Hämmerle, M., Eißmann, M., Hsu, J., Kim, Y., Hung, G., Revenko, A., Arun, G., Stentrup, M., Groß, M., Zörnig, M., MacLeod, A. R., Spector, D. L., & Diederichs, S. (2013). The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Research*, *73*(3), 1180–1189. https://doi.org/10.1158/0008-5472.CAN-12-2850

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

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

Hacisuleyman, E., Shukla, C. J., Weiner, C. L., & Rinn, J. L. (2016). Function and evolution of local repeats in the Firre locus. *Nature Communications*, *7*(1), 11021. https://doi.org/10.1038/ncomms11021

Hainzl, T., Huang, S., & Sauer-Eriksson, A. E. (2005). Structural insights into SRP RNA: An induced fit mechanism for SRP assembly. *RNA*, *11*(7), 1043–1050. https://doi.org/10.1261/rna.2080205

Hajduk, P. J., Huth, J. R., & Tse, C. (2005). Predicting protein druggability. *Drug Discovery Today*, *10*(23–24), 1675–1682. https://doi.org/10.1016/S1359-6446(05)03624-X

Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R., & Hannon, G. J. (2001). Argonaute2, a Link Between Genetic and Biochemical Analyses of RNAi. *Science*, *293*(5532), 1146–1150. https://doi.org/10.1126/science.1064023

Han, X., Luo, S., Peng, G., Lu, J. Y., Cui, G., Liu, L., Yan, P., Yin, Y., Liu, W., Wang, R., Zhang, J., Ai, S., Chang, Z., Na, J., He, A., Jing, N., & Shen, X. (2018). Mouse knockout models reveal largely dispensable but context-dependent functions of lncRNAs during development. *Journal of Molecular Cell Biology*, *10*(2), 175–178. https://doi.org/10.1093/jmcb/mjy003

Han, X., Zhang, J., Liu, Y., Fan, X., Ai, S., Luo, Y., Li, X., Jin, H., Luo, S., Zheng, H., Yue, Y., Chang, Z., Yang, Z., Tang, F., He, A., & Shen, X. (2019). The lncRNA Hand2os1/Uph locus orchestrates heart development through regulation of precise expression of Hand2. *Development (Cambridge)*, *146*(13), dev176198. https://doi.org/10.1242/dev.176198

Haussmann, I. U., Bodi, Z., Sanchez-Moran, E., Mongan, N. P., Archer, N., Fray, R. G., & Soller, M. (2016). M6A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature*, *540*(7632), 301–304. https://doi.org/10.1038/nature20577

Hawkes, E. J., Hennelly, S. P., Novikova, I. v., Irwin, J. A., Dean, C., & Sanbonmatsu, K. Y. (2016). COOLAIR Antisense RNAs Form Evolutionarily Conserved Elaborate Secondary Structures. *Cell Reports*, *16*(12), 3087–3096. https://doi.org/10.1016/j.celrep.2016.08.045

He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D. C., Wilusz, J. E., Garcia, B. A., & Bonasio, R. (2016). High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Molecular Cell*, *64*(2), 416–430. https://doi.org/10.1016/j.molcel.2016.09.034

Heard, E., Mongelard, F., Arnaud, D., Chureau, C., Vourc'h, C., & Avner, P. (1999). Human XIST yeast artificial chromosome transgenes show partial X inactivation center function in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(12), 6841–6846. https://doi.org/10.1073/pnas.96.12.6841

Heinen, T. J. A. J., Staubach, F., Häming, D., & Tautz, D. (2009). Emergence of a New Gene from an Intergenic Region. *Current Biology*, *19*(18), 1527–1531. https://doi.org/10.1016/j.cub.2009.07.049

Henley, M. J., & Koehler, A. N. (2021). Advances in targeting 'undruggable' transcription factors with small molecules. *Nature Reviews Drug Discovery*, *20*(9), 669–688. https://doi.org/10.1038/s41573-021-00199-0

Hentze, M. W., Castello, A., Schwarzl, T., & Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology*, *19*(5), 327–341. https://doi.org/10.1038/nrm.2017.130

Hezroni, H., Ben-Tov Perry, R., Meir, Z., Housman, G., Lubelsky, Y., & Ulitsky, I. (2017). A subset of conserved mammalian long non-coding RNAs are fossils of ancestral protein-coding genes. *Genome Biology*, *18*(1). https://doi.org/10.1186/s13059-017-1293-0

Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Ulitsky, I. (2015). Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Reports*, *11*(7), 1110–1122. https://doi.org/10.1016/j.celrep.2015.04.023

Hezroni, H., Perry, R. B. T., & Ulitsky, I. (2019). Long noncoding RNAs in development and regeneration of the neural lineage. *Cold Spring Harbor Symposia on Quantitative Biology*, *84*, 165–177. https://doi.org/10.1101/sqb.2019.84.039347

Holmström, T. H., Mialon, A., Kallio, M., Nymalm, Y., Mannermaa, L., Holm, T., Johansson, H., Black, E., Gillespie, D., Salminen, T. A., Langel, Ü., Valdez, B. C., & Westermarck, J. (2008). c-Jun supports ribosomal RNA processing and nucleolar localization of RNA helicase DDX21. *Journal of Biological Chemistry*, *283*(11), 7046–7053. https://doi.org/10.1074/jbc.M709613200

Hon, C. C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J. L., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T. M., Severin, J., Lizio, M., Kawaji, H., Kasukawa, T., Itoh, M., Burroughs, A. M., Noma, S., Djebali, S.,

Alam, T., Medvedeva, Y. A., … Forrest, A. R. R. (2017). An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature*, *543*(7644), 199–204. https://doi.org/10.1038/nature21374

Hongay, C. F., & Orr-Weaver, T. L. (2011). Drosophila inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(36), 14855–14860. https://doi.org/10.1073/pnas.1111577108

Hornbeck, P. v., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., & Skrzypek, E. (2015). PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Research*, *43*(D1), D512–D520. https://doi.org/10.1093/nar/gku1267

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., Chu, S. C., Saleh, S., Sankar, K., Su, F., Guo, S., Qiao, Y., Freier, S. M., Bui, H. H., Cao, X., … Chinnaiyan, A. M. (2017). Oncogenic Role of THOR, a Conserved Cancer/Testis Long Non-coding RNA. *Cell*, *171*(7), 1559–1572. https://doi.org/10.1016/j.cell.2017.11.040

Howe, J. A., Wang, H., Fischmann, T. O., Balibar, C. J., Xiao, L., Galgoci, A. M., Malinverni, J. C., Mayhood, T., Villafania, A., Nahvi, A., Murgolo, N., Barbieri, C. M., Mann, P. A., Carr, D., Xia, E., Zuck, P., Riley, D., Painter, R. E., Walker, S. S., … Roemer, T. (2015). Selective small-molecule inhibition of an RNA structural element. *Nature*, *526*(7575), 672–677. https://doi.org/10.1038/nature15542

Hua, Y., Sahashi, K., Hung, G., Rigo, F., Passini, M. A., Bennett, C. F., & Krainer, A. R. (2010). Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes and Development*, *24*(15), 1634–1644. https://doi.org/10.1101/gad.1941310

Hua, Y., Sahashi, K., Rigo, F., Hung, G., Horev, G., Bennett, C. F., & Krainer, A. R. (2011). Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model. *Nature*, *478*(7367), 123–126. https://doi.org/10.1038/nature10485

Huang, C. Y., & Tan, T. H. (2012). DUSPs, to MAP kinases and beyond. *Cell and Bioscience*, *2*(1), 24. https://doi.org/10.1186/2045-3701-2-24

Huang, J., & Yin, P. (2018). Structural Insights into N6-methyladenosine (m6A) Modification in the Transcriptome. *Genomics, Proteomics and Bioinformatics*, *16*(2), 85–98. https://doi.org/10.1016/j.gpb.2018.03.001

Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nature Medicine*, *21*(11), 1253–1261. https://doi.org/10.1038/nm.3981

Hudson, W. H., & Ortlund, E. A. (2014). The structure, function and evolution of proteins that bind DNA and RNA. *Nature Reviews Molecular Cell Biology*, *15*(11), 749–760. https://doi.org/10.1038/nrm3884

Hughes, J. P., Rees, S. S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, *162*(6), 1239–1249. https://doi.org/10.1111/j.1476-5381.2010.01127.x

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J. R. J., & Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology*, *31*(3), 227–229. https://doi.org/10.1038/nbt.2501

Hyeon Hong, S., Han, G., Jae Lee, S., Cocquet, J., & Cho, C. (2021). Testicular germ cell-specific lncRNA, Teshl, is required for complete expression of Y chromosome genes and a normal offspring sex ratio. *Science Advances*, *7*(9), eabg5177. https://doi.org/10.1126/sciadv.abg5177

Ilik, I. A., Quinn, J. J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R. C., Luscombe, N., Backofen, R., Chang, H. Y., & Akhtar, A. (2013). Tandem stem-loops in roX RNAs act together to mediate X Chromosome dosage compensation in Drosophila. *Molecular Cell*, *51*(2), 156–173. https://doi.org/10.1016/j.molcel.2013.07.001

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

Iuchi, S. (2001). Three classes of C2H2 zinc finger proteins. *Cellular and Molecular Life Sciences*, *58*(4), 625–635. https://doi.org/10.1007/PL00000885

Ivanova, I., Much, C., di Giacomo, M., Azzi, C., Morgan, M., Moreira, P. N., Monahan, J., Carrieri, C., Enright, A. J., & O'Carroll, D. (2017). The RNA m6A Reader YTHDF2 Is Essential for the Post-transcriptional Regulation of the Maternal Transcriptome and Oocyte Competence. *Molecular Cell*, *67*(6), 1059–1067. https://doi.org/10.1016/j.molcel.2017.08.003

Jaric, I., Voelkl, B., Clerc, M., Schmid, M. W., Novak, J., Rosso, M., Rufener, R., Tabea Von Kortzfleisch, V., Richter, S. H., Buettner, M., Bleich, A., Amrein, I., Wolfer, D. P., Touma, C., Sunagawa, S., & Würbel, H. (2022). Rearing environment persistently modulates the phenotype of mice. *Preprint at BioRxiv*. https://doi.org/10.1101/2022.02.11.480070

Järvelin, A. I., Noerenberg, M., Davis, I., & Castello, A. (2016). The new (dis)order in RNA regulation. *Cell Communication and Signaling*, *14*(1). https://doi.org/10.1186/s12964-016-0132-3

Jeffrey, K. L., Camps, M., Rommel, C., & Mackay, C. R. (2007). Targeting dual-specificity phosphatases: manipulating MAP kinase signalling and immune responses. *Nature Reviews Drug Discovery*, *6*(5), 391–403. https://doi.org/10.1038/nrd2289

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y. G., & He, C. (2011). N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chemical Biology*, *7*(12), 885–887. https://doi.org/10.1038/nchembio.687

Jiang, C., Li, Y., Zhao, Z., Lu, J., Chen, H., Ding, N., Wang, G., Xu, J., & Li, X. (2016). Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs. *Oncotarget*, *7*(6), 7120–7133. https://doi.org/10.18632/oncotarget.6859

Jin, F., Li, J., Zhang, Y. B., Liu, X., Cai, M., Liu, M., Li, M., Ma, C., Yue, R., Zhu, Y., Lai, R., Wang, Z., Ji, X., Wei, H., Dong, J., Liu, Z., Wang, Y., Sun, Y., & Wang, X. (2021). A functional motif of long noncoding RNA Nron against osteoporosis. *Nature Communications*, *12*(1), 3319. https://doi.org/10.1038/s41467-021-23642-7

Johansson, J. A., Marie, K. L., Lu, Y., Brombin, A., Santoriello, C., Zeng, Z., Zich, J., Gautier, P., von Kriegsheim, A., Brunsdon, H., Wheeler, A. P., Dreger, M., Houston, D. R., Dooley, C. M., Sims, A. H., Busch-Nentwich, E. M., Zon, L. I., Illingworth, R. S., & Patton, E. E. (2020). PRL3-DDX21 Transcriptional Control of Endolysosomal Genes Restricts Melanocyte Stem Cell Differentiation. *Developmental Cell*, *54*(3), 317–332. https://doi.org/10.1016/j.devcel.2020.06.013

Johnson, R., & Guigó, R. (2014). The RIDL hypothesis: Transposable elements as functional domains of long noncoding RNAs. *RNA*, *20*(7), 959–976. https://doi.org/10.1261/rna.044560.114

Julio, A. R., & Backus, K. M. (2021). New approaches to target RNA binding proteins. *Current Opinion in Chemical Biology*, *62*, 13–23. https://doi.org/10.1016/j.cbpa.2020.12.006

Juvvuna, P. K., Mondal, T., di Marco, M., Kosalai, S. T., Kanduri, M., & Kanduri, C. (2021). NBAT1/CASC15-003/USP36 control MYCN expression and its downstream pathway genes in neuroblastoma. *Neuro-Oncology Advances*, *3*(1), vdab056. https://doi.org/10.1093/noajnl/vdab056

Kappelmann-Fenzl, M., Gebhard, C., Matthies, A. O., Kuphal, S., Rehli, M., & Bosserhoff, A. K. (2019). C-Jun drives melanoma progression in PTEN wild type melanoma cells. *Cell Death and Disease*, *10*(8). https://doi.org/10.1038/s41419-019-1821-9

Kapusta, A., & Feschotte, C. (2014). Volatile evolution of long noncoding RNA repertoires: Mechanisms and biological implications. *Trends in Genetics*, *30*(10), 439–452. https://doi.org/10.1016/j.tig.2014.08.004

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

Karner, H., Webb, C. H., Carmona, S., Liu, Y., Lin, B., Erhard, M., Chan, D., Baldi, P., Spitale, R. C., & Sun, S. (2020). Functional Conservation of LncRNA JPX Despite Sequence and Structural Divergence. *Journal of Molecular Biology*, *432*(2), 283–300. https://doi.org/10.1016/j.jmb.2019.09.002

Kasowitz, S. D., Ma, J., Anderson, S. J., Leu, N. A., Xu, Y., Gregory, B. D., Schultz, R. M., & Wang, P. J. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. *PLoS Genetics*, *14*(5), e1007412. https://doi.org/10.1371/journal.pgen.1007412

Kastelic, N., & Landthaler, M. (2017). mRNA interactome capture in mammalian cells. *Methods*, *126*, 38–43. https://doi.org/10.1016/j.ymeth.2017.07.006

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., Ciarlo, C., White, R. M., Matos, D. A., Puller, A. C., Santoriello, C., Liao, E. C., Young, R. A., & Zon, L. I. (2016). A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. *Science*, *351*(6272). https://doi.org/10.1126/science.aad2197

Kelley, D., & Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biology*, *13*(11), R107. https://doi.org/10.1186/gb-2012-13-11-r107

Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B. E., van Oudenaarden, A., Regev, A., Lander, E. S., & Rinn, J. L. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(28), 11667–11672. https://doi.org/10.1073/pnas.0904715106

Kim, J., Hu, C., Moufawad El Achkar, C., Black, L. E., Douville, J., Larson, A., Pendergast, M. K., Goldkind, S. F., Lee, E. A., Kuniholm, A., Soucy, A., Vaze, J., Belur, N. R., Fredriksen, K., Stojkovska, I., Tsytsykova, A., Armant, M., DiDonato, R. L., Choi, J., … Yu, T. W. (2019). Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. *New England Journal of Medicine*, *381*(17), 1644–1652. https://doi.org/10.1056/nejmoa1813279

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

Kittler, R., Surendranath, V., Heninger, A. K., Slabicki, M., Theis, M., Putz, G., Franke, K., Caldarelli, A., Grabner, H., Kozak, K., Wagner, J., Rees, E., Korn, B., Frenzel, C., Sachse, C., Sönnichsen, B., Guo, J., Schelter, J., Burchard, J., … Buchholz, F. (2007). Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nature Methods*, *4*(4), 337–344. https://doi.org/10.1038/nmeth1025

Kjems, J., Brown, M., Changt, D. D., & Sharp, P. A. (1991). Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(3), 683–687. https://doi.org/10.1073/pnas.88.3.683

Klattenhoff, C. A., Scheuermann, J. C., Surface, L. E., Bradley, R. K., Fields, P. A., Steinhauser, M. L., Ding, H., Butty, V. L., Torrey, L., Haas, S., Abo, R., Tabebordbar, M., Lee, R. T., Burge, C. B., & Boyer, L. A. (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*, *152*(3), 570–583. https://doi.org/10.1016/j.cell.2013.01.003

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

Kopp, F., Elguindy, M. M., Yalvac, M. E., Zhang, H., Chen, B., Gillett, F. A., Lee, S., Sivakumar, S., Yu, H., Xie, Y., Mishra, P., Sahenk, Z., & Mendell, J. T. (2019). PUMILIO hyperactivity drives premature aging of Norad-deficient mice. *ELife*, *8*, e42650. https://doi.org/10.7554/eLife.42650.001

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M., & Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nature Chemical Biology*, *3*(9), 570–575. https://doi.org/10.1038/nchembio.2007.16

Kramer, E. T., Godoy, P. M., & Kaufman, C. K. (2022). Transcriptional profile and chromatin accessibility in zebrafish melanocytes and melanoma tumors. *G3 (Bethesda, Md.)*, *12*(1), jkab379. https://doi.org/10.1093/G3JOURNAL/JKAB379

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., & Cech, T. R. (1982). Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell*, *31*(1), 147–157. https://doi.org/10.1016/0092-8674(82)90414-7

Kubic, J. D., Little, E. C., Lui, J. W., Iizuka, T., & Lang, D. (2015). PAX3 and ETS1 synergistically activate MET expression in melanoma cells. *Oncogene*, *34*(38), 4964–4974. https://doi.org/10.1038/onc.2014.420

Kulkarni, M. M., & Arnosti, D. N. (2003). Information display by transcriptional enhancers. *Development*, *130*(26), 6569–6575. https://doi.org/10.1242/dev.00890

Kuprys, P. v., Davis, S. M., Hauer, T. M., Meltser, M., Tzfati, Y., & Kirk, K. E. (2013). Identification of Telomerase RNAs from Filamentous Fungi Reveals Conservation with Vertebrates and Yeasts. *PLoS ONE*, *8*(3), e58661. https://doi.org/10.1371/journal.pone.0058661

Kutter, C., Watt, S., Stefflova, K., Wilson, M. D., Goncalves, A., Ponting, C. P., Odom, D. T., & Marques, A. C. (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genetics*, *8*(7), e1002841. https://doi.org/10.1371/journal.pgen.1002841

Lai, K. M. V., Gong, G., Atanasio, A., Rojas, J., Quispe, J., Posca, J., White, D., Huang, M., Fedorova, D., Grant, C., Miloscio, L., Droguett, G., Poueymirou, W. T., Auerbach, W., Yancopoulos, G. D., Frendewey, D., Rinn, J., & Valenzuela, D. M. (2015). Diverse phenotypes and specific transcription patterns in twenty mouse lines with ablated lincRNAs. *PLoS ONE*, *10*(4), e0125522. https://doi.org/10.1371/journal.pone.0125522

Lasman, L., Krupalnik, V., Viukov, S., Mor, N., Aguilera-Castrejon, A., Schneir, D., Bayerl, J., Mizrahi, O., Peles, S., Tawil, S., Sathe, S., Nachshon, A., Shani, T., Zerbib, M., Kilimnik, I., Aigner, S., Shankar, A., Mueller, J. R., Schwartz, S., … Hanna, J. H. (2020). Context-dependent compensation between functional Ythdf m6A reader proteins. *Genes and Development*, *34*(19–20), 1373–1391. https://doi.org/10.1101/gad.340695.120

Laurette, P., Strub, T., Koludrovic, D., Céline Keime, le Gras, S., Seberg, H., Otterloo, E. van, Imrichova, H., Siddaway, R., Aerts, S., Cornell, R. A., Mengus, G., & Davidson, I. (2015). Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. *ELife*, *4*, e06857. https://doi.org/10.7554/eLife.06857.001

Lavalou, P., Eckert, H., Damy, L., Constanty, F., Majello, S., Bitetti, A., Graindorge, A., & Shkumatava, A. (2019). Strategies for genetic inactivation of long noncoding RNAs in zebrafish. *RNA*, *25*(8), 897–904. https://doi.org/10.1261/rna

Lee, H., Bao, S., Qian, Y., Geula, S., Leslie, J., Zhang, C., Hanna, J. H., & Ding, L. (2019). Stage-specific requirement for Mettl3-dependent m6A mRNA methylation during haematopoietic stem cell differentiation. *Nature Cell Biology*, *21*(6), 700–709. https://doi.org/10.1038/s41556-019-0318-1

Lee, S., Kopp, F., Chang, T. C., Sataluri, A., Chen, B., Sivakumar, S., Yu, H., Xie, Y., & Mendell, J. T. (2016). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell*, *164*(1–2), 69–80. https://doi.org/10.1016/j.cell.2015.12.017

Leelananda, S. P., & Lindert, S. (2016). Computational methods in drug discovery. *Beilstein Journal of Organic Chemistry*, *12*, 2694–2718. https://doi.org/10.3762/bjoc.12.267

Lefebvre, V., Dumitriu, B., Penzo-Méndez, A., Han, Y., & Pallavi, B. (2007). Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *International Journal of Biochemistry and Cell Biology*, *39*(12), 2195–2214. https://doi.org/10.1016/j.biocel.2007.05.019

Leighton, P. A., Ingram, R. S., Eggenschwiler, J., Efstratiadis, A., & Tilghman, S. M. (1995). Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature*, *375*, 34–39. https://doi.org/10.1038/375034a0

Leontis, N. B., Lescoute, A., & Westhof, E. (2006). The building blocks and motifs of RNA architecture. *Current Opinion in Structural Biology*, *16*(3), 279–287. https://doi.org/10.1016/j.sbi.2006.05.009

Lessard, L., Liu, M., Marzese, D. M., Wang, H., Chong, K., Kawas, N., Donovan, N. C., Kiyohara, E., Hsu, S., Nelson, N., Izraely, S., Sagi-Assif, O., Witz, I. P., Ma, X. J., Luo, Y., & Hoon, D. S. B. (2015). The CASC15 Long Intergenic Noncoding RNA Locus Is Involved in Melanoma Progression and Phenotype Switching. *Journal of Investigative Dermatology*, *135*(10), 2464–2474. https://doi.org/10.1038/jid.2015.200

Leucci, E., Vendramin, R., Spinazzi, M., Laurette, P., Fiers, M., Wouters, J., Radaelli, E., Eyckerman, S., Leonelli, C., Vanderheyden, K., Rogiers, A., Hermans, E., Baatsen, P., Aerts, S., Amant, F., van Aelst, S., van den Oord, J., de Strooper, B., Davidson, I., … Marine, J. C. (2016). Melanoma addiction to the long non-coding RNA SAMMSON. *Nature*, *531*(7595), 518–522. https://doi.org/10.1038/nature17161

Leulliot, N., & Varani, G. (2001). Current topics in RNA-protein recognition: Control of specificity and biological function through induced fit and conformational capture. *Biochemistry*, *40*(27), 7947–7956. https://doi.org/10.1021/bi010680y

Lewandowski, J. P., Dumbović, G., Watson, A. R., Hwang, T., Jacobs-Palmer, E., Chang, N., Much, C., Turner, K. M., Kirby, C., Rubinstein, N. D., Groff, A. F., Liapis, S. C., Gerhardinger, C., Bester, A., Pandolfi, P. P., Clohessy, J. G., Hoekstra, H. E., Sauvageau, M., & Rinn, J. L. (2020). The Tug1 lncRNA locus is essential for male fertility. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-02081-5

Lewandowski, J. P., Lee, J. C., Hwang, T., Sunwoo, H., Goldstein, J. M., Groff, A. F., Chang, N. P., Mallard, W., Williams, A., Henao-Meija, J., Flavell, R. A., Lee, J. T., Gerhardinger, C., Wagers, A. J., & Rinn, J. L. (2019). The Firre locus produces a trans-acting RNA molecule that functions in hematopoiesis. *Nature Communications*, *10*(1), 5137. https://doi.org/10.1038/s41467-019-12970-4

Li, C. Y., Zhang, Y., Wang, Z., Zhang, Y., Cao, C., Zhang, P. W., Lu, S. J., Li, X. M., Yu, Q., Zheng, X., Du, Q., Uhl, G. R., Liu, Q. R., & Wei, L. (2010). A human-specific de novo protein-coding gene associated with human brain functions. *PLoS Computational Biology*, *6*(3), e1000734. https://doi.org/10.1371/journal.pcbi.1000734

Li, F., Zhao, D., Wu, J., & Shi, Y. (2014). Structure of the YTH domain of human YTHDF2 in complex with an m6A mononucleotide reveals an aromatic cage for m6A recognition. *Cell Research*, *24*(12), 1490–1492. https://doi.org/10.1038/cr.2014.153

Li, M., Zhao, X., Wang, W., Shi, H., Pan, Q., Lu, Z., Perez, S. P., Suganthan, R., He, C., Bjørås, M., & Klungland, A. (2018). Ythdf2-mediated m6A mRNA clearance modulates neural development in mice. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1436-y

Li, R., Zhu, H., & Luo, Y. (2016). Understanding the functions of long non-coding RNAs through their higher-order structures. *International Journal of Molecular Sciences*, *17*(5), 702. https://doi.org/10.3390/ijms17050702

Li, Y., Bedi, R. K., Moroz-Omori, E. v., & Caflisch, A. (2020). Structural and Dynamic Insights into Redundant Function of YTHDF Proteins. *Journal of Chemical Information and Modeling*, *60*(12), 5932–5935. https://doi.org/10.1021/acs.jcim.0c01029

Li, Y., Tan, Z., Zhang, Y., Zhang, Z., Hu, Q., Liang, K., Jun, Y., Ye, Y., Li, Y.-C., Li, C., Liao, L., Xu, J., Xing, Z., Pan, Y., Chatterjee, S. S., Nguyen, T. K., Hsiao, H., Egranov, S. D., Putluri, N., … Yang, L. (2021). A noncoding RNA modulator potentiates phenylalanine metabolism in mice. *Science*, *373*(6555), 662–673. https://doi.org/10.1126/science.aba4991

Li, Z., Weng, H., Su, R., Weng, X., Zuo, Z., Li, C., Huang, H., Nachtergaele, S., Dong, L., Hu, C., Qin, X., Tang, L., Wang, Y., Hong, G. M., Huang, H., Wang, X., Chen, P., Gurbuxani, S., Arnovitz, S., … Chen, J. (2017). FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a N6-Methyladenosine RNA Demethylase. *Cancer Cell*, *31*(1), 127–141. https://doi.org/10.1016/j.ccell.2016.11.017

Lim, D., Byun, W. G., Koo, J. Y., Park, H., & Park, S. B. (2016). Discovery of a Small-Molecule Inhibitor of Protein-MicroRNA Interaction Using Binding Assay with a Site-Specifically Labeled Lin28. *Journal of the American Chemical Society*, *138*(41), 13630–13638. https://doi.org/10.1021/jacs.6b06965

Lin, Z., Hsu, P. J., Xing, X., Fang, J., Lu, Z., Zou, Q., Zhang, K. J., Zhang, X., Zhou, Y., Zhang, T., Zhang, Y., Song, W., Jia, G., Yang, X., He, C., & Tong, M. H. (2017). Mettl3-/Mettl14-mediated mRNA N 6-methyladenosine modulates murine spermatogenesis. *Cell Research*, *27*(10), 1216–1230. https://doi.org/10.1038/cr.2017.117

Linder, B., Grozhik, A. v., Olarerin-George, A. O., Meydan, C., Mason, C. E., & Jaffrey, S. R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nature Methods*, *12*(8), 767–772. https://doi.org/10.1038/nmeth.3453

Linder, P., & Jankowsky, E. (2011). From unwinding to clamping – the DEAD box RNA helicase family. *Nature Reviews Molecular Cell Biology*, *12*(8), 505–516. https://doi.org/10.1038/nrm3154

Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., Arenas-Huertero, C., & Chua, N. H. (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in arabidopsis. *Plant Cell*, *24*(11), 4333–4345. https://doi.org/10.1105/tpc.112.102855

Liu, J., Li, K., Cai, J., Zhang, M., Zhang, X., Xiong, X., Meng, H., Xu, X., Huang, Z., Peng, J., Fan, J., & Yi, C. (2020). Landscape and Regulation of m6A and m6Am Methylome across Human and Mouse Tissues. *Molecular Cell*, *77*(2), 426–440. https://doi.org/10.1016/j.molcel.2019.09.032

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., & Pan, T. (2015). N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature*, *518*(7540), 560–564. https://doi.org/10.1038/nature14234

Liu, S. J., Nowakowski, T. J., Pollen, A. A., Lui, J. H., Horlbeck, M. A., Attenello, F. J., He, D., Weissman, J. S., Kriegstein, A. R., Diaz, A. A., & Lim, D. A. (2016). Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biology*, *17*(1). https://doi.org/10.1186/s13059-016-0932-1

Lo, K. Y., Li, Z., Bussiere, C., Bresson, S., Marcotte, E. M., & Johnson, A. W. (2010). Defining the pathway of cytoplasmic maturation of the 60S ribosomal subunit. *Molecular Cell*, *39*(2), 196–208. https://doi.org/10.1016/j.molcel.2010.06.018

Lourenço, A. R., & Coffer, P. J. (2017). SOX4: Joining the Master Regulators of Epithelial-to-Mesenchymal Transition? *Trends in Cancer*, *3*(8), 571–582. https://doi.org/10.1016/j.trecan.2017.06.002

Lu, Z., Zhang, Q. C., Lee, B., Flynn, R. A., Smith, M. A., Robinson, J. T., Davidovich, C., Gooding, A. R., Goodrich, K. J., Mattick, J. S., Mesirov, J. P., Cech, T. R., & Chang, H. Y. (2016). RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell*, *165*(5), 1267–1279. https://doi.org/10.1016/j.cell.2016.04.028

Lubelsky, Y., & Ulitsky, I. (2018). Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature*, *555*(7694), 107–111. https://doi.org/10.1038/nature25757

Lucas, B., Grigo, K., Erdmann, S., Lausen, J., Klein-Hitpass, L., & Ryffel, G. U. (2005). HNF4α reduces proliferation of kidney cells and affects genes deregulated in renal cell carcinoma. *Oncogene*, *24*(42), 6418–6431. https://doi.org/10.1038/sj.onc.1208794

Luke, B., Panza, A., Redon, S., Iglesias, N., Li, Z., & Lingner, J. (2008). The Rat1p 5′ to 3′ Exonuclease Degrades Telomeric Repeat-Containing RNA and Promotes Telomere Elongation in Saccharomyces cerevisiae. *Molecular Cell*, *32*(4), 465–477. https://doi.org/10.1016/j.molcel.2008.10.019

Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: Modular design for efficient function. *Nature Reviews Molecular Cell Biology*, *8*(6), 479–490. https://doi.org/10.1038/nrm2178

Maamar, H., Cabili, M. N., Rinn, J., & Raj, A. (2013). linc-HOXA1 is a noncoding RNA that represses HoXa1 transcription in cis. *Genes and Development*, *27*(11), 1260–1271. https://doi.org/10.1101/gad.217018.113

Macarron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T., Green, D. V. S., Hertzberg, R. P., Janzen, W. P., Paslay, J. W., Schopfer, U., & Sittampalam, G. S. (2011). Impact of high-throughput screening in biomedical research. *Nature Reviews Drug Discovery*, *10*(3), 188–195. https://doi.org/10.1038/nrd3368

Macleod, Kay., Leprince, D., & Stehelin, D. (1992). The ets gene family. *Trends in Biochemical Sciences*, *17*(7), 251–256. https://doi.org/10.1016/0968-0004(92)90404-w

Malim, M. H., Hauber, J., Le, S.-Y., Maizel, J. v., & Cullen, B. R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature*, *338*, 254–257. https://doi.org/10.1038/338254a0

Mallory, A. C., & Shkumatava, A. (2015). LncRNAs in vertebrates: Advances and challenges. *Biochimie*, *117*, 3–14. https://doi.org/10.1016/j.biochi.2015.03.014

Mao, P., Brown, A. J., Esaki, S., Lockwood, S., Poon, G. M. K., Smerdon, M. J., Roberts, S. A., & Wyrick, J. J. (2018). ETS transcription factors induce a unique UV damage signature that drives recurrent mutagenesis in melanoma. *Nature Communications*, *9*(1), 2626. https://doi.org/10.1038/s41467-018-05064-0

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., & Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes and Development*, *11*(2), 156–166. https://doi.org/10.1101/gad.11.2.156

Marcaida, M. J., Kauzlaric, A., Duperrex, A., Sülzle, J., Moncrieffe, M. C., Adebajo, D., Manley, S., Trono, D., & Dal Peraro, M. (2020). The Human RNA Helicase DDX21 Presents a Dimerization Interface Necessary for Helicase Activity. *IScience*, *23*(12), 101811. https://doi.org/10.1016/j.isci.2020.101811

Marcheschi, R. J., Mouzakis, K. D., & Butcher, S. E. (2009). Selection and characterization of small molecules that bind the HIV-1 frameshift site RNA. *ACS Chemical Biology*, *4*(10), 844–854. https://doi.org/10.1021/cb900167m

Marín-Béjar, O., Marchese, F. P., Athie, A., Sánchez, Y., González, J., Segura, V., Huang, L., Moreno, I., Navarro, A., Monzó, M., García-Foncillas, J., Rinn, J. L., Guo, S., & Huarte, M. (2013). Pint lincrna connects the p53 pathway with epigenetic silencing by the polycomb repressive complex 2. *Genome Biology*, *14*(9). https://doi.org/10.1186/gb-2013-14-9-r104

Marín-Béjar, O., Mas, A. M., González, J., Martinez, D., Athie, A., Morales, X., Galduroz, M., Raimondi, I., Grossi, E., Guo, S., Rouzaut, A., Ulitsky, I., & Huarte, M. (2017). The human lncRNA LINC-PINT inhibits tumor cell invasion through a highly conserved sequence element. *Genome Biology*, *18*(1). https://doi.org/10.1186/s13059-017-1331-y

Marinello, J., Delcuratolo, M., & Capranico, G. (2018). Anthracyclines as Topoisomerase II poisons: From early studies to new perspectives. *International Journal of Molecular Sciences*, *19*(11), 3480. https://doi.org/10.3390/ijms19113480

Mariño-Ramírez, L., Lewis, K. C., Landsman, D., & Jordan, I. K. (2005). Transposable elements donate lineage-specific regulatory sequences to host genomes. *Cytogenetic and Genome Research*, *110*(1–4), 333–341. https://doi.org/10.1159/000084965

Maris, C., Dominguez, C., & Allain, F. H. T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal*, *272*(9), 2118–2131. https://doi.org/10.1111/j.1742-4658.2005.04653.x

Marti, F., Krause, A., Post, N. H., Lyddane, C., Dupont, B., Sadelain, M., & King, P. D. (2001). Negative-Feedback Regulation of CD28 Costimulation by a Novel Mitogen-Activated Protein Kinase Phosphatase, MKP6. *The Journal of Immunology*, *166*(1), 197–206. https://doi.org/10.4049/jimmunol.166.1.197

Martins-Teixeira, M. B., & Carvalho, I. (2020). Antitumour Anthracyclines: Progress and Perspectives. *ChemMedChem*, *15*(11), 933–948. https://doi.org/10.1002/cmdc.202000131

Matthews, M. M., Thomas, J. M., Zheng, Y., Tran, K., Phelps, K. J., Scott, A. I., Havel, J., Fisher, A. J., & Beal, P. A. (2016). Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nature Structural and Molecular Biology*, *23*(5), 426–433. https://doi.org/10.1038/nsmb.3203

Mattioli, K., Volders, P. J., Gerhardinger, C., Lee, J. C., Maass, P. G., Melé, M., & Rinn, J. L. (2019). High-throughput functional analysis of lncRNA core promoters elucidates rules governing tissue specificity. *Genome Research*, *29*(3), 344–355. https://doi.org/10.1101/gr.242222.118

McHugh, C. A., Chen, C. K., Chow, A., Surka, C. F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M. J., Shishkin, A. A., Su, J., Lander, E. S., Hess, S., Plath, K., & Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*, *521*(7551), 232–236. https://doi.org/10.1038/nature14443

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

McRae, E. K. S., Booy, E. P., Moya-Torres, A., Ezzati, P., Stetefeld, J., & McKenna, S. A. (2017). Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic Acids Research*, *45*(11), 6656–6668. https://doi.org/10.1093/nar/gkx380

Mei, H.-Y., Cui, M., Heldsinger, A., Lemrow, S. M., Loo, J. A., Sannes-Lowery, K. A., Sharmeen, L., & Czarnik, A. W. (1998). Inhibitors of Protein-RNA Complexation That Target the RNA: Specific Recognition of Human Immunodeficiency Virus Type 1 TAR RNA by Small Organic Molecules. *Biochemistry*, *37*(40), 14204–14212. https://doi.org/10.1021/bi981308u

Meisner, N. C., Hintersteiner, M., Mueller, K., Bauer, R., Seifert, J. M., Naegeli, H. U., Ottl, J., Oberer, L., Guenat, C., Moss, S., Harrer, N., Woisetschlaeger, M., Buehler, C., Uhl, V., & Auer, M. (2007). Identification and mechanistic characterization of low-molecular-weight inhibitors for HuR. *Nature Chemical Biology*, *3*(8), 508–515. https://doi.org/10.1038/nchembio.2007.14

Meller, V. H., Kwok, §, Wu, H., & Roman, G. (1997). roX1 RNA Paints the X Chromosome of Male Drosophila and Is Regulated by the Dosage Compensation System. *Cell*, *88*(4), 445–457. https://doi.org/10.1016/S0092-8674(00)81885-1

Meller, V. H., & Rattner, B. P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *The EMBO Journal*, *21*(5), 1084–1091. https://doi.org/10.1093/emboj/21.5.1084

Mendel, M., Delaney, K., Pandey, R. R., Chen, K. M., Wenda, J. M., Vågbø, C. B., Steiner, F. A., Homolka, D., & Pillai, R. S. (2021). Splice site m6A methylation prevents binding of U2AF35 to inhibit RNA splicing. *Cell*, *184*(12), 3125–3142. https://doi.org/10.1016/j.cell.2021.03.062

Meng, Q., Stoyko, D., Andrews, C. M., Konstantinidou, P., Genzor, P., O, T., Elchert, A. R., Benner, L., Sobti, S., Katz, E. Y., & Haase, A. D. (2022). Functional editing of endogenous genes through rapid selection of cell pools (Rapid generation of endogenously tagged genes in Drosophila ovarian somatic sheath cells). *Nucleic Acids Research*, *27*, gkac448. https://doi.org/10.1093/nar/gkac448

Mercuri, E., Darras, B. T., Chiriboga, C. A., Day, J. W., Campbell, C., Connolly, A. M., Iannaccone, S. T., Kirschner, J., Kuntz, N. L., Saito, K., Shieh, P. B., Tulinius, M., Mazzone, E. S., Montes, J., Bishop, K. M., Yang, Q., Foster, R., Gheuens, S., Bennett, C. F., … Finkel, R. S. (2018). Nusinersen versus Sham Control in Later-Onset Spinal Muscular Atrophy. *New England Journal of Medicine*, *378*(7), 625–635. https://doi.org/10.1056/nejmoa1710504

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., & Jaffrey, S. R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell*, *149*(7), 1635–1646. https://doi.org/10.1016/j.cell.2012.05.003

Mialon, A., Thastrup, J., Kallunki, T., Mannermaa, L., Westermarck, J., & Holmström, T. H. (2008). Identification of nucleolar effects in JNK-deficient cells. *FEBS Letters*, *582*(20), 3145–3151. https://doi.org/10.1016/j.febslet.2008.08.004

Minuesa, G., Albanese, S. K., Xie, W., Kazansky, Y., Worroll, D., Chow, A., Schurer, A., Park, S. M., Rotsides, C. Z., Taggart, J., Rizzi, A., Naden, L. N., Chou, T., Gourkanti, S., Cappel, D., Passarelli, M. C., Fairchild, L., Adura, C., Glickman, J. F., … Kharas, M. G. (2019). Small-molecule targeting of MUSASHI RNA-binding activity in acute myeloid leukemia. *Nature Communications*, *10*(1), 2691. https://doi.org/10.1038/s41467-019-10523-3

Mitra, S., Muralidharan, S. V., Marco, M. di, Juvvuna, P. K., Kosalai, S. T., Reischl, S., Jachimowicz, D., Subhash, S., Raimondi, I., Kurian, L., Huarte, M., Kogner, P., Fischer, M., Johnsen, J. I., Mondal, T., & Kanduri, C. (2021). Subcellular distribution of p53 by the p53-Responsive lncRNA NBAT1 Determines Chemotherapeutic Response in Neuroblastoma. *Cancer Research*, *81*(6), 1457–1471. https://doi.org/10.1158/0008-5472.CAN-19-3499

Moffat, J. G., Vincent, F., Lee, J. A., Eder, J., & Prunotto, M. (2017). Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nature Reviews Drug Discovery*, *16*(8), 531–543. https://doi.org/10.1038/nrd.2017.111

Mohibi, S., Chen, X., & Zhang, J. (2019). Cancer the'RBP'eutics–RNA-binding proteins as therapeutic targets for cancer. *Pharmacology and Therapeutics*, *203*. https://doi.org/10.1016/j.pharmthera.2019.07.001

Mondal, T., Juvvuna, P. K., Kirkeby, A., Mitra, S., Kosalai, S. T., Traxler, L., Hertwig, F., Wernig-Zorc, S., Miranda, C., Deland, L., Volland, R., Bartenhagen, C., Bartsch, D., Bandaru, S., Engesser, A., Subhash, S., Martinsson, T., Carén, H., Akyürek, L. M., … Kanduri, C. (2018). Sense-Antisense lncRNA Pair Encoded by Locus 6p22.3 Determines Neuroblastoma Susceptibility via the USP36-CHD7-SOX9 Regulatory Axis. *Cancer Cell*, *33*(3), 417–434. https://doi.org/10.1016/j.ccell.2018.01.020

Monfort, A., di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., & Wutz, A. (2015). Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. *Cell Reports*, *12*(4), 554–561. https://doi.org/10.1016/j.celrep.2015.06.067

Morandi, E., Manfredonia, I., Simon, L. M., Anselmi, F., van Hemert, M. J., Oliviero, S., & Incarnato, D. (2021). Genome-scale deconvolution of RNA structure ensembles. *Nature Methods*, *18*(3), 249–252. https://doi.org/10.1038/s41592-021-01075-w

Mort, R. L., Jackson, I. J., & Elizabeth Patton, E. (2015). The melanocyte lineage in development and disease. *Development (Cambridge)*, *142*(4), 620–632. https://doi.org/10.1242/dev.106567

Much, C., Lasda, E. L., Lewandowski, J. P., Smallegan, M. J., & Rinn, J. L. (2022). The lncRNA Firre functions as a transcriptional activator from a distance. *Preprint at BioRxiv*. https://doi.org/10.1101/2022.05.15.492001

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

Muda, M., Manning, E. R., Orth, K., & Dixon, J. E. (1999). Identification of the Human YVH1 Protein-tyrosine Phosphatase Orthologue Reveals a Novel Zinc Binding Domain Essential for in Vivo Function. *Journal of Biological Chemistry*, *274*(34), 23991–23995. https://doi.org/10.1074/jbc.274.34.23991

Mukherjee, N., Calviello, L., Hirsekorn, A., de Pretis, S., Pelizzola, M., & Ohler, U. (2017). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nature Structural and Molecular Biology*, *24*(1), 86–96. https://doi.org/10.1038/nsmb.3325

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, *320*(5881), 1344–1349. https://doi.org/10.1126/science.1158441

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

Nakagawa, S., Naganuma, T., Shioi, G., & Hirose, T. (2011). Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *Journal of Cell Biology*, *193*(1), 31–39. https://doi.org/10.1083/jcb.201011110

Nakagawa, S., Shimada, M., Yanaka, K., Mito, M., Arai, T., Takahashi, E., Fujita, Y., Fujimori, T., Standaert, L., Marine, J. C., & Hirose, T. (2014). The lncRNA Neat1 is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. *Development (Cambridge)*, *141*(23), 4618–4627. https://doi.org/10.1242/dev.110544

Nakayama, T., Shimmura, T., Shinomiya, A., Okimura, K., Takehana, Y., Furukawa, Y., Shimo, T., Senga, T., Nakatsukasa, M., Nishimura, T., Tanaka, M., Okubo, K., Kamei, Y., Naruse, K., & Yoshimura, T. (2019). Seasonal regulation of the lncRNA LDAIR modulates self-protective behaviours during the breeding season. *Nature Ecology and Evolution*, *3*(5), 845–852. https://doi.org/10.1038/s41559-019-0866-6

Nam, J. W., & Bartel, D. P. (2012). Long noncoding RNAs in C. elegans. *Genome Research*, *22*(12), 2529–2540. https://doi.org/10.1101/gr.140475.112

Naryshkin, N. A., Weetall, M., Dakka, A., Narasimhan, J., Zhao, X., Feng, Z., Ling, K. K. Y., Karp, G. M., Qi, H., Woll, M. G., Chen, G., Zhang, N., Gabbeta, V., Vazirani, P., Bhattacharyya, A., Furia, B., Risher, N., Sheedy, J., Kong, R., … Metzger, F. (2014). SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science*, *345*(6197), 688–693. https://doi.org/10.1126/science.1250127

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

Neme, R., & Tautz, D. (2013). Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. *BMC Genomics*, *14*(1). https://doi.org/10.1186/1471-2164-14-117

Nesterova, T. B., Slobodyanyuk, S. Y., Elisaphenko, E. A., Shevchenko, A. I., Johnston, C., Pavlova, M. E., Rogozin, I. B., Kolesnikov, N. N., Brockdorff, N., & Zakian, S. M. (2001). Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Research*, *11*(5), 833–849. https://doi.org/10.1101/gr.174901

Nguyễn, L. B., Diskin, S. J., Capasso, M., Wang, K., Diamond, M. A., Glessner, J., Kim, C., Attiyeh, E. F., Mosse, Y. P., Cole, K., Iolascon, A., Devoto, M., Hakonarson, H., Li, H. K., & Maris, J. M. (2011). Phenotype restricted genome-wide association study using a gene-centric approach identifies three low-risk neuroblastoma susceptibility loci. *PLoS Genetics*, *7*(3). https://doi.org/10.1371/journal.pgen.1002026

Nitiss, J. L. (2009). Targeting DNA topoisomerase II in cancer chemotherapy. *Nature Reviews Cancer*, *9*(5), 338–350. https://doi.org/10.1038/nrc2607

Nitsche, A., Rose, D., Fasold, M., Reiche, K., & Stadler, P. F. (2015). Comparison of splice sites reveals that long noncoding RNAs are evolutionarily well conserved. *RNA*, *21*(5), 801–812. https://doi.org/10.1261/rna.046342.114

Novikova, I. v., Hennelly, S. P., & Sanbonmatsu, K. Y. (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Research*, *40*(11), 5034–5051. https://doi.org/10.1093/nar/gks071

Novikova, I. v., Hennelly, S. P., Tung, C. S., & Sanbonmatsu, K. Y. (2013). Rise of the RNA machines: Exploring the structure of long non-coding RNAs. *Journal of Molecular Biology*, *425*(19), 3731–3746. https://doi.org/10.1016/j.jmb.2013.02.030

Olazagoitia-Garmendia, A., Garcia-Moreno, F., & Castellanos-Rubio, A. (2022). Nonlinear sequence similarity analysis and validation of evolutionary convergent long noncoding RNAs involved in embryonic development. *Preprint at BioRxiv*. https://doi.org/10.1101/2022.06.15.496228

Omer, S., Harlow, T. J., & Gogarten, J. P. (2017). Does Sequence Conservation Provide Evidence for Biological Function? *Trends in Microbiology*, *25*(1), 11–18. https://doi.org/10.1016/j.tim.2016.09.010

Ounzain, S., Micheletti, R., Arnan, C., Plaisance, I., Cecchi, D., Schroen, B., Reverter, F., Alexanian, M., Gonzales, C., Ng, S. Y., Bussotti, G., Pezzuto, I., Notredame, C., Heymans, S., Guigó, R., Johnson, R., & Pedrazzini, T. (2015). CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. *Journal of Molecular and Cellular Cardiology*, *89*, 98–112. https://doi.org/10.1016/j.yjmcc.2015.09.016

Owen, R., & Cooper, W. W. (1843). *Lectures on the Comparative Anatomy and Physiology of the Invertebrate Animals: Delivered at the Royal College of Surgeons, in 1843*. London, Longman, Brown, Green, and Longmans. https://doi.org/10.5962/bhl.title.6788

Palacino, J., Swalley, S. E., Song, C., Cheung, A. K., Shu, L., Zhang, X., van Hoosear, M., Shin, Y., Chin, D. N., Keller, C. G., Beibel, M., Renaud, N. A., Smith, T. M., Salcius, M., Shi, X., Hild, M., Servais, R., Jain, M., Deng, L., … Sivasankaran, R. (2015). SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nature Chemical Biology*, *11*(7), 511–517. https://doi.org/10.1038/nchembio.1837

Pallett, M. A., Lu, Y., & Smith, G. L. (2022). DDX50 Is a Viral Restriction Factor That Enhances IRF3 Activation. *Viruses*, *14*(2), 316. https://doi.org/10.3390/v14020316

Pandey, G. K., Mitra, S., Subhash, S., Hertwig, F., Kanduri, M., Mishra, K., Fransson, S., Ganeshram, A., Mondal, T., Bandaru, S., Östensson, M., Akyürek, L. M., Abrahamsson, J., Pfeifer, S., Larsson, E., Shi, L., Peng, Z., Fischer, M., Martinsson, T., … Kanduri, C. (2014). The Risk-Associated Long Noncoding RNA NBAT-1 Controls Neuroblastoma Progression by Regulating Cell Proliferation and Neuronal Differentiation. *Cancer Cell*, *26*(5), 722–737. https://doi.org/10.1016/j.ccell.2014.09.014

Pang, K. C., Frith, M. C., & Mattick, J. S. (2006). Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in Genetics*, *22*(1), 1–5. https://doi.org/10.1016/j.tig.2005.10.003

Paris, J., Morgan, M., Campos, J., Spencer, G. J., Shmakova, A., Ivanova, I., Mapperley, C., Lawson, H., Wotherspoon, D. A., Sepulveda, C., Vukovic, M., Allen, L., Sarapuu, A., Tavosanis, A., Guitart, A. v., Villacreces, A., Much, C., Choe, J., Azar, A., … Kranc, K. R. (2019). Targeting the RNA m6A Reader YTHDF2 Selectively Compromises Cancer Stem Cells in Acute Myeloid Leukemia. *Cell Stem Cell*, *25*(1), 137–148. https://doi.org/10.1016/j.stem.2019.03.021

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., Aster, J. C., Granter, S. R., Look, A. T., Lee, C., Fisher, D. E., & Zon, L. I. (2005). BRAF Mutations Are Sufficient to Promote Nevi Formation and Cooperate with p53 in the Genesis of Melanoma. *Current Biology*, *15*(3), 249–254. https://doi.org/10.1016/j.cub.2005.01.031

Patzelt, E., Blaas, D., & Kuechler, E. (1983). CAP binding proteins associated with the nucleus. *Nucleic Acids Research*, *11*(17), 5821–5835. https://doi.org/10.1093/nar/11.17.5821

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

Paunovska, K., Loughrey, D., & Dahlman, J. E. (2022). Drug delivery systems for RNA therapeutics. *Nature Reviews Genetics*, *23*, 265–280. https://doi.org/10.1038/s41576-021-00439-4

Pearson, H. (2002). Surviving a knockout blow. *Nature*, *415*, 8–9. https://doi.org/10.1038/415008a

Pendleton, K. E., Chen, B., Liu, K., Hunter, O. v., Xie, Y., Tu, B. P., & Conrad, N. K. (2017). The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell*, *169*(5), 824–835. https://doi.org/10.1016/j.cell.2017.05.003

Pérez-Rico, Y. A., Barillot, E., & Shkumatava, A. (2020). Demarcation of Topologically Associating Domains Is Uncoupled from Enriched CTCF Binding in Developing Zebrafish. *IScience*, *23*(5). https://doi.org/10.1016/j.isci.2020.101046

Perry, R. B. T., Hezroni, H., Goldrich, M. J., & Ulitsky, I. (2018). Regulation of Neuroregeneration by Long Noncoding RNAs. *Molecular Cell*, *72*(3), 553-567.e5. https://doi.org/10.1016/j.molcel.2018.09.021

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

Podlevsky, J. D., & Chen, J. J. L. (2016). Evolutionary perspectives of telomerase RNA structure and function. *RNA Biology*, *13*(8), 720–732. https://doi.org/10.1080/15476286.2016.1205768

Ponting, C. P., & Haerty, W. (2022). Genome-Wide Analysis of Human Long Noncoding RNAs: A Provocative Review. *Annual Review of Genomics and Human Genetics*, *23*, 6.1-6.20. https://doi.org/10.1146/annurev-genom-112921

Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and Functions of Long Noncoding RNAs. *Cell*, *136*(4), 629–641. https://doi.org/10.1016/j.cell.2009.02.006

Potzner, M. R., Tsarovina, K., Binder, E., Penzo-Méndez, A., Lefebvre, V., Rohrer, H., Wegner, M., & Sock, E. (2010). Sequential requirement of Sox4 and Sox11 during development of the sympathetic nervous system. *Development (Cambridge)*, *137*(5), 775–784. https://doi.org/10.1242/dev.042101

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., McNamee, C., Norris, A., Sanseau, P., Cavalla, D., & Pirmohamed, M. (2018). Drug repurposing: Progress, challenges and recommendations. *Nature Reviews Drug Discovery*, *18*(1), 41–58. https://doi.org/10.1038/nrd.2018.168

Qi, X., Li, Y., Honda, S., Hoffmann, S., Marz, M., Mosig, A., Podlevsky, J. D., Stadler, P. F., Selker, E. U., & Chen, J. J. L. (2013). The common ancestral core of vertebrate and fungal telomerase RNAs. *Nucleic Acids Research*, *41*(1), 450–462. https://doi.org/10.1093/nar/gks980

Quattrone, A. (2020). *ROles of ePitranscriptomic in diseasES*. Horizon 2020 - MSCA.

Quinn, J. J., Zhang, Q. C., Georgiev, P., Ilik, I. A., Akhtar, A., & Chang, H. Y. (2016). Rapid evolutionary turnover underlies conserved lncRNA-genome interactions. *Genes and Development*, *30*(2), 191–207. https://doi.org/10.1101/gad.272187

Rahmouni, S., Cerignoli, F., Alonso, A., Tsutji, T., Henkens, R., Zhu, C., Louis-Dit-sully, C., Moutschen, M., Jiang, W., & Mustelin, T. (2006). Loss of the VHR dual-specific phosphatase causescell-cycle arrest and senescence. *Nature Cell Biology*, *8*(5), 524–531. https://doi.org/10.1038/ncb1398

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

Ramírez-Colmenero, A., Oktaba, K., & Fernandez-Valverde, S. L. (2020). Evolution of Genome-Organizing Long Non-coding RNAs in Metazoans. *Frontiers in Genetics*, *30*(11), 589697. https://doi.org/10.3389/fgene.2020.589697

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., & Lim, D. A. (2015). The long noncoding RNA Pnky regulates neuronal differentiation of embryonic and postnatal neural stem cells. *Cell Stem Cell*, *16*(4), 439–447. https://doi.org/10.1016/j.stem.2015.02.007

Ransohoff, J. D., Wei, Y., & Khavari, P. A. (2018). The functions and unique features of long intergenic non-coding RNA. *Nature Reviews Molecular Cell Biology*, *19*(3), 143–157. https://doi.org/10.1038/nrm.2017.104

Ravasi, T., Suzuki, H., Pang, K. C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M. C., Gongora, M. M., Grimmond, S. M., Hume, D. A., Hayashizaki, Y., & Mattick, J. S. (2006). Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Research*, *16*(1), 11–19. https://doi.org/10.1101/gr.4200206

Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L. H., Dale, R. K., Smith, S. A., Yarosh, C. A., Kelly, S. M., Nabet, B., Mecenas, D., … Hughes, T. R. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, *499*(7457), 172–177. https://doi.org/10.1038/nature12311

Reboll, M. R., Korf-Klingebiel, M., Klede, S., Polten, F., Brinkmann, E., Reimann, I., Schönfeld, H. J., Bobadilla, M., Faix, J., Kensah, G., Gruh, I., Klintschar, M., Gaestel, M., Niessen, H. W., Pich, A., Bauersachs, J., Gogos, J. A., Wang, Y., & Wollert, K. C. (2017). EMC10 (Endoplasmic reticulum membrane protein complex subunit 10) is a bone marrow-derived angiogenic growth factor promoting tissue repair after myocardial infarction. *Circulation*, *136*(19), 1809–1823. https://doi.org/10.1161/CIRCULATIONAHA.117.029980

Ribeiro, D. M., Zanzoni, A., Cipriano, A., Delli Ponti, R., Spinelli, L., Ballarino, M., Bozzoni, I., Tartaglia, G. G., & Brun, C. (2018). Protein complex scaffolding predicted as a prevalent function of long non-coding RNAs. *Nucleic Acids Research*, *46*(2), 917–928. https://doi.org/10.1093/nar/gkx1169

Rinn, J. L., & Chang, H. Y. (2020). Long Noncoding RNAs: Molecular Modalities to Organismal Functions. *Annual Review of Biochemistry*, *89*, 283–308. https://doi.org/10.1146/annurev-biochem-062917

Ritter, N., Ali, T., Kopitchinski, N., Schuster, P., Beisaw, A., Hendrix, D. A., Schulz, M. H., Müller-McNicoll, M., Dimmeler, S., & Grote, P. (2019). The lncRNA Locus Handsdown Regulates Cardiac Gene Programs and Is Essential for Early Mouse Development. *Developmental Cell*, *50*(5), 644–657. https://doi.org/10.1016/j.devcel.2019.07.013

Rivas, E. (2020). RNA structure prediction using positive and negative evolutionary information. *PLoS Computational Biology*, *16*(10), e1008387. https://doi.org/10.1371/journal.pcbi.1008387

Rivas, E., Clements, J., & Eddy, S. R. (2016). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nature Methods*, *14*(1), 45–48. https://doi.org/10.1038/nmeth.4066

Roberts, T. C., Langer, R., & Wood, M. J. A. (2020). Advances in oligonucleotide drug delivery. *Nature Reviews Drug Discovery*, *19*(10), 673–694. https://doi.org/10.1038/s41573-020-0075-7

Rom, A., Melamed, L., Gil, N., Goldrich, M. J., Kadir, R., Golan, M., Biton, I., Perry, R. B. T., & Ulitsky, I. (2019). Regulation of CHD2 expression by the Chaserr long noncoding RNA gene is essential for viability. *Nature Communications*, *10*(1), 5092. https://doi.org/10.1038/s41467-019-13075-8

Romito, A., & Rougeulle, C. (2011). Origin and evolution of the long non-coding genes in the X-inactivation center. In *Biochimie* (Vol. 93, Issue 11, pp. 1935–1942). https://doi.org/10.1016/j.biochi.2011.07.009

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

Ross, C. J., & Ulitsky, I. (2022). Discovering functional motifs in long noncoding RNAs. *WIREs RNA*, e1708. https://doi.org/10.1002/wrna.1708

Rothhammer, T., Hahne, J. C., Florin, A., Poser, I., Soncin, F., Wernert, N., & Bosserhoff, A. K. (2004). The Ets-1 transcription factor is involved in the development and invasion of malignant melanoma. *Cellular and Molecular Life Sciences*, *61*(1), 118–128. https://doi.org/10.1007/s00018-003-3337-8

Roundtree, I. A., Luo, G.-Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., Sha, J., Huang, X., Guerrero, L., Xie, P., He, E., Shen, B., & He, C. (2017). YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *ELife*, *6*, e31311. https://doi.org/10.7554/eLife.31311.001

Runtuwene, V., van Eekelen, M., Overvoorde, J., Rehmann, H., Yntema, H. G., Nillesen, W. M., van Haeringen, A., van der Burgt, I., Burgering, B., & den Hertog, J. (2011). Noonan syndrome gain-of-function mutations in NRAS cause zebrafish gastrulation defects. *Disease Models and Mechanisms*, *4*(3), 393–399. https://doi.org/10.1242/dmm.007112

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., Zhang, S., Gnanchandran, J., Seeger, R., Asgharzadeh, S., Khan, J., Diskin, S. J., Maris, J. M., & Cole, K. A. (2015). CASC15-S is a tumor suppressor lncRNA at the 6p22 neuroblastoma susceptibility locus. *Cancer Research*, *75*(15), 3155–3166. https://doi.org/10.1158/0008-5472.CAN-14-3613

Sabaté-Cadenas, X., & Shkumatava, A. (2020). In-Cell Discovery of RNA–Protein Interactions. *Trends in Biochemical Sciences*, *45*(3), 272–273. https://doi.org/10.1016/j.tibs.2019.12.003

Sallam, M. A., Prakash, S., Kumbhojkar, N., Shields, C. W., & Mitragotri, S. (2021). Formulation-based approaches for dermal delivery of vaccines and therapeutic nucleic acids: Recent advances and future perspectives. *Bioengineering and Translational Medicine*, *6*(3), e10215. https://doi.org/10.1002/btm2.10215

Salta, E., & de Strooper, B. (2017). Noncoding RNAs in neurodegeneration. *Nature Reviews Neuroscience*, *18*(10), 627–640. https://doi.org/10.1038/nrn.2017.90

Sanborn, A. L., Rao, S. S. P., Huang, S. C., Durand, N. C., Huntley, M. H., Jewett, A. I., Bochkov, I. D., Chinnappan, D., Cutkosky, A., Li, J., Geeting, K. P., Gnirke, A., Melnikov, A., McKenna, D., Stamenova, E. K., Lander, E. S., & Aiden, E. L. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(47), E6456–E6465. https://doi.org/10.1073/pnas.1518552112

Sanchez de Groot, N., Armaos, A., Graña-Montes, R., Alriquet, M., Calloni, G., Vabulas, R. M., & Tartaglia, G. G. (2019). RNA structure drives interaction with proteins. *Nature Communications*, *10*(1), 3246. https://doi.org/10.1038/s41467-019-10923-5

Sander, J. D., Dahlborg, E. J., Goodwin, M. J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S. J., Blackburn, J. S., Thibodeau-Beganny, S., Qi, Y., Pierick, C. J., Hoffman, E., Maeder, M. L., Khayter, C., Reyon, D., Dobbs, D., Langenau, D. M., Stupar, R. M., Giraldez, A. J., … Joung, J. K. (2011). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature Methods*, *8*(1), 67–69. https://doi.org/10.1038/nmeth.1542

Santoriello, C., Sporrij, A., Yang, S., Flynn, R. A., Henriques, T., Dorjsuren, B., Custo Greig, E., McCall, W., Stanhope, M. E., Fazio, M., Superdock, M., Lichtig, A., Adatto, I., Abraham, B. J., Kalocsay, M., Jurynec, M., Zhou, Y., Adelman, K., Calo, E., & Zon, L. I. (2020). RNA helicase DDX21 mediates nucleotide stress responses in neural crest and melanoma cells. *Nature Cell Biology*, *22*(4), 372–379. https://doi.org/10.1038/s41556-020-0493-0

Sasaki, Y. T. F., Ideue, T., Sano, M., Mituyama, T., & Hirose, T. (2009). MENε/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(8), 2525–2530. https://doi.org/10.1073/pnas.0807899106.

Sauvageau, M., Goff, L. A., Lodato, S., Bonev, B., Groff, A. F., Gerhardinger, C., Sanchez-Gomez, D. B., Hacisuleyman, E., Li, E., Spence, M., Liapis, S. C., Mallard, W., Morse, M., Swerdel, M. R., D'Ecclessis, M. F., Moore, J. C., Lai, V., Gong, G., Yancopoulos, G. D., … Rinn, J. L. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *ELife*, *2*, e01749. https://doi.org/10.7554/eLife.01749

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., & Proudfoot, N. J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Molecular Cell*, *65*(1), 25–38. https://doi.org/10.1016/j.molcel.2016.11.029

Schmidt, K., Joyce, C. E., Buquicchio, F., Brown, A., Ritz, J., Distel, R. J., Yoon, C. H., & Novina, C. D. (2016). The lncRNA SLNCR1 Mediates Melanoma Invasion through a Conserved SRA1-like Region. *Cell Reports*, *15*(9), 2025–2037. https://doi.org/10.1016/j.celrep.2016.04.018

Schmidt, K., Weidmann, C. A., Hilimire, T. A., Yee, E., Hatfield, B. M., Schneekloth, J. S., Weeks, K. M., & Novina, C. D. (2020). Targeting the Oncogenic Long Non-coding RNA SLNCR1 by Blocking Its Sequence-Specific Binding to the Androgen Receptor. *Cell Reports*, *30*(2), 541–554. https://doi.org/10.1016/j.celrep.2019.12.011

Schmidt, N., Lareau, C. A., Keshishian, H., Ganskih, S., Schneider, C., Hennig, T., Melanson, R., Werner, S., Wei, Y., Zimmer, M., Ade, J., Kirschner, L., Zielinski, S., Dölken, L., Lander, E. S., Caliskan, N., Fischer, U., Vogel, J., Carr, S. A., … Munschauer, M. (2021). The SARS-CoV-2 RNA–protein interactome in infected human cells. *Nature Microbiology*, *6*(3), 339–353. https://doi.org/10.1038/s41564-020-00846-z

Schor, I. E., Bussotti, G., Maleš, M., Forneris, M., Viales, R. R., Enright, A. J., & Furlong, E. E. M. (2018). Non-coding RNA Expression, Function, and Variation during Drosophila Embryogenesis. *Current Biology*, *28*(22), 3547-3561.e9. https://doi.org/10.1016/j.cub.2018.09.026

Schreyer, A. M., & Blundell, T. (2012). USRCAT: Real-time ultrafast shape recognition with pharmacophoric constraints. *Journal of Cheminformatics*, *4*(11). https://doi.org/10.1186/1758-2946-4-27

Schubert, M. S., Thommandru, B., Woodley, J., Turk, R., Yan, S., Kurgan, G., McNeill, M. S., & Rettig, G. R. (2021). Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-98965-y

Schwartz, S., Mumbach, M. R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G. G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., Sanjana, N. E., Freinkman, E., Pacold, M. E., Satija, R., Mikkelsen, T. S., Hacohen, N., Zhang, F., Carr, S. A., Lander, E. S., & Regev, A. (2014). Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Reports*, *8*(1), 284–296. https://doi.org/10.1016/j.celrep.2014.05.048

Seemann, S. E., Mirza, A. H., Hansen, C., Bang-Berthelsen, C. H., Garde, C., Christensen-Dalsgaard, M., Torarinsson, E., Yao, Z., Workman, C. T., Pociot, F., Nielsen, H., Tommerup, N., Ruzzo, W. L., & Gorodkin, J. (2017). The identification and functional annotation of RNA structures conserved in vertebrates. *Genome Research*, *27*(8), 1371–1383. https://doi.org/10.1101/gr.208652

Sehgal, P., Mathew, S., Sivadas, A., Ray, A., Tanwar, J., Vishwakarma, S., Ranjan, G., Shamsudheen, K. v, Bhoyar, R. C., Pateria, A., Leonard, E., Lalwani, M., Vats, A., Pappuru, R. R., Tyagi, M., Jakati, S., Sengupta, S., B K, B., Chakrabarti, S., … Sivasubbu, S. (2021). LncRNA VEAL2 regulates PRKCB2 to modulate endothelial permeability in diabetic retinopathy . *The EMBO Journal*, *40*(15), e107134. https://doi.org/10.15252/embj.2020107134

Setten, R. L., Rossi, J. J., & Han, S. ping. (2019). The current state and future directions of RNAi-based therapeutics. *Nature Reviews Drug Discovery*, *18*(6), 421–446. https://doi.org/10.1038/s41573-019-0017-4

Seufert, L., Benzing, T., Ignarski, M., & Müller, R. U. (2022). RNA-binding proteins and their role in kidney disease. *Nature Reviews Nephrology*, *18*(3), 153–170. https://doi.org/10.1038/s41581-021-00497-1

Shen, C., Sheng, Y., Zhu, A. C., Robinson, S., Jiang, X., Dong, L., Chen, H., Su, R., Yin, Z., Li, W., Deng, X., Chen, Y., Hu, Y. C., Weng, H., Huang, H., Prince, E., Cogle, C. R., Sun, M., Zhang, B., … Chen, J. (2020). RNA Demethylase ALKBH5 Selectively Promotes Tumorigenesis and Cancer Stem Cell Self-Renewal in Acute Myeloid Leukemia. *Cell Stem Cell*, *27*(1), 64–80. https://doi.org/10.1016/j.stem.2020.04.009

Shi, H., Wei, J., & He, C. (2019). Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Molecular Cell*, *74*(4), 640–650. https://doi.org/10.1016/j.molcel.2019.04.025

Shi, Y., Parag, S., Patel, R., Lui, A., Murr, M., Cai, J., & Patel, N. A. (2019). Stabilization of lncRNA GAS5 by a Small Molecule and Its Implications in Diabetic Adipocytes. *Cell Chemical Biology*, *26*(3), 319–330. https://doi.org/10.1016/j.chembiol.2018.11.012

Singh, G., Pratt, G., Yeo, G. W., & Moore, M. J. (2015). The clothes make the mRNA: Past and present trends in mRNP fashion. *Annual Review of Biochemistry*, *84*, 325–354. https://doi.org/10.1146/annurev-biochem-080111-092106

Sioud, M. (2011). Promises and challenges in developing RNAi as a research tool and therapy. *Methods in Molecular Biology* , *703*, 173–187. https://doi.org/10.1007/978-1-59745-248-9_12

Sivaramakrishnan, M., McCarthy, K. D., Campagne, S., Huber, S., Meier, S., Augustin, A., Heckel, T., Meistermann, H., Hug, M. N., Birrer, P., Moursy, A., Khawaja, S., Schmucki, R., Berntenis, N., Giroud, N., Golling, S., Tzouros, M., Banfai, B., Duran-Pacheco, G., … Metzger, F. (2017). Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. *Nature Communications*, *8*(1), 1476. https://doi.org/10.1038/s41467-017-01559-4

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

Smola, M. J., Christy, T. W., Inoue, K., Nicholson, C. O., Friedersdorf, M., Keene, J. D., Lee, D. M., Calabrese, J. M., & Weeks, K. M. (2016). SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(37), 10322–10327. https://doi.org/10.1073/pnas.1600008113

Snetkova, V., Pennacchio, L. A., Visel, A., & Dickel, D. E. (2022). Perfect and imperfect views of ultraconserved sequences. *Nature Reviews Genetics*, *23*(3), 182–194. https://doi.org/10.1038/s41576-021-00424-x

Spitale, R. C., Crisalli, P., Flynn, R. A., Torre, E. A., Kool, E. T., & Chang, H. Y. (2013). RNA SHAPE analysis in living cells. *Nature Chemical Biology*, *9*(1), 18–20. https://doi.org/10.1038/nchembio.1131

Sprague, D., Waters, S. A., Kirk, J. M., Wang, J. R., Samollow, P. B., Waters, P. D., & Calabrese, J. M. (2019). Nonlinear sequence similarity between the Xist and Rsx long noncoding RNAs suggests shared functions of tandem repeat domains. *RNA*, *25*(8), 1004–1019. https://doi.org/10.1261/rna

Stadler, P. F. (2010). Evolution of the Long Non-coding RNAs MALAT1 and MENβ/ε. In C. E. Ferreira, S. Miyano, & P. F. Stadler (Eds.), *Advances in Bioinformatics and Computational Biology. BSB 2010. Lecture Notes in Computer Science*. Springer. https://doi.org/10.1007/978-3-642-15060-9_1

Standaert, L., Adriaens, C., Radaelli, E., van Keymeulen, A., Blanpain, C., Hirose, T., Nakagawa, S., & Marine, J. C. (2014). The long noncoding RNA Neat1 is required for mammary gland development and lactation. *RNA*, *20*(12), 1844–1849. https://doi.org/10.1261/rna.047332.114

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

Stelzer, A. C., Frank, A. T., Kratz, J. D., Swanson, M. D., Gonzalez-Hernandez, M. J., Lee, J., Andricioaei, I., Markovitz, D. M., & Al-Hashimi, H. M. (2011). Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nature Chemical Biology*, *7*(8), 553–559. https://doi.org/10.1038/nchembio.596

Story, R. M., Li, H., & Abelson, J. N. (2001). Crystal structure of a DEAD box protein from the hyperthermophile Methanococcus jannaschii. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(4), 1465–1470. https://doi.org/10.1073/pnas.98.4.1465

Su, R., Dong, L., Li, Y., Gao, M., Han, L., Wunderlich, M., Deng, X., Li, H., Huang, Y., Gao, L., Li, C., Zhao, Z., Robinson, S., Tan, B., Qing, Y., Qin, X., Prince, E., Xie, J., Qin, H., … Chen, J. (2020). Targeting FTO Suppresses Cancer Stem Cell Maintenance and Immune Evasion. *Cancer Cell*, *38*(1), 79–96. https://doi.org/10.1016/j.ccell.2020.04.017

Swift, L. P., Rephaeli, A., Nudelman, A., Phillips, D. R., & Cutts, S. M. (2006). Doxorubicin-DNA adducts induce a non-topoisomerase II-mediated form of cell death. *Cancer Research*, *66*(9), 4863–4871. https://doi.org/10.1158/0008-5472.CAN-05-3410

Swinney, D. C. (2011). Molecular Mechanism of Action (MMoA) in Drug Discovery. In *Annual Reports in Medicinal Chemistry* (Vol. 46, pp. 301–317). Academic Press Inc. https://doi.org/10.1016/B978-0-12-386009-5.00009-6

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

Teeli, A. S., Łuczyńska, K., Haque, E., Gayas, M. A., Winiarczyk, D., & Taniguchi, H. (2021). Disruption of tumor suppressors hnf4α /hnf1α causes tumorigenesis in liver. *Cancers*, *13*(21), 5357. https://doi.org/10.3390/cancers13215357

The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

The GTEx Consortium. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776

Theimer, C. A., Blois, C. A., & Feigon, J. (2005). Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Molecular Cell*, *17*(5), 671–682. https://doi.org/10.1016/j.molcel.2005.01.017

Theler, D., Dominguez, C., Blatter, M., Boudet, J., & Allain, F. H. T. (2014). Solution structure of the YTH domain in complex with N6-methyladenosine RNA: A reader of methylated RNA. *Nucleic Acids Research*, *42*(22), 13911–13919. https://doi.org/10.1093/nar/gku1116

Tichon, A., Gil, N., Lubelsky, Y., Solomon, T. H., Lemze, D., Itzkovitz, S., Stern-Ginossar, N., & Ulitsky, I. (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nature Communications*, *7*(1), 12209. https://doi.org/10.1038/ncomms12209

Tilgner, H., Knowles, D. G., Johnson, R., Davis, C. A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T. R., & Guigó, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Research*, *22*(9), 1616–1625. https://doi.org/10.1101/gr.134445.111

Tiwari, N., Tiwari, V. K., Waldmeier, L., Balwierz, P. J., Arnold, P., Pachkov, M., Meyer-Schaller, N., Schübeler, D., vanNimwegen, E., & Christofori, G. (2013). Sox4 Is a Master Regulator of Epithelial-Mesenchymal Transition by Controlling Ezh2 Expression and Epigenetic Reprogramming. *Cancer Cell*, *23*(6), 768–783. https://doi.org/10.1016/j.ccr.2013.04.020

Toll-Riera, M., Bosch, N., Bellora, N., Castelo, R., Armengol, L., Estivill, X., & Mar Albà, M. (2009). Origin of primate orphan genes: A comparative genomics approach. *Molecular Biology and Evolution*, *26*(3), 603–612. https://doi.org/10.1093/molbev/msn281

Tupy, J. L., Bailey, A. M., Dailey, G., Evans-Holm, M., Siebel, C. W., Misra, S., Celniker, S. E., & Rubin, G. M. (2005). Identification of putative noncoding polyadenylated transcripts in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(15), 5495–5500. https://doi.org/10.1073/pnas.0501422102

Udd, B., & Krahe, R. (2012). The myotonic dystrophies: molecular, clinical, and therapeutic challenges. *Lancet Neurology*, *11*, 891–905. https://doi.org/10.1016/S1474-4422(12)70204-1

Ulitsky, I. (2016). Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs. *Nature Reviews Genetics*, *17*(10), 601–614. https://doi.org/10.1038/nrg.2016.85

Ulitsky, I., & Bartel, D. P. (2013). lincRNAs: Genomics, evolution, and mechanisms. In *Cell* (Vol. 154, Issue 1, pp. 26–46). Elsevier B.V. https://doi.org/10.1016/j.cell.2013.06.020

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H., & Bartel, D. P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*, *147*(7), 1537–1550. https://doi.org/10.1016/j.cell.2011.11.055

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

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

Ursu, A., Vézina-Dawod, S., & Disney, M. D. (2019). Methods to identify and optimize small molecules interacting with RNA (SMIRNAs). In *Drug Discovery Today* (Vol. 24, Issue 10, pp. 2002–2016). Elsevier Ltd. https://doi.org/10.1016/j.drudis.2019.06.019

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R., & Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. *Nature Reviews Genetics*, *19*(9), 535–548. https://doi.org/10.1038/s41576-018-0017-y

Uzonyi, A., Slobodin, B., & Schwartz, S. (2022). Exon-intron architecture determines mRNA stability by dictating m6A deposition. *Preprint at BioRxiv*. https://doi.org/10.1101/2022.06.29.498130

Vakirlis, N., Carvunis, A. R., & McLysaght, A. (2020). Synteny-based analyses indicate that sequence divergence is not the main source of orphan genes. *ELife*, *9*(e53500). https://doi.org/10.7554/eLife.53500

Valdez, B. C., Perlaky, L., & Henning, D. (2002). Expression, cellular localization, and enzymatic activities of RNA helicase II/Guβ. *Experimental Cell Research*, *276*(2), 249–263. https://doi.org/10.1006/excr.2002.5538

Vallot, C., & Rougeulle, C. (2013). Long non-coding RNAs and human X-chromosome regulation: A coat for the active X chromosome. *RNA Biology*, *10*(8), 1262–1265. https://doi.org/10.4161/rna.25802

van Oss, S. B., & Carvunis, A. R. (2019). De novo gene birth. *PLoS Genetics*, *15*(5), e1008160. https://doi.org/10.1371/journal.pgen.1008160

Velagapudi, S. P., Costales, M. G., Vummidi, B. R., Nakai, Y., Angelbello, A. J., Tran, T., Haniff, H. S., Matsumoto, Y., Wang, Z. F., Chatterjee, A. K., Childs-Disney, J. L., & Disney, M. D. (2018). Approved Anti-cancer Drugs Target Oncogenic Non-coding RNAs. *Cell Chemical Biology*, *25*(9), 1086-1094.e7. https://doi.org/10.1016/j.chembiol.2018.05.015

Venkatesan, A. M., Vyas, R., Gramann, A. K., Dresser, K., Gujja, S., Bhatnagar, S., Chhangawala, S., Gomes, C. B. F., Xi, H. S., Lian, C. G., Houvras, Y., Edwards, Y. J. K., Deng, A., Green, M., & Ceol, C. J. (2018). Ligand-activated BMP signaling inhibits cell differentiation and death to promote melanoma. *Journal of Clinical Investigation*, *128*(1), 294–308. https://doi.org/10.1172/JCI92513

Vu, L. P., Pickering, B. F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., Chou, T., Chow, A., Saletore, Y., Mackay, M., Schulman, J., Famulare, C., Patel, M., Klimek, V. M., Garrett-Bakelman, F. E., Melnick, A., Carroll, M., Mason, C. E., Jaffrey, S. R., & Kharas, M. G. (2017). The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nature Medicine*, *23*(11), 1369–1376. https://doi.org/10.1038/nm.4416

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

Wang, S., Drummond, M. L., Guerrero-Juarez, C. F., Tarapore, E., MacLean, A. L., Stabell, A. R., Wu, S. C., Gutierrez, G., That, B. T., Benavente, C. A., Nie, Q., & Atwood, S. X. (2020). Single cell transcriptomics of human epidermis identifies basal stem cell transition states. *Nature Communications*, *11*(1), 4239. https://doi.org/10.1038/s41467-020-18075-7

Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., Ren, B., Pan, T., & He, C. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, *505*(7481), 117–120. https://doi.org/10.1038/nature12730

Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., & He, C. (2015). N6-methyladenosine modulates messenger RNA translation efficiency. *Cell*, *161*(6), 1388–1399. https://doi.org/10.1016/j.cell.2015.05.014

Wang, Y., Yesselman, J. D., Zhang, Q., Kang, M., & Feigon, J. (2016). Structural conservation in the template/pseudoknot domain of vertebrate telomerase RNA from teleost fish to human. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(35), E5125–E5134. https://doi.org/10.1073/pnas.1607411113

Wang, Z., Li, Y., Wu, D., Yu, S., Wang, Y., & Leung Chan, F. (2020). Nuclear receptor HNF4α performs a tumor suppressor function in prostate cancer via its induction of p21-driven cellular senescence. *Oncogene*, *39*(7), 1572–1589. https://doi.org/10.1038/s41388-019-1080-3

Wang, Z., Zhang, X. J., Ji, Y. X., Zhang, P., Deng, K. Q., Gong, J., Ren, S., Wang, X., Chen, I., Wang, H., Gao, C., Yokota, T., Ang, Y. S., Li, S., Cass, A., Vondriska, T. M., Li, G., Deb, A., Srivastava, D., … Wang, Y. (2016). The long noncoding RNA Chaer defines an epigenetic checkpoint in cardiac hypertrophy. *Nature Medicine*, *22*(10), 1131–1139. https://doi.org/10.1038/nm.4179

Warner, K. D., Hajdin, C. E., & Weeks, K. M. (2018). Principles for targeting RNA with drug-like small molecules. *Nature Reviews Drug Discovery*, *17*(8), 547–558. https://doi.org/10.1038/nrd.2018.93

Washietl, S., Kellis, M., & Garber, M. (2014). Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Research*, *24*(4), 616–628. https://doi.org/10.1101/gr.165035.113

Wei, S., Chen, H., Dzakah, E. E., Yu, B., Wang, X., Fu, T., Li, J., Liu, L., Fang, S., Liu, W., & Shan, G. (2019). Systematic evaluation of C. elegans lincRNAs with CRISPR knockout mutants. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-018-1619-6

Wen, K., Yang, L., Xiong, T., Di, C., Ma, D., Wu, M., Xue, Z., Zhang, X., Long, L., Zhang, W., Zhang, J., Bi, X., Dai, J., Zhang, Q., Lu, Z. J., & Gao, G. (2016). Critical roles of long noncoding RNAs in Drosophila Spermatogenesis. *Genome Research*, *26*(9), 1233–1244. https://doi.org/10.1101/gr.199547.115

Wen, W., Pillai-Kastoori, L., Wilson, S. G., & Morris, A. C. (2015). Sox4 regulates choroid fissure closure by limiting Hedgehog signaling during ocular morphogenesis. *Developmental Biology*, *399*(1), 139–153. https://doi.org/10.1016/j.ydbio.2014.12.026

Weng, H., Huang, H., Wu, H., Qin, X., Zhao, B. S., Dong, L., Shi, H., Skibbe, J., Shen, C., Hu, C., Sheng, Y., Wang, Y., Wunderlich, M., Zhang, B., Dore, L. C., Su, R., Deng, X., Ferchen, K., Li, C., … Chen, J. (2018). METTL14 Inhibits Hematopoietic Stem/Progenitor Differentiation and Promotes Leukemogenesis via mRNA m6A Modification. *Cell Stem Cell*, *22*(2), 191-205.e9. https://doi.org/10.1016/j.stem.2017.11.016

Westermarck, J., Weiss, C., Saffrich, R., Kast, J., Musti, A.-M., Wessely, M., Ansorge, W., Séraphin, B., Wilm, M., Valdez, B. C., & Bohmann, D. (2002). The DEXD/H-box RNA helicase RHII/Gu is a co-factor for c-Jun-activated transcription. *The EMBO Journal*, *21*(3), 451–460. https://doi.org/10.1093/emboj/21.3.451.

Wilson, D. N., Schluenzen, F., Harms, J. M., Starosta, A. L., Connell, S. R., & Fucini, P. (2008). The oxazolidinone antibiotics perturb the ribosomal peptidyl-transferase center and effect tRNA positioning. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(36), 13339–13344. https://doi.org/10.1073/pnas.0804276105

Wilson, K. D., Ameen, M., Guo, H., Abilez, O. J., Tian, L., Mumbach, M. R., Diecke, S., Qin, X., Liu, Y., Yang, H., Ma, N., Gaddam, S., Cunningham, N. J., Gu, M., Neofytou, E., Prado, M., Hildebrandt, T. B., Karakikes, I., Chang, H. Y., & Wu, J. C. (2020). Endogenous Retrovirus-Derived lncRNA BANCR Promotes Cardiomyocyte Migration in Humans and Non-human Primates. *Developmental Cell*, *54*(6), 694–709. https://doi.org/10.1016/j.devcel.2020.07.006

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

Winkle, M., El-Daly, S. M., Fabbri, M., & Calin, G. A. (2021). Noncoding RNA therapeutics — challenges and potential solutions. *Nature Reviews Drug Discovery*, *20*(8), 629–651. https://doi.org/10.1038/s41573-021-00219-z

Wojtas, M. N., Pandey, R. R., Mendel, M., Homolka, D., Sachidanandam, R., & Pillai, R. S. (2017). Regulation of m6A Transcripts by the 3′→5′ RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline. *Molecular Cell*, *68*(2), 374-387.e12. https://doi.org/10.1016/j.molcel.2017.09.021

Wong, E. S., Zheng, D., Tan, S. Z., Bower, N. I., Garside, V., Vanwalleghem, G., Gaiti, F., Scott, E., Hogan, B. M., Kikuchi, K., McGlinn, E., Francois, M., & Degnan, B. M. (2020). Deep conservation of the enhancer regulatory code in animals. *Science*, *370*(6517), eaax8137. https://doi.org/10.1126/science.aax8137

Woo Nam, S., Clair, T., Campo, C. K., Young Lee, H., Liotta, L. A., & Stracke, M. L. (2000). Autotaxin (ATX), a potent tumor motogen, augments invasive and metastatic potential of ras-transformed cells. *Oncogene*, *19*, 241–247. https://doi.org/10.1038/sj.onc.1203263

Wouters, J., Kalender-Atak, Z., Minnoye, L., Spanier, K. I., de Waegeneer, M., Bravo González-Blas, C., Mauduit, D., Davie, K., Hulselmans, G., Najem, A., Dewaele, M., Pedri, D., Rambow, F., Makhzami, S., Christiaens, V., Ceyssens, F., Ghanem, G., Marine, J. C., Poovathingal, S., & Aerts, S. (2020). Robust gene expression programs underlie recurrent cell states and phenotype switching in melanoma. *Nature Cell Biology*, *22*(8), 986–998. https://doi.org/10.1038/s41556-020-0547-3

Wu, M., Xu, G., Han, C., Luan, P.-F., Xing, Y.-H., Nan, F., Yang, L.-Z., Huang, Y., Yang, Z.-H., Shan, L., Yang, L., Liu, J., & Chen, L.-L. (2021). lncRNA SLERT controls phase separation of FC/DFCs to facilitate Pol I transcription. *Science*, *373*(6554), 547–555. https://doi.org/10.1126/science.abf6582

Wu, M., Yang, L.-Z., & Chen, L.-L. (2021). Long noncoding RNA and protein abundance in lncRNPs. *RNA*, *27*(12), 1427–1440. https://doi.org/10.1261/rna

Wu, P. (2020). Inhibition of RNA-binding proteins with small molecules. *Nature Reviews Chemistry*, *4*(9), 441–458. https://doi.org/10.1038/s41570-020-0201-4

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

Xia, H., Zhong, C., Wu, X., Chen, J., Tao, B., Xia, X., Shi, M., Zhu, Z., Trudeau, V. L., & Hu, W. (2018). Mettl3 mutation disrupts gamete maturation and reduces fertility in zebrafish. *Genetics*, *208*(2), 729–743. https://doi.org/10.1534/genetics.117.300574

Xiao, W., Adhikari, S., Dahal, U., Chen, Y. S., Hao, Y. J., Sun, B. F., Sun, H. Y., Li, A., Ping, X. L., Lai, W. Y., Wang, X., Ma, H. L., Huang, C. M., Yang, Y., Huang, N., Jiang, G. bin, Wang, H. L., Zhou, Q., Wang, X. J., … Yang, Y. G. (2016). Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Molecular Cell*, *61*(4), 507–519. https://doi.org/10.1016/j.molcel.2016.01.012

Xie, C., Zhang, Y. E., Chen, J. Y., Liu, C. J., Zhou, W. Z., Li, Y., Zhang, M., Zhang, R., Wei, L., & Li, C. Y. (2012). Hominoid-Specific De Novo Protein-Coding Genes Originating from Long Non-Coding RNAs. *PLoS Genetics*, *8*(9), e1002942. https://doi.org/10.1371/journal.pgen.1002942

Xie, W., Chen, B., & Wong, J. (2021). Evolution of the market for mRNA technology. *Nature Reviews Drug Discovery*, *20*(10), 735–736. https://doi.org/10.1038/d41573-021-00147-y

Xing, Y. H., Yao, R. W., Zhang, Y., Guo, C. J., Jiang, S., Xu, G., Dong, R., Yang, L., & Chen, L. L. (2017). SLERT Regulates DDX21 Rings Associated with Pol I Transcription. *Cell*, *169*(4), 664-678.e16. https://doi.org/10.1016/j.cell.2017.04.011

Yang, Y., Declerck, N., Manival, X., Aymerich, S., & Kochoyan, M. (2002). Solution structure of the LicT-RNA antitermination complex: CAT clamping RAT. *The EMBO Journal*, *21*(8), 1987–1997. https://doi.org/10.1093/emboj/21.8.1987

Yankova, E., Blackaby, W., Albertella, M., Rak, J., de Braekeleer, E., Tsagkogeorga, G., Pilka, E. S., Aspris, D., Leggate, D., Hendrick, A. G., Webster, N. A., Andrews, B., Fosbeary, R., Guest, P., Irigoyen, N., Eleftheriou, M., Gozdecka, M., Dias, J. M. L., Bannister, A. J., … Kouzarides, T. (2021). Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature*, *593*(7860), 597–601. https://doi.org/10.1038/s41586-021-03536-w

Yao, N., Hesson, T., Cable, M., Hong, Z., Kwong, A., Le, H., & Weber, P. C. (1997). Structure of the hepatitis C virus RNA helicase domain. *Nature Structural Biology*, *4*(5), 463–467. https://doi.org/10.1038/nsb0697-463

Yao, R. W., Wang, Y., & Chen, L. L. (2019). Cellular functions of long noncoding RNAs. *Nature Cell Biology*, *21*(5), 542–551. https://doi.org/10.1038/s41556-019-0311-8

Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., Zhao, Y., Shen, B., Huang, X., Zhu, H., Orkin, S. H., & Shen, X. (2015). Opposing roles for the lncRNA haunt and its genomic locus in regulating HOXA gene activation during embryonic stem cell differentiation. *Cell Stem Cell*, *16*(5), 504–516. https://doi.org/10.1016/j.stem.2015.03.007

You, X., Ryu, M. J., Cho, E., Sang, Y., Damnernsawad, A., Zhou, Y., Liu, Y., Zhang, J., & Lee, Y. (2021). Embryonic Expression of NrasG12D Leads to Embryonic Lethality and Cardiac Defects. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/fcell.2021.633661

Zaccara, S., & Jaffrey, S. R. (2020). A Unified Model for the Function of YTHDF Proteins in Regulating m6A-Modified mRNA. *Cell*, *181*(7), 1582–1595. https://doi.org/10.1016/j.cell.2020.05.012

Zaccara, S., Ries, R. J., & Jaffrey, S. R. (2019). Reading, writing and erasing mRNA methylation. *Nature Reviews Molecular Cell Biology*, *20*(10), 608–624. https://doi.org/10.1038/s41580-019-0168-5

Zama, T., Aoki, R., Kamimoto, T., Inoue, K., Ikeda, Y., & Hagiwara, M. (2002). A novel dual specificity phosphatase SKRP1 interacts with the MAPK kinase MKK7 and inactivates the JNK MAPK pathway: Implication for the precise regulation of the particular MAPK pathway. *Journal of Biological Chemistry*, *277*(26), 23909–23918. https://doi.org/10.1074/jbc.M200837200

Zapp, M. L., Stern, S., & Green, M. R. (1993). Small Molecules That Selectively Block RNA Binding of HIV-1 Rev Protein Inhibit Rev Function and Viral Production. *Cell*, *74*, 969–976. https://doi.org/10.1016/0092-8674(93)90720-b

Zhang, B., Arun, G., Mao, Y. S., Lazar, Z., Hung, G., Bhattacharjee, G., Xiao, X., Booth, C. J., Wu, J., Zhang, C., & Spector, D. L. (2012). The lncRNA malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Reports*, *2*(1), 111–123. https://doi.org/10.1016/j.celrep.2012.06.003

Zhang, B., Mao, Y. S., Diermeier, S. D., Novikova, I. v., Nawrocki, E. P., Jones, T. A., Lazar, Z., Tung, C. S., Luo, W., Eddy, S. R., Sanbonmatsu, K. Y., & Spector, D. L. (2017). Identification and Characterization of a Class of MALAT1-like Genomic Loci. *Cell Reports*, *19*(8), 1723–1738. https://doi.org/10.1016/j.celrep.2017.05.006

Zhang Lab. (2013). *Zhang Lab CRISPR. Target Sequence Cloning Protocol*. https://media.addgene.org/cms/filer_public/6d/d8/6dd83407-3b07-47db-8adb-4fada30bde8a/zhang-lab-general-cloning-protocol-target-sequencing_1.pdf

Zhang, S., Hu, Z., Tanji, H., Jiang, S., Das, N., Li, J., Sakaniwa, K., Jin, J., Bian, Y., Ohto, U., Shimizu, T., & Yin, H. (2018). Small-molecule inhibition of TLR8 through stabilization of its resting state. *Nature Chemical Biology*, *14*(1), 58–64. https://doi.org/10.1038/nchembio.2518

Zhao, B. S., Wang, X., Beadell, A. v., Lu, Z., Shi, H., Kuuspalu, A., Ho, R. K., & He, C. (2017). M6 A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature*, *542*(7642), 475–478. https://doi.org/10.1038/nature21355

Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., Vågbø, C. B., Shi, Y., Wang, W. L., Song, S. H., Lu, Z., Bosmans, R. P. G., Dai, Q., Hao, Y. J., Yang, X., Zhao, W. M., Tong, W. M., Wang, X. J., Bogdan, F., … He, C. (2013). ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Molecular Cell*, *49*(1), 18–29. https://doi.org/10.1016/j.molcel.2012.10.015

Zheng, S., Chen, Y., Donahue, C. P., Wolfe, M. S., & Varani, G. (2009). Structural Basis for Stabilization of the Tau Pre-mRNA Splicing Regulatory Element by Novantrone (Mitoxantrone). *Chemistry and Biology*, *16*(5), 557–566. https://doi.org/10.1016/j.chembiol.2009.03.009

Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M., & Fray, R. G. (2008). MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell*, *20*(5), 1278–1288. https://doi.org/10.1105/tpc.108.058883

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S. R., & Qian, S. B. (2015). Dynamic m6 A mRNA methylation directs translational control of heat shock response. *Nature*, *526*(7574), 591–594. https://doi.org/10.1038/nature15377

Zhou, S., Li, T., Zhang, M., Chen, C., Gao, X., Zhang, C., Hu, C., Zuo, Q., Chen, G., & Li, B. (2021). Epigenetic modification cooperates with Zeb1 transcription factor to regulate Bmp4 to promote chicken PGCs formation. *Gene*, *794*, 145760. https://doi.org/10.1016/j.gene.2021.145760

Zhu, P., Wu, J., Wang, Y., Zhu, X., Lu, T., Liu, B., He, L., Ye, B., Wang, S., Meng, S., Fan, D., Wang, J., Yang, L., Qin, X., Du, Y., Li, C., He, L., Ren, W., Wu, X., … Fan, Z. (2018). LncGata6 maintains stemness of intestinal stem cells and promotes intestinal tumorigenesis. *Nature Cell Biology*, *20*(10), 1134–1144. https://doi.org/10.1038/s41556-018-0194-0

Zhu, T., Roundtree, I. A., Wang, P., Wang, X., Wang, L., Sun, C., Tian, Y., Li, J., He, C., & Xu, Y. (2014). Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Research*, *24*(12), 1493–1496. https://doi.org/10.1038/cr.2014.152

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

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

RÉSUMÉ

L'importance des interactions ARN-protéines dans la biologie du cancer a été de plus en plus reconnue au cours des dernières années. Les protéines de liaison à l'ARN (RBP) régulent la maturation, la stabilité et la localisation subcellulaire des transcrits d'ARN, et sont également les moteurs de la fonction des ARN non codants impliqués dans de multiples processus biologiques. En outre, les interactions ARN-RBP sont apparues comme des régulateurs de l'oncogenèse et donc comme des cibles thérapeutiques potentielles.

Le long ARN non codant (lncRNA) humain CASC15 a été signalé comme étant dérégulé dans le mélanome métastatique humain. Notre laboratoire a récemment identifié son orthologue chez le poisson zèbre, dont la séquence primaire est peu conservée au cours de l'évolution. À l'aide d'un modèle inductible de cancer de la peau chez le poisson zèbre, qui correspond étroitement au processus humain, nous avons démontré que *CASC15* chez le poisson zèbre atténue l'initiation et la progression du mélanome. Ainsi, l'expression de *CASC15* humain chez le poisson zèbre mutant pour ce lncRNA permet le sauvetage du phénotype de progression du mélanome. Nous supposons que chez les deux orthologues, les mêmes interactions ARN-RBPs expliquent leur conservation fonctionnelle. Dans mon premier projet, j'ai étudié le mécanisme d'action moléculaire de *CASC15* dans une lignée cellulaire humaine. Comme pour le poisson zèbre, la délétion de *CASC15* dans les cellules de mélanome humain accélère leur migration. De plus, en combinant différentes approches moléculaires et biochimiques, j'ai identifié un ensemble d'interacteurs protéiques commun chez les orthologues de *CASC15* qui conduisent à leur conservation fonctionnelle. Ces résultats mettent en lumière la relation entre le mécanisme d'action et la compatibilité fonctionnelle entre les lncRNA de différentes espèces animales divergentes au niveau de leur séquence.

En parallèle, j'ai également étudié YTHDF2, une RBP impliquée dans de multiples cancers humains. Cette protéine reconnaît une marque de méthylation présente dans des transcrits d'ARN spécifiques qui régule leur stabilité. La délétion de YTHDF2 empêche l'apparition et le développement de la leucémie aiguë myéloïde tout en préservant l'hématopoïèse normale in vivo. En raison de cette spécificité cancéreuse, YTHDF2 est apparu comme une cible thérapeutique unique. Cependant, l'identification de petites molécules qui ciblent les interactions ARN-RBPs dans les cellules est actuellement limitée par l'absence de méthodes de criblage non biaisées. Dans mon deuxième projet, j'ai adapté notre technologie d'interaction ARN-RBPs à une plateforme de criblage de médicaments. J'ai identifié plusieurs composés approuvés par la FDA qui inhibent spécifiquement la liaison de YTHDF2 à ses ARN cibles et j'ai établi la procédure pour contrôler la spécificité des résultats identifiés. Afin de trouver de nouvelles molécules, j'ai appliqué un crible computationnel basé sur les ligands identifiés ainsi que des fragments de ces derniers. L'ensemble de ce travail démontre que la protéine YTHDF2 peut être traitée par des médicaments et corrobore la fiabilité de notre méthode pour l'identification d'inhibiteurs ciblant des interactions ARN-RBPs spécifiques.

En conclusion, j'ai établi des modèles in vivo, des systèmes cellulaires et des outils moléculaires robustes, menant à l'identification et au ciblage d'interactions fonctionnelles ARN-RBPs ayant des rôles cruciaux dans les cancers humains.

MOTS CLÉS

Interaction ARN-protéine, long ARN non-codants, mélanome, découverte de médicaments

ABSTRACT

The importance of RNA-protein interactions in cancer biology has been increasingly acknowledged in the past years. RNA-binding proteins (RBPs) regulate the maturation, stability, and subcellular localization of RNA transcripts. RBPs are also the drivers of the function of long non-coding RNAs implicated in multiple biological processes. In addition, RNA-RBP interactions have arisen as regulators of tumorigenesis and therefore as potential therapeutic targets.

The human long non-coding RNA (lncRNA) *CASC15* has been reported to be dysregulated in human metastatic melanoma. Our laboratory recently identified its zebrafish ortholog, whose primary sequence is poorly conserved throughout evolution. Using an inducible zebrafish skin cancer model that closely parallels the human process, we demonstrated that the zebrafish *casc15* attenuates melanoma initiation and progression. Importantly, the expression of human *CASC15* in the zebrafish lncRNA mutant rescued the melanoma progression phenotype. We hypothesized that the same RBP interactions in both orthologous transcripts underlie their functional conservation. In my first project, I investigated the molecular mechanism of action of *CASC15* in a human cell line. Similar to zebrafish, the depletion of CASC15 in human melanoma cells mutant increases their migration behavior. In addition, by combining different molecular and biochemical approaches, I identified a common set of functional protein interactors that drive the functional conservation of *CASC15* orthologs. These findings shed light on the relationship between mechanism of action and functional compatibility between lncRNA from different animal species divergent in sequence.

In parallel, I also studied YTHDF2, an important RBP implicated in multiple human cancers. This protein recognizes an abundant methylation mark present in specific RNA transcripts and regulates their stability. YTHDF2 knockout compromises leukemic stem cells while preserving normal hematopoiesis in vivo. Because of this cancer-specificity property, YTHDF2 has emerged as a unique therapeutic target. However, the identification of small molecules that target RNA-protein interactions in cell is currently limited by the absence of unbiased scalable screening methods. In my second project, I have adapted our RNA-protein interaction technology to a drug screening platform. I identified several FDA-approved compounds that specifically inhibit YTHDF2 binding to its target RNAs and established the pipeline to control for the specificity of the identified hits. In order to find novel small molecule inhibitors, I expanded the collection of tested drugs by following a ligand-based computational screen as well as including fragments of the hit compounds. In conclusion, this work demonstrates that YTHDF2 protein is druggable and corroborates the reliability of our method to be used for the identification of inhibitors targeting specific RNA-protein interactions.

KEYWORDS

RNA-protein interactions, lncRNA, melanoma, drug discovery