



HAL
open science

Structural, functional and evolutionary characterisation of the long non-coding RNA MEG3

Eleni Anastasakou

► **To cite this version:**

Eleni Anastasakou. Structural, functional and evolutionary characterisation of the long non-coding RNA MEG3. Structural Biology [q-bio.BM]. Université Grenoble Alpes [2020-..], 2021. English. NNT : 2021GRALV048 . tel-03793810

HAL Id: tel-03793810

<https://theses.hal.science/tel-03793810>

Submitted on 2 Oct 2022

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

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



THESIS / THÈSE

To obtain the title of / Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Discipline / Spécialité : **Biologie Structurale et Nanobiologie**

Arrêté ministériel : 25 mai 2016

Presented by / Présentée par

ELENI ANASTASAKOU

Thesis supervisor / Thèse dirigée par **Dr. Marco Marcia**

Co-supervisor / co-encadrée par **Dr. Isabel Chillon**

Thesis prepared at / Thèse préparée au sein du **European Molecular Biology Laboratory (EMBL), Grenoble Outstation**
in / dans **l'École Doctorale de Chimie et Sciences du Vivant**

Structural, functional and evolutionary characterisation of the long non-coding RNA MEG3

Caractérisation structurelle, fonctionnelle et évolutive de l'ARN long non-codant MEG3

Public defense on / Thèse soutenue publiquement le
21 Septembre 2021

Jury members / Devant le jury composé de:

Dr. Robert Feil Group Leader, IGMM, Montpellier, France.	Reviewer / Rapporteur
Dr. Alena Shkumatava Group Leader, Institut Curie, Paris, France.	Reviewer / Rapporteur
Dr. Maria Garcia Alai Group Leader, EMBL-Hamburg, Germany.	Examiner / Examinateur
Dr. Montserrat Soler Lopez Group Leader, University of Grenoble Alpes, France.	Examiner / Examinateur
Dr. André Verdel Group Leader, University of Grenoble Alpes, France.	President / Président



Structural, functional and evolutionary
characterisation of the long non-coding
RNA *MEG3*

Eleni Anastasakou

Université Grenoble Alpes

and

European Molecular Biology Laboratory

PhD supervisor: Dr. Marco Marcia

Co-supervisor: Dr. Isabel Chillón

A thesis submitted for the degree of

Doctor of Philosophy

Structural biology and nanobiology

July 2021

Declaration

I, Eleni Anastasakou confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

In this thesis I describe my PhD work on the long non-coding RNA maternally expressed gene 3 (*MEG3*). The main aim of my project was to understand the molecular mechanism of *MEG3* in regulating p53-dependent gene expression. Through detailed experimental analysis of the secondary structure of three human *MEG3* splicing isoforms, characterisation of common tertiary structural elements and use of a systematic *in cellulo* functional luciferase reporter assay, I contributed to establishing a structure-to-function axis for the mechanism of *MEG3* function. Furthermore, I studied the evolution of *MEG3* in mammals and I provide theoretical and experimental evidence for the conservation of a functional mechanism of *MEG3* relying on its structural features.

Résumé en français

Dans cette thèse, je décris mon travail de doctorat sur le long ARN non codant, le gène 3 exprimé par la mère (*MEG3*). L'objectif principal de mon projet était de comprendre le mécanisme moléculaire de *MEG3* dans la régulation de l'expression génique dépendante de p53. Grâce à une analyse expérimentale détaillée de la structure secondaire de trois isoformes d'épissage de *MEG3* humain, à la caractérisation d'éléments structuraux tertiaires communs et à l'utilisation d'un test rapporteur de luciférase fonctionnel *in cellulo* systématique, j'ai contribué à établir un axe structure-fonction pour le mécanisme de la fonction de *MEG3*. De plus, j'ai étudié l'évolution de *MEG3* chez les mammifères et j'ai fourni des preuves théoriques et expérimentales de la conservation d'un mécanisme fonctionnel de *MEG3* reposant sur ses caractéristiques structurelles.

Acknowledgement

There many people I wish to thank and too few words that seem suitable for the purpose. Bringing my PhD to completion was an immense task which I wouldn't have even dreamt of accomplishing had it not been for the support and help of those around me.

I want to thank Marco Marcia for giving me the opportunity to embark on this journey. I also want to thank Tina Uroda, Isabel Chillón and all the other past and present members of the Marcia lab for their input, advice and company during my PhD. I would like to express my gratitude to everyone else who contributed to my PhD work; my thesis advisory committee members, our collaborators at the different PSB platforms, in Italy and at the EBI. I also wish to thank my thesis jury members for their patience and careful assessment of my work. Finally, none of this work would have been possible without the tireless work of our lab support staff and facility managers at EMBL Grenoble.

I am ever thankful to my previous supervisors Jernej Ule and Cristina Militti who inspired me to do a PhD in the first place and whom I will always consider as my mentors and scientific role models.

Words fail me to express my gratitude to everyone who helped me keep my sanity during those past four years. Daniel, my family in Greece, Sally and David, thank you for everything! I also feel I should thank Eva Frickel for being so understanding and allowing Daniel to visit me in France so often! To all my wonderful friends, Moritz, Wanda, Cecilia and Théo, Alice, Jasmine, Linnea, Anna, Sunny, Antonia, thank you so much for being there for me!

Table of Contents

Abstract	3
Résumé en français	4
Acknowledgement	5
Table of Contents	6
Table of figures	9
List of tables	11
Abbreviations	12
Chapter 1. Introduction	15
1.1 The functional non-coding genome	15
1.2 Long non-coding RNAs	17
1.2.1 Functions of long non-coding RNAs	18
1.2.2 Clinical implications of long non-coding RNAs	32
1.2.3 The role of RNA structure in their functional mechanisms	34
1.2.4 Evolution of long non-coding RNAs	36
1.3 Maternally Expressed Gene 3	39
1.3.1 Introduction to Maternally Expressed Gene 3	39
1.3.2 Roles and mechanisms of <i>MEG3</i> in embryonic development	41
1.3.3 Roles of <i>MEG3</i> in the regulation of cell cycle and its clinical implications	42
1.3.4 Current state of research on <i>MEG3</i>	44
1.4 The p53 pathway	46
1.4.1 The major cell-cycle regulator: p53.....	46
1.4.2 p53 response elements	48
1.4.3 Interactions of p53 with RNA.....	50
1.4.4 Conservation of p53 protein and of p53 pathway across mammals ..	52
Chapter 2. Materials & Methods	53
2.1 Cell culture and treatments	53
2.1.1 Cell culture materials	53
2.1.2 Cell culture	54
2.1.3 RNA extraction and real time quantitative PCR (RT-qPCR)	54
2.2 Molecular biology	54
2.2.1 Materials for molecular biology techniques.....	54
2.2.2 DNA and RNA Agarose electrophoresis.....	55
2.2.3 Sequence- and ligation-independent cloning (SLIC).....	56
2.2.4 Quick-change polymerase chain reaction cloning.....	56
2.2.5 Restriction cloning	65
2.2.6 Transformation and plasmid purification	65
2.3 RNA probing	66
2.3.1 Materials for RNA probing	66
2.3.2 Fluorescent dye coupling to primers used for RNA probing	69
2.3.3 Generation of sequencing ladders	70
2.3.4 Preparation of NTPs	71
2.3.5 <i>In vitro</i> RNA transcription	73
2.3.6 Native purification of <i>in vitro</i> transcribed RNA	73
2.3.7 Selective Hydroxyl-alkylation and primer extension (SHAPE) using fragment length analysis (FLA)	74

2.3.8 Hydroxyl-radical foot-printing (HRF)	75
2.3.9 Terbium (III) RNA probing	78
2.3.10 Reverse transcription of RNA probing assays	78
2.3.11 Sequencing of RNA fragments from RNA probing assays	79
2.3.12 Fragment length analysis (FLA) of probing assays	79
2.3.13 RNA secondary structure modelling	79
2.4 Biochemical and biophysical methods	80
2.4.1 Sedimentation velocity analytical ultracentrifugation (AUC)	80
2.5 Dual Luciferase Assay	81
2.5.1 Cell seeding	81
2.5.2 Transfections	82
2.5.3 Luciferase Assay	82
2.5.4 Luciferase assay signal normalisation	82
2.6 Computational alignments	83
2.6.1 RNA sequence alignments	83
2.6.2 RNA structural alignments	83
2.7 Crystallisation trials	84
Chapter 3. Results	85
3.1 Identification of the active core of the lncRNA <i>MEG3</i>	85
3.1.1 Résumé en français	85
3.1.2 Chapter Introduction	86
3.1.3 <i>In vitro</i> SHAPE probing of <i>MEG3</i> variants v1, v3 and v9	88
3.1.4 Cell-based functional assays reveal the <i>MEG3</i> active core	96
3.1.5 Chapter Discussion	106
3.2 Linking the active core of <i>MEG3</i> to the tertiary structural organisation of the lncRNA	110
3.2.1 Résumé en français	110
3.2.2 Chapter introduction	111
3.2.3 The H11-TR pseudoknot affects the overall folding profile of <i>MEG3</i>	112
3.2.4 Chapter discussion	118
3.3 Towards a high-resolution structure of the <i>MEG3</i> core	120
3.3.1 Résumé en français	120
3.3.2 Chapter introduction	121
3.3.3 Terbium foot-printing reveals divalent ion binding sites essential for tertiary structure formation.	122
3.3.4 Manipulating the H11-TR pseudoknot motif can tweak the activity profile of <i>MEG3</i> in the p53 pathway	125
3.3.5 Chapter discussion	128
3.4 Investigation of the evolutionary conservation of the <i>MEG3</i> functional mechanism	131
3.4.1 Résumé en français	131
3.4.2 Chapter introduction	132
3.4.3 Sequence and structure alignment provide evidence supporting that the <i>MEG3</i> functional mechanism is conserved through evolution ...	133
3.4.4 Structural characterisation of mouse <i>Meg3</i> provides unbiased proof of structural conservation of <i>MEG3</i> lncRNA in mammals	138

3.4.5 Functional characterisation of murine <i>Meg3</i> in p53-dependent gene activation.....	146
3.4.6 Chapter discussion	150
Chapter 4.Discussion.....	153
4.1 Résumé en français.....	153
4.2 Summary, discussion and future outlook	154
Chapter 5.Appendix.....	161
5.1 Appendix 1: Human <i>MEG3</i> splicing isoforms	161
5.2 Appendix 2: Reporters plasmids with p53 response elements	162
5.3 Appendix 3: Sequence alignment of human and murine p53 proteins.....	163
5.4 Appendix 4: Publications.....	164
Reference List	174

Table of figures

Figure 1:1 The central dogma of molecular biology	15
Figure 1:2 Modes of lncRNA activities.....	19
Figure 1:3 Transcriptional interference prevents simultaneous <i>Airn</i> lncRNA and <i>Igfr2</i> gene transcription	20
Figure 1:4 Mechanism of expression regulation and imprinting of the lncRNA <i>Airn</i> and its neighbouring protein-coding genes.....	22
Figure 1:5 Regulation of transcription by <i>Evl2-Dlx2</i>	24
Figure 1:6 Mechanism of <i>ZEB2</i> mRNA splicing regulation by <i>NAT Zeb2</i>	26
Figure 1:7 The micro RNA biogenesis and function pathway	28
Figure 1:8 Mechanism of <i>PHO2</i> mRNA metabolism regulation by the lncRNA <i>IPS1</i>	31
Figure 1:9 Regulation of the <i>p27</i> mRNA by 3' UTR structures and protein binding	35
Figure 1:10 Graphic representation of the <i>Dlk1/Gtl2</i> locus in mice	40
Figure 1:11 Schematic representation of the p53 pathway.....	47
Figure 1:12 Three graphic representations of p53.....	48
Figure 1:13 Recognition of p53 response elements	50
Figure 2:1 Workflow of RNA probing assays	72
Figure 2:2 SHAPE chemistry and downstream analysis.....	75
Figure 2:3 HRF chemistry	77
Figure 2:4 Tb(III) probing chemistry	78
Figure 3:1 <i>MEG3</i> gene structure and known splicing isoforms.....	87
Figure 3:2 Functional characterisation of human <i>MEG3</i> splicing isoforms in the activation of p53-dependent gene expression	88
Figure 3:3 Purification of <i>MEG3v3</i>	89
Figure 3:4 Correlation of SHAPE reactivity values across <i>MEG3v3</i> between n = 3 independent replicates	90
Figure 3:5 Schematic representation of the secondary structure maps of human <i>MEG3</i> variants v1, v3 and v9	91
Figure 3:6 1M7 SHAPE reactivities of individual nucleotides in <i>MEG3v1</i> , v3 and v9	94
Figure 3:7 Close up on 1M7 reactivities of individual nucleotides in <i>MEG3v1</i> , v3 and v9 at the region helix H11	95
Figure 3:8 Luciferase reporter assay.....	97
Figure 3:9 Activity assay for human <i>MEG3</i> isoforms v1, v3 and v9 and their mutants targeting helix H11.....	98
Figure 3:10 Illustration showing the potential interactions forming between each one of the tandem repeats and helix H11.....	99
Figure 3:11 Activity assays with <i>MEG3</i> mutants targeting the H11-TR pseudoknot interaction	100
Figure 3:12 Activity assays on <i>MEG3v1</i> mutants targeting the H11-TRs interaction by altering the TRs sequence.....	103
Figure 3:13 Activity assay screening the effect of pseudoknot formation by each tandem repeat on different p53-responsive promoters	104
Figure 3:14 RT-qPCR quantification of transfected <i>MEG3</i> mutant-expressing plasmids compared to <i>MEG3v1</i> wildtype control.....	105
Figure 3:15 Correlation of 3 replicates of <i>MEG3v1</i> hydroxyl-radical probing at 17.5 mM MgCl ₂	113

Figure 3:16 Hydroxyl-radical foot-print of individual nucleotides across wild-type <i>MEG3v1</i> and the helix H11 loop poly(A) mutant.....	116
Figure 3:17 Hydroxyl-radical reactivity values of helix H11.....	117
Figure 3:18 Correlation of 3 replicas of the Terbium probing experiment	123
Figure 3:19 Purification of domains D2-D3 and D2-D3 with Δ H8	126
Figure 3:20 Purification of <i>MEG3</i> domains D2-D3 Δ H8 with forced pseudoknot formation on TR1 receptor	127
Figure 3:21 Sequence and structural alignment of 41 mammalian <i>MEG3</i> analogues mapped on the secondary structure model of human <i>MEG3v1</i>	134
Figure 3:22 Genomic location of bovine <i>MEG3</i> and analysis of its sequence.....	137
Figure 3:23 Exon organisation and splicing isoforms of mouse <i>Meg3</i>	139
Figure 3:24 Purification and initial biophysical characterisation of <i>mMeg3v3</i>	140
Figure 3:25 Sedimentation-velocity analytical ultracentrifugation of <i>mMeg3</i>	141
Figure 3:26 Summary of <i>mMeg3v3</i> <i>in vitro</i> SHAPE probing	142
Figure 3:27 Secondary structural map of <i>mMeg3v3</i> highlighting its domain organisation	143
Figure 3:28 1M7 reactivities of individual nucleotides in the helix H11/H15 region	145
Figure 3:29 Activity assay for human <i>MEG3v1</i> and mouse <i>Meg3</i> in human HCT116 cells.....	146
Figure 3:30 Activity assay using HCT116 p53 ^{-/-} cells.....	147
Figure 3:31 RT-qPCR quantification of endogenously expressed <i>mMeg3</i> in the murine cell line Hepa1-6	148
Figure 3:32 Activity assay of human <i>MEG3</i> and mouse <i>Meg3v3</i> on mouse Hepa1-6 cells.....	148
Figure 3:33 Activity assay of individual exon deletion constructs of murine <i>Meg3v3</i>	149
Figure 4:1 Cartoon representation of the proposed mechanism for <i>MEG3</i> -driven gene expression regulation in trans.....	160
Figure 5:1 Representation of p53-responsive reporter plasmids	162
Figure 5:2 Sequence alignment of human and murine p53 proteins.....	163

List of tables

Table 2:1 List of materials used in cell culture	53
Table 2:2 List of materials used for molecular cloning	54
Table 2:3: List of selected primers used for SLIC	56
Table 2:4 List of primers used for quick-change site-directed mutagenesis.....	57
Table 2:5 List of essential instruments used in RNA probing experiments	66
Table 2:6 List of materials used in RNA probing experiments	66
Table 2:7 List of primers used in RNA probing experiments.....	69
Table 2:8 PCR cyclor program for ladder generation reaction.....	70
Table 2:9 List of materials used for analytical ultracentrifugation.....	80
Table 2:10 List of materials used for luciferase reporter assays.....	81
Table 2:11 Software used for structure and sequence-based alignments	83
Table 2:12 Materials used for crystallisation trials	84
Table 3:1 Nucleotides involved in coordinating divalent ions.....	124
Table 3:2 Tandem repeats identified in different mammals	135
Table 5:1 Annotations of human MEG3 splicing isoforms	161
Table 5:2 List of genomic p53 response elements used in this work	162

Abbreviations

Full word	Abbreviation
Capture Hybridization Analysis of RNA Targets	CHART
5-Carboxyfluorescein, succinimidyl ester	5-FAM
6-Carboxy-4', 5'-dichloro-2', 7'-dimethoxyfluorescein	6-JOE
Analytical Ultracentrifugation	AUC
Antisense Noncoding RNA in the INK4 locus	<i>ANRIL</i>
Argonaute 2	Ago2
Atomic force microscopy	AFM
Cancer lncRNA Census database	CLC
Chromatin Immunoprecipitation	ChIP
Competing endogenous RNA	ceRNA
Cryo-electron microscopy	cryo-EM
C-terminal domain	CTD
Deletion	Δ
Delta-like	<i>Dlk</i>
Deoxyribonucleic acid	DNA
DNA-binding domain	DBD
double-stranded RNA-binding protein	dsRBP
Dulbecco's Modified Eagle Medium	DMEM
Endogenous siRNA	endo-siRNA
Epithelial-mesenchymal transition	EMT
Exogenous siRNA	exo-siRNA
Exon	E
Fragment length analysis	FLA
Gamma-Aminobutyric acid	GABA
Gene trap locus 2	<i>Gtl2</i>
Green fluorescent protein	GFP
Helix	H
Hepatocellular carcinoma	HCC
high throughput crystallisation	HTX
HOX transcript antisense RNA	<i>HOTAIR</i>
HOXA cluster antisense RNA 3	<i>HOXAAS3</i>
HOXB cluster antisense RNA 1	<i>HOXBAS1</i>
Human <i>MEG3</i>	<i>MEG3</i>
Hydrodynamic radius	Rh
Hydroxyl-radical foot-printing	HRF
Individual-nucleotide cross-linking and immunoprecipitation	iCLIP
Insulin-like growth factor 2 receptor	Igfr
Internal ribosome entry site	IRES

Kilobase	kb
light activated structural examination of RNA by high throughout sequencing	LASER-seq
Locked nucleic acid	LNA
Long intergenic non-coding RNA	lincRNA
Long non-coding RNA	lncRNA
Maternally expressed gene 3	<i>MEG3</i>
Menin 1	MEN1
Messenger RNA	mRNA
Metastasis Associated Lung Adenocarcinoma Transcript 1	<i>MALAT1</i>
Methyl-CpG binding protein 2	MECP2
Micro RNA	miRNA
Mouse embryonic fibroblasts	MEF
Murine <i>Meg3</i>	<i>mMeg3</i>
MS2 coat proteins	MS2-CP
Natural antisense transcript	NAT
Non-coding RNA	ncRNA
Non-coding RNA Activated by DNA damage	<i>NORAD</i>
<i>NORAD</i> repeat unit	NRU
N-terminal domain	NTD
Nuclear magnetic resonance	NMR
Nucleotides	nt
Online Mendelian Inheritance in Man	OMIM
Oxygen radicals	•OH
Piwi interacting RNA	piRNA
Polycomb repressive complex	PRC
psoralen analysis of RNA interactions and structures	PARIS
Pumilio-1	PUM1
Pumilio recognition sequence	PRE
Rapid amplification of cDNA ends	RACE
Real time quantitative PCR	RT-qPCR
Response element	RE
Ribonucleic acid	RNA
Ribonucleoprotein	RNP
Ribosomal RNA	rRNA
RNA immunoprecipitations	RIP
RNA sequencing	RNA-seq
RNA-binding protein	RBP
RNA-induced silencing complex	RISC
Sedimentation velocity – analytical ultracentrifugation	SV-AUC
Selective hydroxyl alkylation and primer extension	SHAPE
Sequence- and ligation-independent cloning	SLIC

Short interfering RNA	siRNA
Size exclusion chromatography and multi-angle light scattering	SEC-MALS
Small angle X-ray scattering	SAXS
Small nuclear RNA	snRNA
Small nucleolar RNA	snoRNA
Tandem repeat	TR
Target RNA-directed miRNA degradation	TDMD
Transfer RNA	tRNA
Transformation-related protein 53	TRP53/ p53
Transforming growth factor- β	TGF- β
Uniparental disomy of chromosome 14	UPD14
Untranslated region	UTR
Variant X	vX
Wildtype	WT
X chromosome inactive specific transcript	<i>XIST</i>

Chapter 1. Introduction

1.1 The functional non-coding genome

Four billion years ago, life existed in a completely different form than it does today. According to “*the RNA world hypothesis*,” the ancestors of modern cells are believed to have encoded their genetic information and performed their life processes by a single type of molecule: ribonucleic acid (RNA) (Higgs & Lehman, 2015). With the passage of time and the slow process of evolution, cells gradually exchanged RNA for DNA (deoxyribonucleic acid) to carry their genetic information and proteins to perform cellular functions. Thus, we moved from the era of the RNA world to this of the “*central dogma of molecular biology*,” dictating that “*DNA makes RNA, RNA makes protein*” (Crick, 1970) (Figure 1:1). The original interpretation of the central dogma suggested that RNA devolved to become an intermediate in the process of proteins production. Using the knowledge currently at hand, I will present arguments for a different interpretation of the central dogma highlighting the central position of RNA in the simple sentence “*DNA makes RNA, RNA makes protein.*”

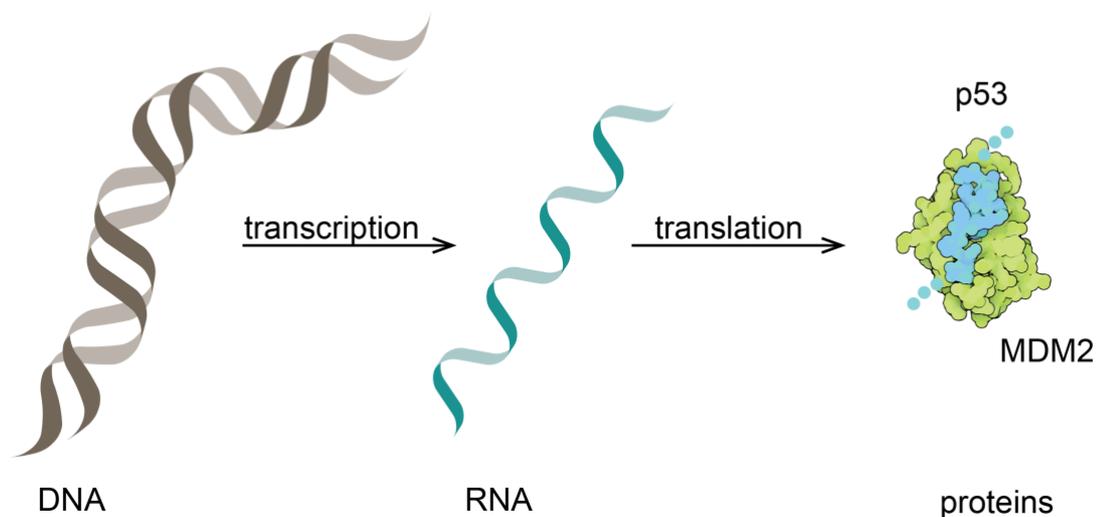


Figure 1:1 The central dogma of molecular biology

Schematic representation of the central dogma of molecular biology “DNA makes RNA, RNA makes protein.” The proteins depicted here are a portion of the tumour suppressor p53 and its regulator MDM2, this image has been sourced from the PDB molecule of the month, June 2019 entry.

Let us start by considering the second part of the central dogma: “*RNA makes protein.*” This could simply mean that the transcribed messenger RNA (mRNA) is translated to produce polypeptides (Figure 1:1). Inarguably, mRNA is a major class

of RNA with an essential function for cells. However, I would like to shift the focus on the actual *process* of protein-production. Translation of mRNA into proteins occurs in the ribosomes. Ribosomes are the molecular machines that translate mRNA into proteins. They are large assemblies of proteins and ribosomal RNA (rRNA) (Ramakrishnan, 2002). The rRNA component of ribosomes occupies roughly two thirds of their total mass and has no protein-coding potential. The role of rRNA is to catalyse the formation of peptide bonds between consecutive amino acids that will eventually form a complete peptide once the process of translation is finished. In this example “*RNA makes protein*” but the contribution of RNA in the making of protein is functional and dynamic in addition to its role as a messenger. In conclusion, this example of a catalytic RNA links the idea of the RNA world to modern cells and demonstrates that, whilst proteins are undoubtedly essential for all processes in the life of the cell, the role of RNA can be more than an intermediate for protein production.

There are more non-coding RNAs other than the well-known rRNAs. In fact, the majority of the genome does not code for any protein but is nevertheless transcribed resulting in more non-coding RNAs than mRNAs inside the cell (ENCODE Project Consortium, 2004; Kapranov *et al*, 2007). The vast number of non-coding RNAs and their functional diversity has led to efforts to classify them. Thus, ncRNAs are split into *families* based on their sequence identities and into *classes* on the basis of common structural and/or functional features (Nieselt & Herbig, 2013). In terms of ncRNA function, to date, there are at least eight known classes of non-coding RNA some of which are well characterised and others which are less so. In addition to the aforementioned rRNAs, some examples of functional non-coding RNAs include, but are not limited to, transfer RNAs (tRNAs) which are also involved in protein production by carrying amino acids to the translation complex (Hoagland *et al*, 1958), small nuclear RNAs (snRNA) comprising the catalytic cores of spliceosomes, small nucleolar RNAs (snoRNA) which are involved in various RNA processing events including but not limited to RNA modifications such as methylation and pseudouridylation (Kiss, 2002), piwi-interacting RNAs (piRNA) which silence transposons in germ cells (reviewed in Ozata *et al*, 2018), short interfering RNAs (siRNAs) which can be either exogenous or endogenous (exo-siRNA and endo-siRNA respectively) (reviewed in Okamura & Lai 2008) and micro RNAs (miRNA)

(Eddy, 2001). Both miRNAs and siRNAs are involved in post-transcriptional regulation of mRNA. (reviewed in Okamura & Lai, 2008). An additional class of non-coding RNA which I would like to mention, is long non-coding RNA (lncRNA). lncRNAs have much more diverse functions and are not as well-characterised as the other classes of non-coding RNAs mentioned above. The focus of my work is on the lncRNA *maternally expressed gene 3 (MEG3)*. A general understanding of lncRNAs is critical for the understanding of this thesis and therefore the following chapter will focus on lncRNAs.

Closing, I hope that in this chapter I have achieved to convince you that the role of RNA in the cell is not limited to its function as a messenger during protein production. From a closer look at the role of different RNA molecules during translation it becomes apparent that RNA has a central role in the life processes of the cell thus linking the RNA world hypothesis with modern cells. In addition to the well-known examples of functional non-coding RNA in the cell, there are other classes of non-coding RNAs that are less well-understood despite their biological importance. As I will discuss in the following section, new tools and technologies have enabled scientists to learn more about those under-studied and somewhat underappreciated classes of transcripts, some of which I will present in the following section.

1.2 Long non-coding RNAs

Long non-coding RNAs (lncRNAs) are transcripts longer than 200 nucleotides (nt) with no protein-coding potential and which perform some sort of regulatory function (Quinn & Chang, 2016). Arguably, this is a rather vague definition and as such, it is bound to have exceptions. There are thousands of lncRNAs inside the cell not all of which have a defined function assigned to them (Kirk *et al*, 2018; Iyer *et al*, 2015). Also of note, the 200 nt cut-off was established as a result of the available protocols for the exclusion of small RNAs from RNA extraction experiments and as a result lncRNAs can range from 200 nts to more than 100 kilobases (Guttman *et al*, 2009; Quinn & Chang, 2016). Similar to protein-coding mRNAs, lncRNAs are transcribed by RNA polymerase II (RPol II) and can be spliced, 3'-polyadenylated and 5'-capped (Quinn & Chang, 2016). Some lncRNAs are made up of different exons that are spliced together to form the final lncRNA and as I will discuss later, some spliced lncRNAs have alternative splicing isoforms that may modulate and/or regulate their

functionality. At first lncRNAs were conceived as transcriptional noise. However, from studies that carefully analysed full-length cDNAs from C25BL/6J mice it became apparent that lncRNAs are tightly regulated and stabilised (Okazaki *et al*, 2002). The regulation and stabilisation of lncRNAs requires the consumption of valuable energy from the cell implying that lncRNAs must have a function, otherwise lncRNA production and maintenance would have devolved over time.

1.2.1 Functions of long non-coding RNAs

There are several additional lines of evidence supporting the functionality of lncRNAs. Firstly, the transcription of lncRNAs appears to be temporally regulated and to follow highly specific tissue localisation patterns (Cabili *et al*, 2011). Secondly, lncRNA gene promoters are associated with regulatory proteins (ENCODE Project Consortium, 2012) and appear to be under higher evolutionary pressure than non-functional sequences such as introns (Hezroni *et al*, 2015). The functions of lncRNAs are as diverse as different lncRNAs themselves and whilst the mechanisms of some are well-studied and thoroughly-understood, some others remain elusive (Figure 1:2). Furthermore, the diversity of lncRNAs in combination with the novelty of their discovery as a functional class of molecule in the cell, makes it very difficult to establish a single experimental strategy to study them and multiplexing is not currently possible for the detailed characterisation of lncRNAs. In this section I will discuss some known functional modes of lncRNAs giving examples for each class as well as the methodologies used to study them.

A common theme for all known functions of lncRNAs is the regulation of gene expression (Figure 1:2). For clarity I have divided this chapter in examples of lncRNAs acting in gene regulation at the stages of (1) chromatin modification, (2) regulation of the process of transcription and (3) splicing, or at the post-transcriptional level by (4) regulating mRNA degradation. This list of examples of lncRNA functional modes presented here is not exhaustive. However, in my opinion it elegantly highlights the diversity of regulatory mechanisms adopted by different lncRNAs and in extent the dynamic nature of this class of non-coding transcripts.

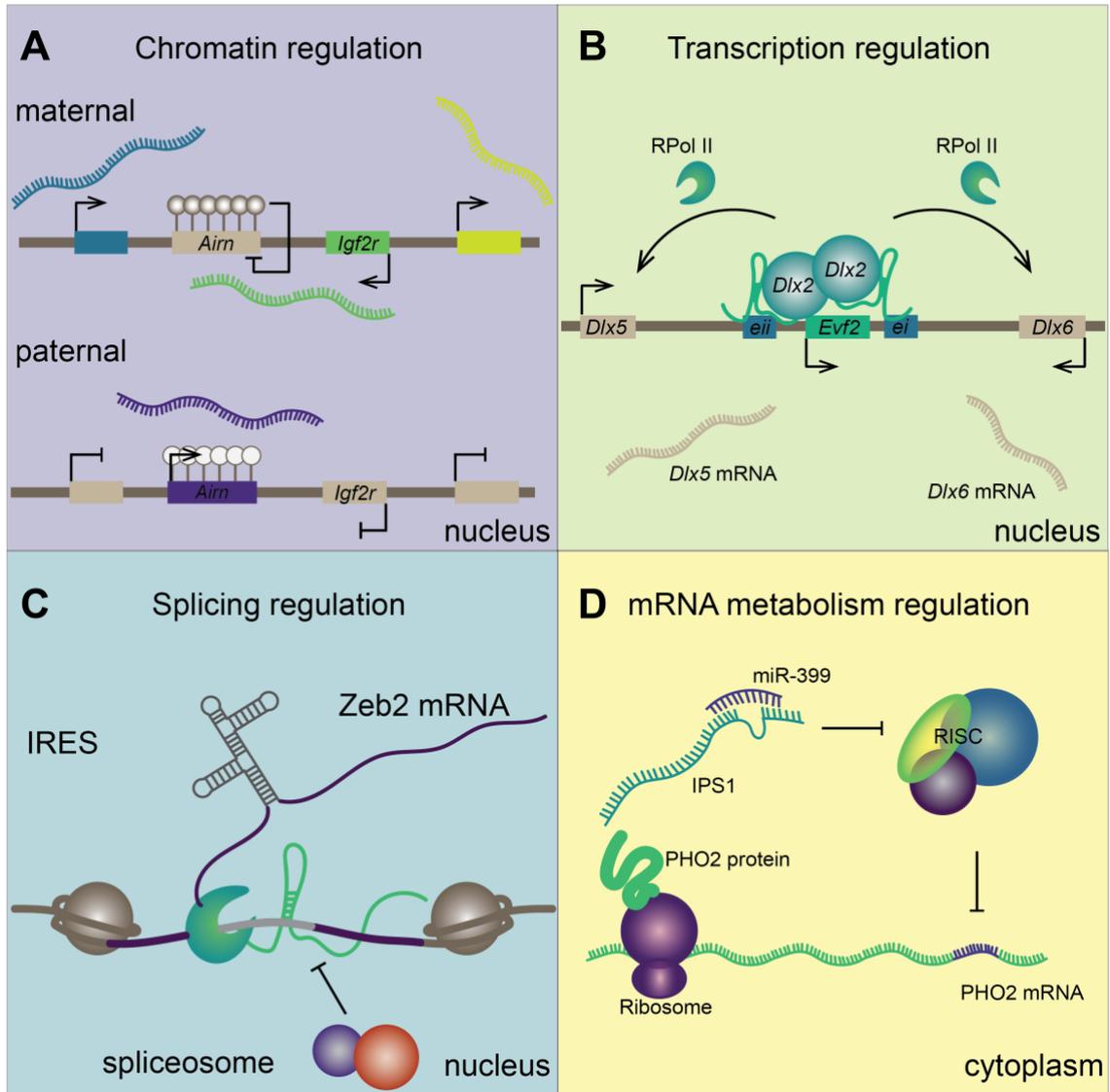


Figure 1:2 Modes of lncRNA activities

Figure depicting different modes of lncRNA activities: (A) lncRNA acting as chromatin regulator by dictating DNA methylation events. (B) lncRNA acting as transcriptional regulator by directing the binding of transcriptional activator protein on their target DNA sequence. (C) lncRNA acting as splicing regulator by preventing spliceosomal association leading to intron inclusion. (D) lncRNA acting as a regulator of mRNA stability by sequestering miRNAs and thus preventing mRNA degradation by RNA-interference. Examples shown in panels (A)-(C) occur inside the cell nucleus whilst example (D) shows a cytoplasmic process. All examples shown in this figure will be explained in detail in the next chapters.

1.2.1.1 Long non-coding RNAs acting at the chromatin level

Initially, lncRNAs were thought to affect transcription by indirectly allowing access of the transcriptional machinery to the locus where they are encoded. For example, transcription of the lncRNA *Airn* (also sometimes known as *Air*) was found to be essential for the epigenetic silencing of the paternal copy of the insulin-like growth factor 2 receptor (*Igf2r*) locus in mouse embryonic stem cells (Latos et al, 2012). The genes coding for *Igf2r* protein and *Airn* lncRNA respectively are found on opposite strands of the same DNA locus. The mRNA of *Igf2r* is maternally expressed whilst the lncRNA *Airn* is paternally expressed. (Reviewed in Quinn & Chang, 2016). *Airn* transcription inhibits *Igf2r* expression by antisense transcriptional interference (Figure 1:3) (Quinn & Chang, 2016). This mechanism of function suggests that the two RPol II complexes moving in opposite directions on the same strand of DNA will inevitably collide with one another and thus either the gene on the sense or the antisense strand only can possibly be transcribed from one allele.

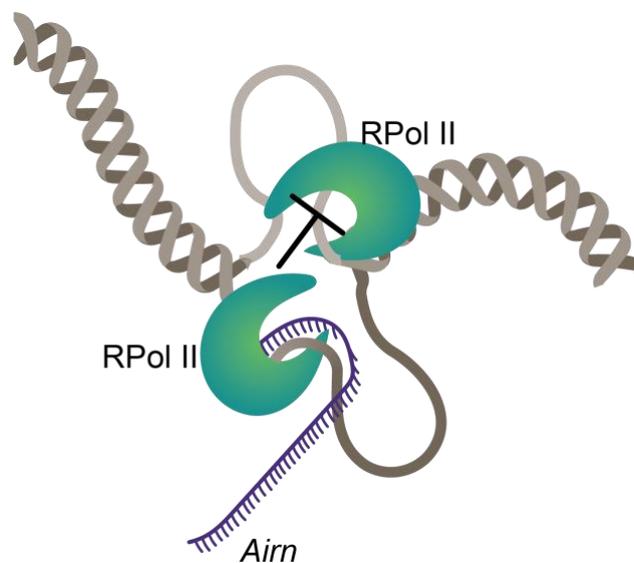


Figure 1:3 Transcriptional interference prevents simultaneous *Airn* lncRNA and *Igfr2* gene transcription

Cartoon representation of the competition of two RNA polymerase II complexes (RPol II in green) moving in opposite directions on the same DNA strand (brown). In this example successful transcription of the lncRNA *Airn* (purple) inhibits the expression of the gene on the opposite stand.

Historically, this was one of the first hypotheses about the function of lncRNAs and in my opinion it does not convey the full potential of lncRNAs as it does not highlight the active role of lncRNAs in regulatory pathways. Perhaps the first example of a

lncRNA to be thoroughly functionally characterised was the X-inactive specific transcript (*XIST*), which is essential for the process of random X-chromosome inactivation in mammalian females to compensate for the absence of a second copy of the X chromosome in males (Brown *et al*, 1991; Borsani *et al*, 1991; Penny *et al*, 1996). After its original identification in the early 1990's, more than 20 years of subsequent research revealed that *XIST* acts at the RNA level to orchestrate the mechanism of random X-chromosome inactivation revealing a complicated, many-layered mechanism of action (reviewed in Augui *et al*, 2011). Briefly, *XIST* coats the X chromosome to be inactivated upon stem cell differentiation, initiating silencing marked by asynchronous replication. Coating of an X chromosome by *XIST* prevents activating factors from associating with the chromosome and recruits repressive proteins. Then long-term chromatin modifications take over propagating chromosomal silencing (reviewed in Avner & Heard, 2001). A full and detailed description of the mechanism of *XIST* would constitute a thesis in its own and therefore I will refrain from discussing the mechanism of action in my thesis. What I would like to highlight with the example of *XIST* is the idea that lncRNAs can act on the genome and regulate chromatin state in a dynamic manner.

Interestingly, the earlier example of *Airn* also extends to the chromatin level. Chromatin conformation capture studies revealed that *Airn* is also responsible for the imprinting of the entire *Igf2r* gene cluster including *Slc22a3*, found 234 kilobases upstream of *Airn* (Andergassen *et al*, 2019) (Figure 1:4). It was shown that *Airn* regulates the *Igf2r* locus by directing the interaction of *Polycomb* group chromatin remodelling complexes with the *Slc22a3* gene (Andergassen *et al*, 2019; Nagano *et al*, 2008). Through this mechanism *Airn* lncRNA acts as a long-term gene expression regulator by facilitating the chromatin modifications which can be carried on through epigenetic memory mechanisms (Figure 1:4).

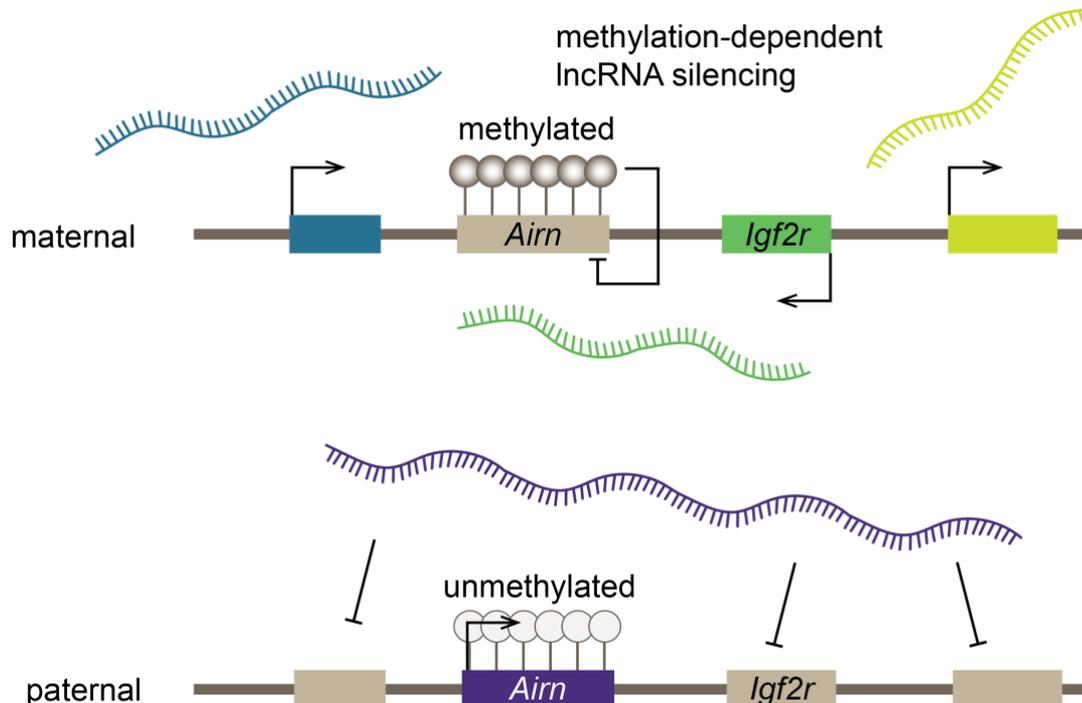


Figure 1:4 Mechanism of expression regulation and imprinting of the lncRNA *Airn* and its neighbouring protein-coding genes.

The lncRNA *Airn* is expressed exclusively from the paternal allele. In the genome, the *Airn* gene is flanked by protein-coding genes which are maternally expressed. Expression of *Airn* is necessary for paternal imprinting of its own locus *in cis*. Figure adapted from (Quinn & Chang, 2016).

Interactions of lncRNAs with the *Polycomb* group proteins are amongst the better-understood examples of lncRNA functions and therefore there are many noteworthy examples for this type of functional mechanism. For instance, the lncRNA HOX transcript antisense RNA (*HOTAIR*) which is transcribed from the developmentally important *HOXC* locus, interacts with the *Polycomb* Repressive Complex 2 (PRC2) and regulates the *HOXD* locus *in trans* (Rinn et al, 2007; Davidovich et al, 2013). Through its interactions with the PRC2 complex, *HOTAIR* lncRNA dictates which genomic regions will be epigenetically silenced via histone methylation. There are 230 more non-coding RNAs derived from the *HOXD* locus. However, interaction experiments have shown that of all those RNAs only *HOTAIR* forms stable complexes with PRC2 supporting that the interaction between *HOTAIR* and the PRC2 is highly specific (Rinn et al, 2007). The regulation of *HOX* gene loci is a very delicate and complex process in which the protein-coding genes of each of the four *HOX* loci (A to D) exhibit unique temporal and spatial expression patterns which curiously match their genomic position within their locus (Kmita & Duboule, 2003). This effect is termed collinearity and can be observed in flies (Bae et al, 2002),

humans (Bernstein et al, 2005) and possibly mice (Bernstein et al, 2005; Schorderet & Duboule, 2011). The other *HOX* loci also encode antisense long non-coding RNAs such as the *HOXA* cluster antisense RNA 3 (*HOXAAS3*) or *HOXB* cluster antisense RNA 1 (*HOXBAS1*) which are less well-studied than *HOTAIR*. Dysregulation of *HOTAIR* as well as *HOXAAS3* and *HOXBAS1* have been linked to disease phenotypes such as cancer (Carlevaro-Fita et al, 2020; Li et al, 2015). Closing, lncRNAs are not the only regulators of the *HOX* locus. Several *HOX* gene mRNAs can be regulated by miRNAs, for instance *Hoxa7*, *Hoxb8*, *Hoxc8* and *Hoxd8* mRNAs are all targets of miR-196 and their dysregulation by blockage of miR-196 could be linked to developmental defects (reviewed in Yekta et al, 2008). Like the lncRNA *HOTAIR*, miR-196 is also encoded within the *HOX* gene cluster it regulates (reviewed in Yekta et al, 2008). Other *HOX* locus-encoded miRNAs have also been found to preferentially regulate the expression of *HOX* mRNAs (reviewed in Yekta et al, 2008). Taken together, although the mechanism of *HOX* locus regulation is not entirely known, it has become apparent that regulatory non-coding RNAs are part of this mechanism and that different types of ncRNA could potentially act in concert or perhaps in opposition to regulate gene expression from the locus. For instance, lncRNAs such as *HOTAIR* regulate *HOX* genes at the transcriptional level while miRNAs exert a post-transcriptional level of regulation together building a complex network for the regulation of important developmental processes.

1.2.1.2 Long non-coding RNAs acting at the transcription level

In rats the 3.8 kilobase (kb) lncRNA *Evf2* can modulate the activity of homeodomain transcription factor proteins Dlx1 and Dlx2 *in trans* regulating the transcription of the important developmental genes of the *Dlx* family (Feng et al, 2006). *Evf2* lncRNA interacts directly with Dlx2 protein exclusively in neuronal cells to activate the *Dlx* genes but not other Dlx2-target genes. Specifically, *Evf2* guides Dlx2 to the *Dlx* intergenic enhancers *ei* and *eii* (Bond et al, 2009; Zerucha et al, 2000) (Figure 1:5). *Evf2* does not form a complex with any other homeodomain proteins in the *Dlx* gene activation pathway (Feng et al, 2006).

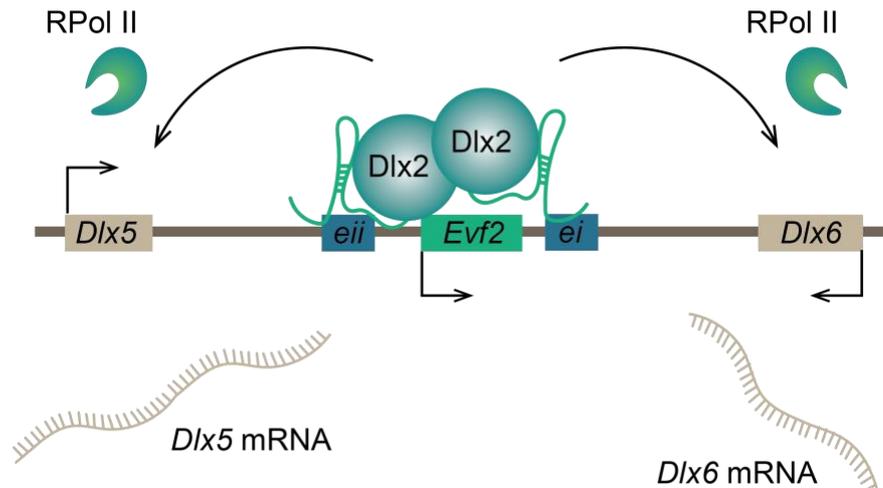


Figure 1:5 Regulation of transcription by *Evf2*-*Dlx2*

The lncRNA *Evf2* (green) interacts with the protein *Dlx2* (teal) to activate the *Dlx* genes (beige) by guiding *Dlx2* onto the *Dlx* intergenic enhancers *ei* and *eii*. Figure adapted from (Wang *et al*, 2011).

However, this is not the only function of *Evf2*. It also interacts with methyl-CpG binding protein 2 (MECP2) targeting it to the *Dlx* gene transcriptional control region where MECP2 acts as a transcriptional repressor (Bond *et al*, 2009). MECP2 acts in a regulatory loop inhibiting the expression of *Evf2* by antagonising its transcriptional activator *Dlx2* (Berghoff *et al*, 2013). The regulatory role of *Evf2* lncRNA is essential for the correct development of GABAergic interneurons (i.e. neurons activated by gamma-aminobutyric acid - GABA) (Bond *et al*, 2009) and thus aberrant *Evf2* expression can be linked to GABA-related pathologies including epilepsy, schizophrenia and autism (Nan *et al*, 1997; Xu *et al*, 2010; Cajigas *et al*, 2015). In fact, mice lacking *Evf2* are more susceptible to seizures upon GABA inhibition (Cajigas *et al*, 2018) and loss of MECP2 from the *Dlx* gene locus is linked to autism spectrum disorders in humans (Nan *et al*, 1997; Xu *et al*, 2010; Cajigas *et al*, 2015). Taken together, the lncRNA *Evf2* regulates the binding of proteins with both transcriptional activating (*Dlx2*) and inhibitory (MECP2) roles to the neurodevelopmentally important *Dlx* gene control region in a close regulatory loop with clinical implications. The interactions of *Evf2* with *Dlx1* extend to a ribonucleoprotein (RNP) network including chromatin remodelling proteins such as BRG1 and BAF170. This RNP complex enhances *Evf2*-*Dlx1* association with the *Dlx* gene locus (Cajigas *et al*, 2015). The BRG1 subunit of the RNP contacts the rest of the complex by its ATPase domain disabling its function as an ATP-dependent chromatin remodeller (Cajigas *et al*, 2015). Thus, the effect of *Evf2* on *Dlx* gene

expression self-propagates in the long-term by inhibiting BRG1-mediated chromatin repression allowing the continuous transcription from the *Dlx* locus. Finally, *Evf2* stabilises the topological organisation of chromosomes in interneurons of mice via interactions with cohesins (Cajigas *et al*, 2018). Thus, *Evf2* constitutes a global regulator for the *Dlx* gene family expression via three levels of control: (1) short-term transcriptional regulation, (2) long-term inhibition of chromatin repression and, (3) chromosome topology guidance. This example suggests how more than one mode of action can be connected with a single lncRNA highlighting the difficulty in delineating the functional mechanisms of different lncRNAs.

1.2.1.3 Long non-coding RNAs regulating splicing of protein-coding genes

lncRNAs can actively regulate gene expression both in short- and long-term. The simplest mechanism by which lncRNAs can exert regulatory function is by complementary base-pairing with coding transcripts (Mercer *et al*, 2009). This is a common mechanism of antisense lncRNAs which are transcribed from the complementary DNA strand of protein-coding genes. For instance, the natural antisense transcript (*NAT*) is derived from the antisense strand of the protein-coding gene *Zeb2* (hereafter referred to as *NAT Zeb2*). ZEB2 protein is a repressor of E-cadherin and its expression is induced upon initiation of epithelial-mesenchymal transition (EMT) in developing embryos (Comijn *et al*, 2001; Vandewalle *et al*, 2005). *NAT Zeb2* is a lncRNA involved in the regulation of ZEB2 protein in humans (Beltran *et al*, 2008) and in mice (Nelles *et al*, 2003). *NAT Zeb2* regulates the splicing of *Zeb2* mRNA by forming complementary base-pairs with the nascent mRNA during transcription. In the presence of *NAT Zeb2*, an internal ribosome entry site (IRES) contained near the 5'untranslated region (5'UTR) of the *Zeb2* gene is retained in mature *Zeb2* mRNA, facilitating efficient downstream ZEB2 protein translation. In the absence of *NAT Zeb2*, the IRES-containing intron is excised from the mRNA and thus whilst there is no evident difference in the available levels of *Zeb2* mRNA, the amount ZEB2 protein is significantly lower. ZEB2 protein is more profoundly expressed in tissues with low levels of E-cadherin for instance breast cells express more ZEB2 than epithelial cells (Beltran *et al*, 2008). The expression of *NAT Zeb2* follows the same pattern suggesting that the expression of *NAT Zeb2* depends on downstream effectors of the E-cadherin signalling pathway (Beltran *et al*, 2008).

Thus, *NAT Zeb2* plays an important role in regulating the expression of ZEB2 protein in response to developmental cues initiated by EMT.

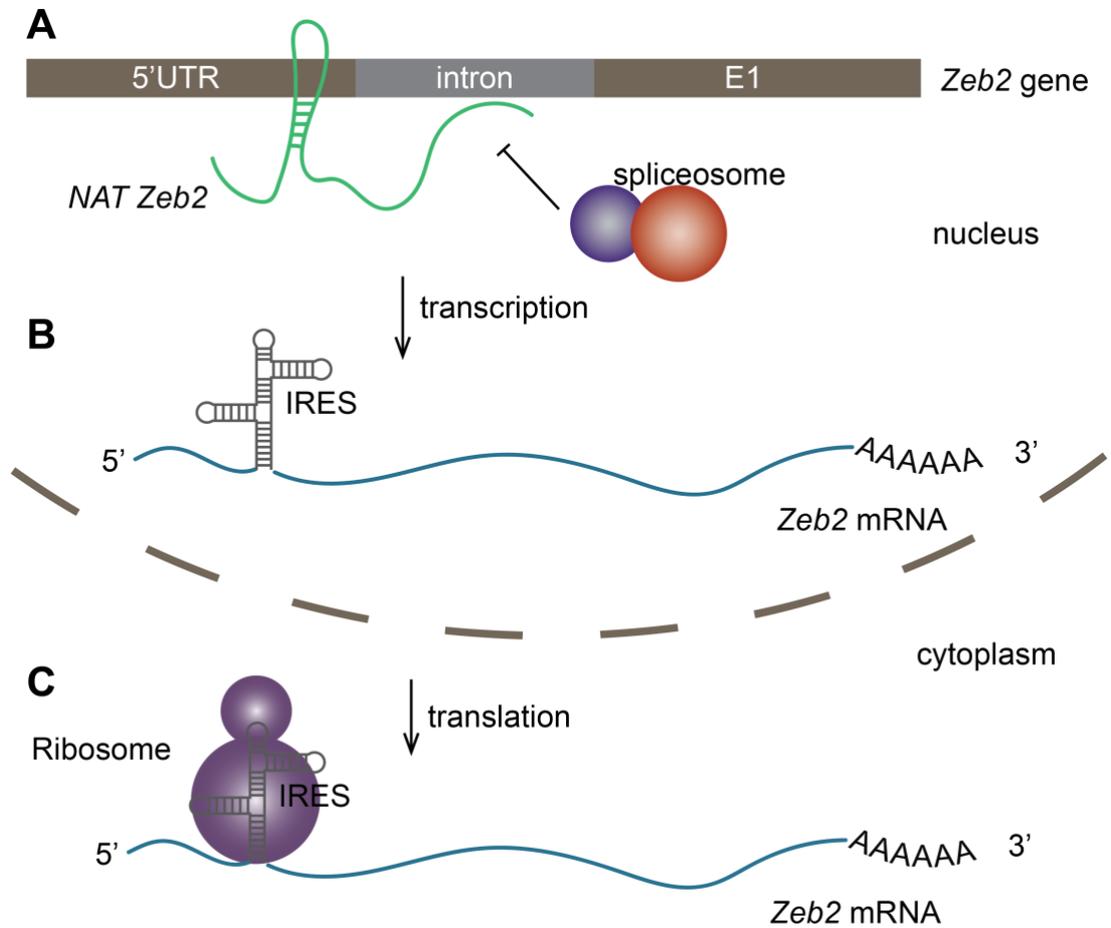


Figure 1:6 Mechanism of ZEB2 mRNA splicing regulation by NAT Zeb2

(A) *NAT Zeb2* (green) is an antisense lncRNA that can base-pair with the first intron (grey) of the *ZEB2* gene (brown) inhibiting it from being spliced out. (B) Transcription of *ZEB2* mRNA (blue) in the presence of *NAT Zeb2* leads to inclusion of the first intron which contains the sequence of an internal ribosomal entry site (IRES, grey). (C) In the presence of the IRES, *ZEB2* mRNA is recognised by the ribosomes in the cytoplasm and becomes translated.

1.2.1.4 Long non-coding RNAs acting at the post-transcriptional level

So far, I have covered some examples of lncRNAs as regulators of gene expression either by directly controlling transcription factors or by orchestrating the chromatin landscape. In other words, all aforementioned mechanisms of lncRNAs act at the transcriptional level. However, lncRNAs can also act post-transcriptionally. A well-known mechanism of ncRNA-mediated, post-transcriptional gene regulation is the mechanism of mRNA regulation by microRNAs (miRNA). Even though, miRNAs do not belong to the same class as lncRNAs, understanding the mechanism of miRNA-dependent post-transcriptional gene regulation is an important first step in understanding the ways by which lncRNAs can influence the post-transcriptional gene regulatory network. Therefore, I will briefly describe the main mechanism of activity of miRNAs. The description provided here is far from being detailed since I simply wish to introduce the context in which lncRNAs may act.

MicroRNAs regulate the stability of mRNAs post-transcriptionally by guiding the RNA-induced silencing complex (RISC) via base-pairing between the miRNA and its mRNA target (Figure 1:7). The RISC in turn either cleaves mRNAs at the site of complementarity with the miRNA or destabilises them by deadenylation, thereby preventing the production of proteins (Hannon, 2002). During the biogenesis and functional cycle of miRNA, a series of cleavage reactions deliver mature miRNAs to Argonaute (AGO) and Dicer proteins, together forming the RISC leading to subsequent mRNA degradation or temporary translational repression (Hannon, 2002) (Figure 1:7). In plants the miRNAs contain areas of high complementarity with the mRNAs and commonly lead to mRNA degradation (Bartel, 2009). In animals, miRNA target sites are most often found on the 3'UTS of mRNAs and show less extensive complementarity leading to translational repression (Bartel, 2009). Many and diverse miRNAs have been identified in humans which have been subsequently linked to the regulation of virtually all developmental and disease mechanisms (Bartel, 2018; Ha & Kim, 2014; Jonas & Izaurralde, 2015).

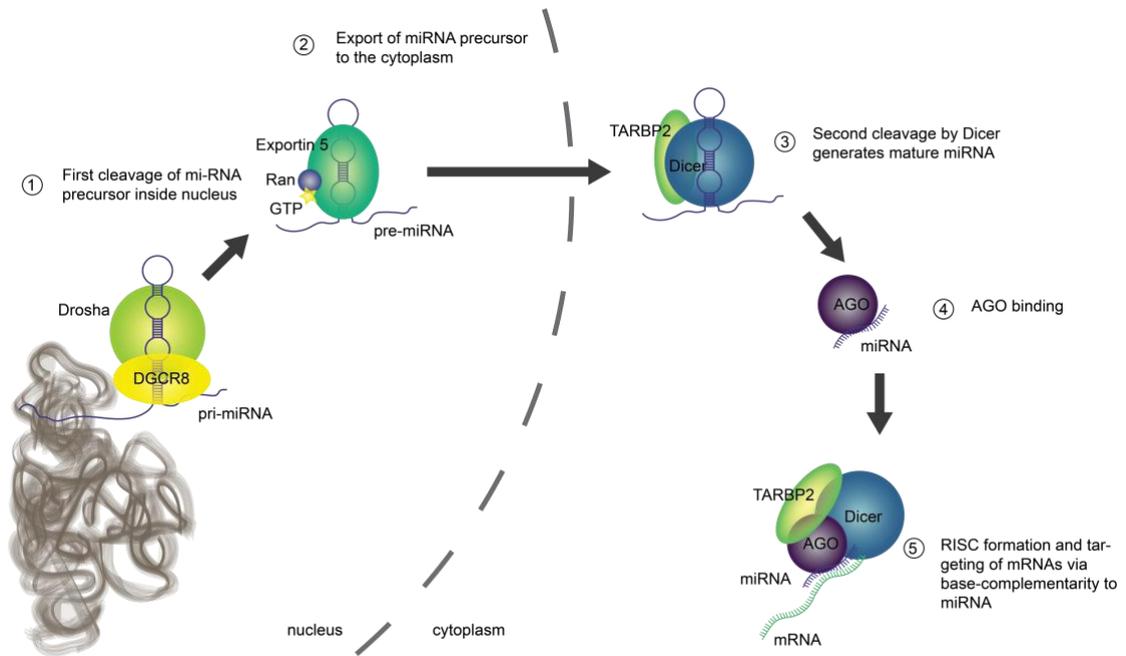


Figure 1:7 The micro RNA biogenesis and function pathway

Schematic representation of miRNA biogenesis and activity mechanism. In the nucleus primary micro RNA (pri-miRNA - purple) is generated, it is recognised by the microprocessor complex including proteins DGCR8 (yellow) and Drosha (light green) owing to a specific bulge in its hairpin structure. Association of the microprocessor leads to cleavage of the pri-miRNA into precursor miRNA (pre-miRNA) which is then exported into the cytosol via Exportin 5 (dark green). In the cytosol the pre-miRNA is further cleaved by Dicer (blue) to its mature form (miRNA). Mature miRNA is loaded onto Argonaute (AGO - purple) guiding the RISC onto its target mRNAs.

There are three main gateways through which lncRNAs can enter the post-transcriptional regulatory network of miRNAs. First, some lncRNAs can be downstream processed giving rise to small functional RNAs such as miRNA or snoRNAs (see chapter 1.1). For example, the miRNAs miR-657-3p and miR-675-5p are encoded within the first intron of the lncRNA *H19* and are important for skeletal growth during embryonic development (Dey *et al*, 2014; Cai & Cullen, 2007; Keniry *et al*, 2012). In this example, the role of *H19* is to control the abundance of miRNAs available thus adding a layer of regulation to the miRNA-target network (Keniry *et al*, 2012).

Some lncRNAs can form complexes with microRNAs (miRNA) leading to different downstream effects (reviewed in Ulitsky, 2018). One possible downstream effect after miRNA:lncRNA complex formation is the so-called competing endogenous RNA (ceRNA) effect. The idea of ceRNA activity is that a lncRNA with complementarity to a miRNA acts as a sponge for the respective miRNA, thereby

preventing target mRNA degradation. This is a rather controversial concept because the (commonly) nuclear localisation and relative low abundance of lncRNAs compared to the high transcriptome-wide number of miRNA binding sites for a single mRNA suggest that it is stoichiometrically unrealistic that a lncRNA could outcompete a target mRNA for miRNA binding.

Recently, a different mechanism for lncRNA-mediated miRNA regulation has been proposed. This is target RNA-directed miRNA degradation (TDMD) (reviewed in Ulitsky, 2018). According to TDMD, binding of lncRNAs with extensive complementarity onto miRNAs loaded on AGO can prevent mRNA degradation (Ameres *et al*, 2010). In *Arabidopsis thaliana*, the lncRNA *IPS1* regulates the stability of the miRNA *miR-399*. In turn, *miR-399* is a regulator of the mRNA coding for PHO2 which is a key regulatory protein required for inorganic phosphate homeostasis (Borah *et al*, 2018). Thus, *IPS1* is involved in fine-tuning the plant's response to phosphate starvation (Franco-Zorrilla *et al*, 2007) (Figure 1:8). Similar mechanisms have been described in the model organisms *Drosophila melanogaster* and in human HeLa cells for the regulation of miRNAs by their own target mRNAs through regions of extensive complementarity. Furthermore, in the past few years, there have been more reports of TDMD in mammals. Some prominent examples include (1) inhibition of miR-7 in the mammalian brain by binding to the lncRNA *Cyran* (Kleaveland *et al*, 2018), (2) degradation of miR-30b-5p and miR-30c-5p by the lncRNA *Serpin1* in mouse fibroblast cells (Ghini *et al*, 2018) and (3) the regulation of miR-29 by the 3'UTR of the *Nerp* mRNA in zebrafish and in mice (Bitetti *et al*, 2018). The latter example is also linked to behavioural phenotypes (Bitetti *et al*, 2018). Interestingly the NERP gene evolved from an ancestral lncRNA gene (Bitetti *et al*, 2018). Finally, the mechanism of TDMD is also used by some viral ncRNAs such as the *herpesvirus saimiri* HSUR1 ncRNA and the murine *cytomegalovirus* m169 mRNA which can both destabilise host miR-27 aiding viral replication (Cazalla *et al*, 2010; Marcinowski *et al*, 2012; Libri *et al*, 2012).

Conversely, some miRNAs have been found to regulate the function of lncRNAs (Yoon *et al*, 2013). For example, the lncRNA *HOTAIR*, whose function in chromatin remodelling was explained earlier (see chapter 1.2.1.1), has been found to be regulated by Argonaute 2 (Ago2) via base complementarity with the miRNA *Let-7i* (Yoon *et al*, 2013). In this system a secondary function of *HOTAIR* is revealed, where

it regulates protein stability at the post-translational level by recruiting ubiquitin ligases to ataxin-1 and snurportin-1 proteins. In turn, the stability of *HOTAIR* lncRNA is regulated by *Let-7i* miRNA which bridges the interaction between *HOTAIR* and Ago2 leading to *HOTAIR* degradation.

This chapter gave some examples of functions of lncRNAs in the regulation of gene expression post-transcriptionally. From these examples it becomes evident that there is no single common mechanism followed by all lncRNAs. On the contrary, lncRNAs are a very diverse class of transcripts both in terms of their nature (including length, structure, biochemistry, etc) and in terms of their functional mechanisms. However, what is common in all the examples above, is the centrality of the lncRNAs in their respective mechanisms. Admittedly, some lncRNAs do not function at the transcript level but are by-products from the passage of polymerase as is required to set the chromatin landscape before transcription of nearby protein-coding genes can commence (reviewed in Quinn & Chang, 2016). Nevertheless, it becomes apparent that a great portion of lncRNAs are functional and that their presence and function can be indispensable for correct gene regulation as demonstrated by severe clinical consequences in cases of aberrant lncRNA expression or mutation.

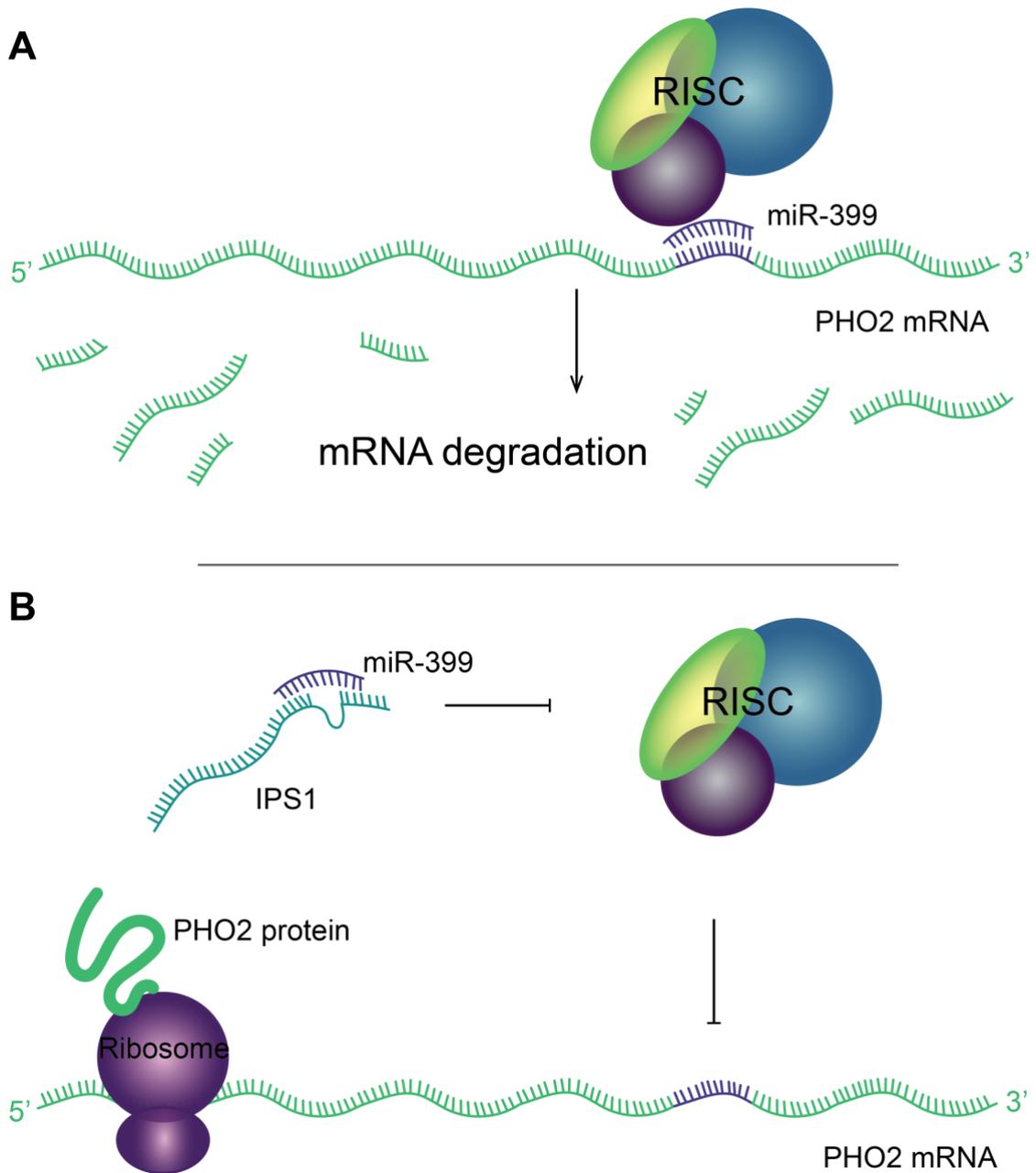


Figure 1:8 Mechanism of *PHO2* mRNA metabolism regulation by the lncRNA *IPS1*
 A) During normal plant growth conditions *PHO2* mRNA (green) is readily degraded by the RISC complex guided by *miRNA-399* (purple). (B) During phosphate starvation, the lncRNA *IPS1* (teal) is present, sequestering *miR-399* and preventing *PHO2* mRNA degradation which allows translation and accumulation of *PHO2* protein.

1.2.2 Clinical implications of long non-coding RNAs

From the functions of lncRNAs explored above, it becomes apparent that lncRNAs have a general role as regulators of gene expression sometimes linked to developmental processes or other times linked to specific cellular responses. It is therefore inevitable that the functions of lncRNAs are also connected to disease mechanisms. The Online Mendelian Inheritance in Man database (OMIM) lists 9,503 lncRNAs connected to genetic conditions and diseases and the Cancer lncRNA Census database (CLC) contains 122, experimentally confirmed cancer-related lncRNAs (Carlevaro-Fita *et al*, 2020). While it would be impossible to provide a complete description on how lncRNAs may contribute to diseases, in this section I will provide some examples of roles of lncRNAs in disease mechanisms and will then present research highlighting their potential as therapeutic targets.

The role of lncRNAs in epigenetic regulation of developmental genes suggests that the deregulation of lncRNA expression or stability can affect the development of organisms. For example, point mutations in the *ZEB2* gene or chromosomal abnormalities in its locus lead to an intellectual disability syndrome known as Hirschsprung disease or more commonly as Mowat-Wilson Syndrome (OMIM identifier #235730). As explained earlier (see chapter 1.2.1.3), *NAT Zeb2* lncRNA is transcribed from the antisense strand of *Zeb2* protein-coding gene and the presence of *NAT Zeb2* lncRNA regulates the splicing and stability of the protein-coding *Zeb2* mRNA. Moreover, *Zeb2* protein accumulates in aged animals while the expression of *NAT Zeb2* is triggered by differentiation cues (De Jesus *et al*, 2018). Finally, treatment with locked nucleic acids (LNA) restored the aged phenotypes of mice (De Jesus *et al*, 2018). Although much of the regulatory mechanism remains to be unearthed, it is already apparent that as we learn more about the mechanisms of lncRNAs, we can see more connections between them and disease mechanisms. In fact, just like *NAT Zeb2* the majority of long intergenic non-coding RNA (lincRNA) are proximal to important developmental protein-coding genes (Cabili *et al*, 2011; Guttman *et al*, 2009). For instance, *lincRNA-p21* is encoded on the antisense strand and is located 16.7 kb upstream of the gene coding for the cell-cycle regulator and tumour-suppressor protein p21 (Huarte *et al*, 2010). *LincRNA-p21* is involved in the regulation of many other genes including *cis*-regulation of the *p21* gene (Dimitrova *et al*, 2014).

Long non-coding RNAs have also been linked to cancer progression. For example, the lncRNAs of the *HOX* locus can become deregulated in breast cancer (Gupta *et al*, 2010). *HOTAIR* (see chapter 1.2.1.1) specifically is described as an oncogene (Arun *et al*, 2018). *HOTAIR* is overexpressed in several types of tumours including lung, pancreas, colorectal and breast cancers and it is involved in triggering metastasis (Arun *et al*, 2018). In primary breast cancer cells, *HOTAIR* can be used as a prediction marker for metastasis (OMIM identifier #611400) (Botti *et al*, 2016). Conversely, inhibition of *HOTAIR* expression can decelerate tumour growth in breast cancer cell lines (Gupta *et al*, 2010). In addition, human-origin colorectal tumours transfected with *HOTAIR* and grafted onto mice failed to invade other tissues (Arun *et al*, 2018). The effect of *HOTAIR* in tumour progression and malignancy has been attributed to its ability to regulate chromatin remodelling and in extent the cancer epigenome. Based on this hypothesis, therapeutic approaches targeting the interaction of *HOTAIR* with *Polycomb* group proteins are currently being researched. Examples of other oncogenic lncRNAs include Antisense Noncoding RNA in the *INK4* locus (*ANRIL*) which is encoded amongst important growth-regulatory protein-coding genes and regulates the chromatin accessibility of its own locus *in cis* via interactions with *Polycomb* group proteins (Reviewed in Arun *et al*, 2018), Metastasis Associated Lung Adenocarcinoma Transcript 1 (*MALAT1*) which was actually first identified through its involvement in metastasis (Ji *et al*, 2003; Gutschner *et al*, 2013) and lincRNA-*p21* which is involved in the regulatory networks of the important tumour suppressor proteins p21 and p53 (Reviewed in Arun *et al*, 2018).

In addition to oncogenic lncRNAs there are also some lncRNAs functioning as tumour suppressors. Some examples of tumour suppressor lncRNAs include maternally expressed gene 3 (*MEG3*) which will be described in detail in the next chapters and Non-coding RNA Activated by DNA damage (*NORAD*) whose downregulation is linked to chromosomal abnormalities and aneuploidy through the regulation of Pumilio-mediated degradation of mRNA coding for tumour suppressor proteins (Lee *et al*, 2016). Interestingly, a lot of the known oncogenic (e.g. *HOTAIR*) and tumour suppressor lncRNAs (e.g. *MEG3*) are also involved in developmental regulation and their own dysregulation is also linked to developmental syndromes. Altogether, it becomes apparent that lncRNAs are deeply involved in regulating healthy development and disease, their dysregulation can have adverse effects on

our health. Finally, since lncRNAs tend to be very tightly regulated and less abundant than proteins inside cells, they could be very promising targets for novel therapeutic approaches. For instance, the potential of *MALAT1* as target for the prevention of breast cancer metastasis has already been tested in mice with promising results (Arun *et al*, 2016).

1.2.3 The role of RNA structure in their functional mechanisms

Thus far, I have given examples of different mechanisms employed by lncRNAs. However, I have not yet stressed the role of RNA structures in those mechanisms. RNA structures are found in different classes of both coding and non-coding RNAs. For example, the mRNA coding for the cell-cycle regulator protein p27 forms a stable hairpin structure on its 3' UTR (Wan *et al*, 2011; Kedde *et al*, 2010). This hairpin contains a specific sequence consensus known as a Pumilio recognition sequence (PRE). Upon binding of the double-stranded RNA-binding protein (dsRBP) protein Pumilio-1 (PUM1) onto the PRE, the 5' part of the hairpin stem induces conformational changes to the structure of the 3' UTR resulting in recognition of the p27 mRNA by the RISC guided by miRNAs *miR-221* and *miR-222* (Kedde *et al*, 2010) (Figure 1:9). This example demonstrates how structural elements in RNA can provide an extra layer of regulation. Indeed, many mRNAs are regulated by structures on their UTRs such as, for example, internal ribosome entry sites (IRES) (Hellen & Sarnow, 2001).

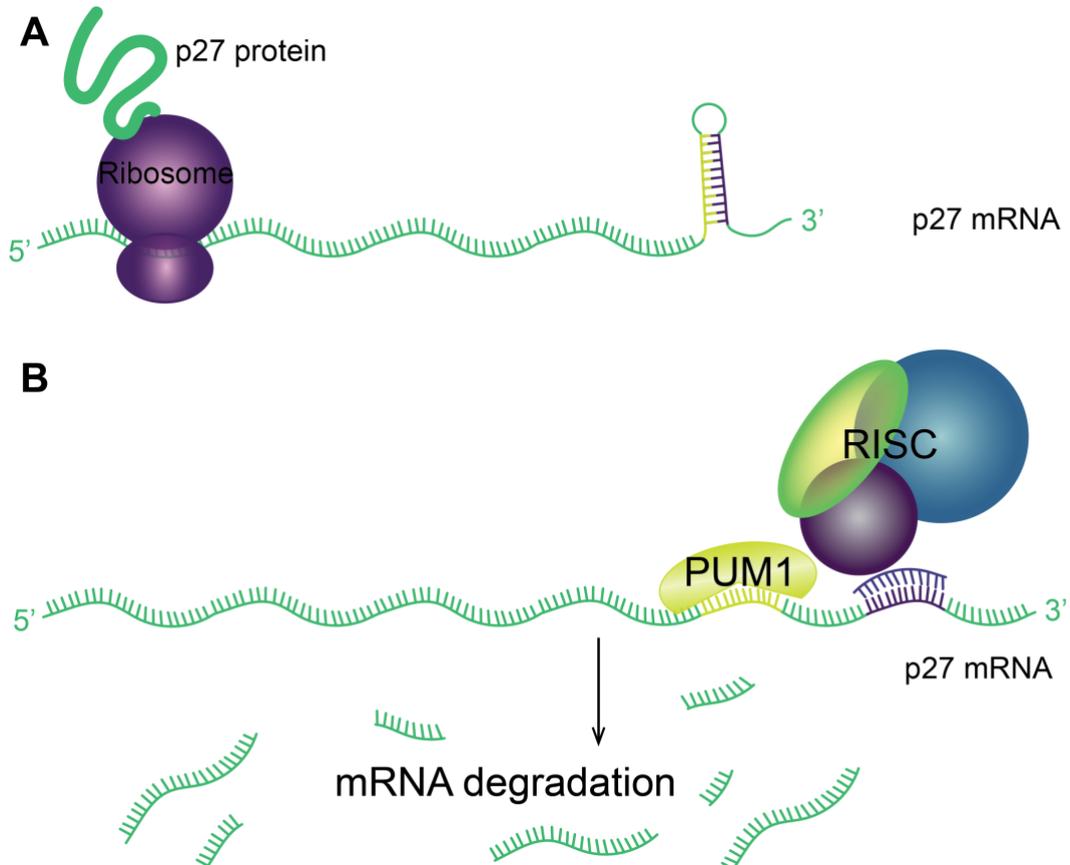


Figure 1:9 Regulation of the *p27* mRNA by 3' UTR structures and protein binding
 (A) Normal translation of the *p27* mRNA allows production of p27 protein by ribosomes in the cell. Translation of *p27* mRNA occurs in the absence of Pumilio-1 (PUM1) when the 3' UTR of *p27* mRNA forms a hairpin structure. (B) In the presence of PUM1 (yellow) the sequence corresponding to the 3' part of the hairpin (purple) is recognised by micro RNA *miR-221* and *miR-222* which guide the RISC complex onto the *p27* mRNA targeting it for degradation and thus preventing protein production of this mRNA.

Inside cells lncRNAs exhibit significantly more structured elements than mRNAs (Spitale *et al*, 2015). Therefore, it is not surprising that lncRNA functions often rely on either global or local structural features. The regulatory mechanism, controlling mRNA stability Pumilio protein binding, constitutes a good example of how structural elements on RNA guiding its function. As mentioned previously, *NORAD* is a lncRNA involved in chromosomal organisation and segregation during mitosis (Lee *et al*, Tichon *et al*, 2016 2016). *NORAD* contains multiple repeats of the PRE consensus sequence through which it sequesters Pumilio proteins (PUM1 and PUM2) preventing their interaction with mRNAs and subsequent degradation of these mRNAs (Lee *et al*, 2016; Tichon *et al*, 2016). The interaction of *NORAD* with Pumilio proteins is itself facilitated by the RBP SAM68. Mutation of *NORAD* on different positions including on either one of its two critical hairpin structures abolished the

ability of SAM68 to bind on *NORAD*, leading to subsequent dysregulation of the entire *NORAD*-PUM1/2 network (Tichon *et al*, 2018). Interestingly *NORAD* is organised in individual functional domains called *NORAD* repeat units (NRUs).

Another example of a structured RNA which folds in a modular fashion into distinct domains is once more the well-studied lncRNA *HOTAIR*. As mentioned above, *HOTAIR* is involved in chromatin remodelling via interactions with the PRCs (see chapter 1.2.1.1). The interaction with the PRC2 complex has been mapped to the 5'-end of the lncRNA (Rinn *et al*, 2007). The secondary structure of *HOTAIR* revealed that specific structural elements form on the sites where proteins such as the PRC2 complex are predicted to bind (Somarowthu *et al*, 2015). The 5'-end of *HOTAIR* as well as the identified recognition elements retain their specific folds even when downstream domains of the RNA have been deleted suggesting that the structural elements are a pre-requisite for protein-binding rather than the structures being stabilised due to protein binding (Somarowthu *et al*, 2015). In addition, the nucleotides forming those PRC2-interacting elements appear to be under higher evolutionary pressure as was determined by sequence alignment and structure-based co-variation analysis. Finally, studies have revealed that different parts of *HOTAIR* lncRNA can interact with different protein subunits of the PRC2 (Tsai *et al*, 2010). Thus, *HOTAIR* acts as a bridge connecting the PRC2 and chromatin in a mechanism that relies on its defined structures.

1.2.4 Evolution of long non-coding RNAs

Some of the examples of lncRNA functions I discussed in section 1.2.1 included functions identified through experiments in rodent models or in rodent embryonic stem cells. To do such experiments and be able to extrapolate the findings into the human system, we must first establish functional and mechanistic conservation of lncRNA activity across humans and the respective model organism(s). Sincerely, not all lncRNAs are under conservative evolutionary pressure. In addition, those lncRNAs that are conserved are several degrees of magnitude less well-conserved when compared to protein-coding genes (Wang *et al*, 2004). The rules governing non-coding genome conservation are different than those regulating protein-coding genes. For example, while mutations in protein-coding open reading frames can be silent due to the codon redundancy during translation, a mutation in a lncRNA gene immediately becomes apparent in the final transcribed molecule. To understand

lncRNA function and the evolution of their functional mechanisms, it is important to understand how lncRNA genes evolved. In this section, I will explain the levels of lncRNA gene conservation and how we can use them to infer functional information.

Identification of lncRNA homologues in distant species includes some hidden difficulties. Firstly, whole genome alignments of non-coding genes are not as straightforward as for protein-coding ones because the non-coding genome is under looser conservational pressure (Ulitsky, 2016). One way to overcome this problem is to align RNA sequencing (RNA-seq) data, provided that they are available for the organisms of interest. The use of RNA-seq data has the advantage that intronic sequences are not taken into consideration and therefore alignment between the lncRNA sequence of interest and its potential homologues is computationally less demanding (Ulitsky, 2016). However, RNA-seq data never contains the full genomic information, since the transcriptomes of different cell lines can vary significantly. Even though lncRNA genes evolve at faster rates than coding genes, the exons of lncRNA genes have been found to evolve slower than their introns (Guttman *et al*, 2009; Marques & Ponting, 2009). Evidence supporting the purifying selection of lncRNA exons in real life come from large-scale experiments using the model organism *Drosophila melanogaster*, which can be bred in large quantities, reaching an effective population size, large enough to identify evolutionary mechanisms (Haerty & Ponting, 2013). Thus, even though lncRNA genes as a whole evolve faster than their coding counterparts in a given genomic locus, we can still distinguish the functional non-coding regions if we exclude the coding sequences and compare the evolutionary conservation between the exons and introns of lncRNA genes.

In the lack of high-quality RNA-seq data, genomic alignment can be limited to syntenic regions assuming that if a given lncRNA is functionally important and has been conserved throughout evolution, it will be encoded in the same position on the ancestral chromosome. This method suffers from one major drawback: it is not possible to know whether identified homologous loci are actively transcribed and produce full-length lncRNA. In fact, research from the Ulitsky lab has revealed that the process of transcription evolved faster than the genome itself (Hezroni *et al*, 2015). Conversely, splicing is also an evolving process and, in some cases, splicing events are unique to one organism. For example, only 20% of lncRNA splicing events

were found to be conserved between humans and other primates and conservation did not extend to other vertebrates or sea urchins (Hezroni *et al*, 2015).

As highlighted earlier, the structure of some lncRNAs is an important determinant of their function (see chapter 1.2.3). If a structural element on a lncRNA is required to regulate a cellular function, then it is possible that overall primary sequence conservation might be insignificant as long as the base-pairing pattern driving structure can be established. This assumption is the main principle of structure-based covariation analysis (Nawrocki *et al*, 2009). For example, the mouse counterpart of the human lncRNA *HOTAIR* was found to be amongst the least conserved genes in the *HOX* loci of the two respective species (Schorderet & Duboule, 2011). Notably there seems to be functional conservation of the lncRNAs in the two species as in both humans and mice the lncRNA is involved in the regulation of epigenetic chromatin remodelling of the *HOXD* locus *in trans*. A structural study by the Pyle lab published four years later, revealed that several structural elements corresponding to potentially functional motifs of human *HOTAIR* lncRNA are under evolutionary pressure to maintain the structure as they exhibit a tendency for co-variation (Somarowthu *et al*, 2015). Finally, independent studies have shown that splicing enhancer *k*-mers are particularly well-conserved across species (Schüler *et al*, 2014; Hezroni *et al*, 2015). If we assume that the folding of lncRNAs in specific domains is important for function, then the high conservation of exon boundaries and exon regulatory sequences gains new gravity.

1.3 Maternally Expressed Gene 3

This thesis is about the long non-coding RNA maternally expressed gene 3 (*MEG3*), also known as Gene trap locus 2 (*Gtl2*) transcript. It is therefore essential for the understanding of the work presented in the following chapters to appreciate the biology of *MEG3*, and the research that has been done prior to the beginning of my own work.

1.3.1 Introduction to Maternally Expressed Gene 3

Meg3 was first identified in mice where it is encoded on the distal part of chromosome 12 and subsequently in humans, where the *MEG3* gene is located on chromosome 14q32, the syntenic chromosome of mouse distal chromosome 12 (Miyoshi *et al*, 2000). *MEG3* is mapped to the *Delta-like (Dlk)/Gtl2* (Schuster-Gossler *et al*, 1996; Zhang *et al*, 2010b) (Figure 1:10). The same locus also encodes 3 protein-coding genes as well as a variety of ncRNAs including a miRNA and a snoRNA cluster (Da Rocha *et al*, 2008). In mouse embryonic stem cells, the ncRNAs of the (*Dlk*)/*Gtl2* locus are expressed as a large (~200 kb) polycistronic RNA (Luo *et al*, 2016). The (*Dlk*)/*Gtl2* locus becomes imprinted during embryonic development so that the protein-coding genes are predominantly paternally expressed whilst the ncRNA, including *MEG3* are predominantly maternally expressed (Da Rocha *et al*, 2008). Imprinting of the (*Dlk*)/*Gtl2* locus is controlled at the imprinting control region (IG-DMR) by a mechanism similar to regulation of *Igf2r* locus via the lncRNA *Airn* (see chapter 1.2.1.1). The maternal chromosome is characterised by hypomethylation at the IG-DMR correlating with expression of the ncRNA polycistron (Sanli *et al*, 2018). *Meg3* is thought to be involved in controlling the silencing of maternal *Dlk1* protein-coding gene and the maintenance of hypomethylated state at the IG-DMR (Sanli *et al*, 2018; Kota *et al*, 2014; Sato *et al*, 2011).

The localisation of *MEG3* lncRNA itself is restricted to the nucleus and expression is tightly controlled at the tissue level. The tissues in which *MEG3* is most profoundly expressed are the brain, endocrine glands such as the pituitary, the kidneys and the placenta. Human *MEG3* shows a high degree of sequence conservation with its mouse homolog with a 67% sequence coverage. Later the conservation of *MEG3* was extended beyond human and mouse. For example, the Rfam database (Griffiths-Jones *et al*, 2003) identifies a region of *MEG3* conserved in at least 36

species (RF01872). As mentioned earlier, evolutionary conservation is not a common feature for lncRNA genes and those lncRNAs that show a high degree of conservation are likely functionally significant. In this work I will therefore describe our lab's efforts, including my own work to connect the evolutionary conservation, the structure and the function of *MEG3* aiming to fully understand its mechanism of action and its biological significance.

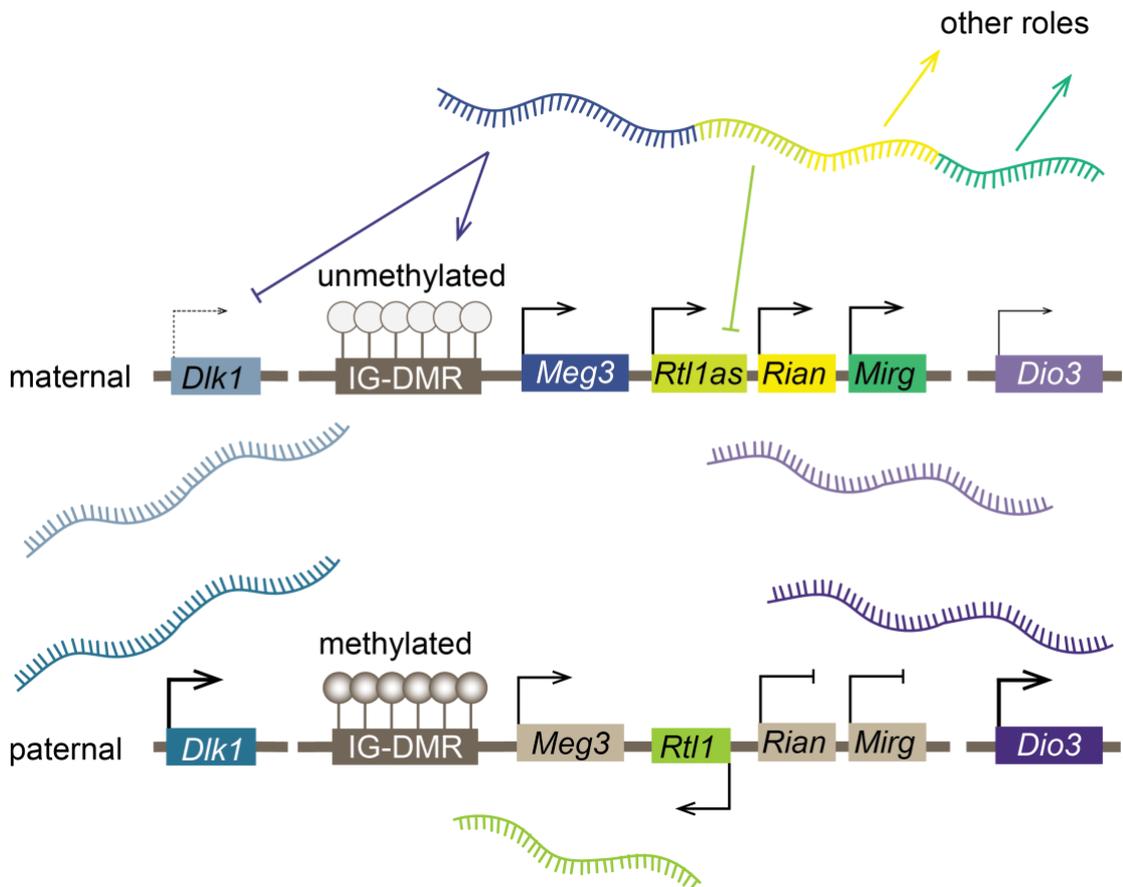


Figure 1:10 Graphic representation of the *Dlk1/Gtl2* locus in mice

Graphic representation of the *Dlk1/Gtl2* locus. The chromosome is represented by a brown line and gaps in the line indicate long distances in the DNA. Individual genes or gene clusters are shown as rectangles. Coloured rectangles indicate transcriptionally active genes and their corresponding transcripts are coloured accordingly. Beige rectangles represent inactive genes and the dull hues of maternal *Dlk1* and *Dio3* represent limited transcriptional activity from those genes. The maternally expressed polycistron is represented by a continuous, multicoloured transcript. Empty or filled lollipops represent the methylation status of the IG-DMR in the maternal and paternal chromosome respectively. *Meg3* lncRNA is involved in the silencing of maternal *Dlk1* gene and the maintenance of hypomethylation at the maternal IG-DMR. Maternally expressed *Rtl1as* is involved in the silencing of maternal *Rtl1* protein-coding gene. The maternally expressed micro- and sno-RNAs function away from the *Dlk1/Gtl2* locus. This figure was adapted from (Sanil *et al*, 2018) and implemented with information found in other references cited in this thesis.

1.3.2 Roles and mechanisms of *MEG3* in embryonic development

Failure to establish the correct imprinting pattern across the *Gtl2* locus has been linked to parental-origin-dependent phenotypic defects. In humans, incorrect imprinting in the *DLK1//Gtl2* locus is linked to diseases including neurodevelopmental defects (Hordijk *et al*, 1999; Kagami *et al*, 2008, 2015). For example, patients with paternal uniparental disomy of chromosome 14 (UPD14) exhibit musculoskeletal growth problems and mental retardation, whilst those with maternal UPD14 show short stature, hydrocephaly and early-onset puberty but no mental defects (Hordijk *et al*, 1999; Martin *et al*, 1999; Berends *et al*, 1999; Healey *et al*, 1994; Antonarakis *et al*, 1993; Temple *et al*, 1991; Coviello *et al*, 1996; Wang *et al*, 1991; Cotter *et al*, 1997). Statistically, cases of paternal UPD14 are much rarer than maternal UPD14s however there is no correlation between paternal UPD14 and increased miscarriage rates. Hence, this observation could be attributed to higher chromosomal translocation incidence in ova than spermatocytes. Similar observations have been made in mice. Specifically, mouse embryos with paternal duplication of chromosome 12 are not viable whilst pups born with maternal chromosome 12 duplication exhibit growth retardation and die soon after birth (Miyoshi *et al*, 2000). Further experiments using mice as a model organism have revealed that homozygous deletion mutants of *Meg3* exhibit the same phenotype as paternal UPD mice (Miyoshi *et al*, 2000). Thus, the developmental importance of the murine *Dlk//Gtl2* locus was linked to *Meg3* directly.

The phenotypes connected to aberrant *MEG3* expression led to further investigation of the mechanisms by which *MEG3* might affect embryonic development. In mouse embryonic stem cells (Iyer *et al*, 2017) and in human induced pluripotent stem cells (Kaneko *et al*, 2014), *MEG3* was found to functionally and potentially also physically interact with the *Polycomb* group proteins. In a study focusing on the potential interaction of *MEG3* with the PRC2, it was found that *MEG3* was involved in PRC2-dependent *trans*-regulation of the *ZIC* family genes which are involved in neuronal differentiation (Kaneko *et al*, 2014). In addition, murine *Meg3* was shown to be important for the PRC-dependent *cis*-regulation of the *Dlk* locus via another *Polycomb* group protein, Exh2 (Sanli *et al*, 2018). Mouse embryonic stem cells with undetectable levels of *Meg3* exhibit poor neuronal differentiation profiles (Mo *et al*, 2015). The developmental role of *Meg3* is strictly connected to the nervous system

as *Meg3* mutant mice did not exhibit any defect in haematopoiesis (Sommerkamp *et al*, 2019). Finally, *Meg3* was also dispensable for muscle development in goats even though genes neighbouring *Meg3* were required, suggesting that the *Dlk/Gtl2* locus is very tightly regulated and that its different genes have distinct functions (Fleming-Waddell *et al*, 2009). All in all, there is strong evidence supporting that *MEG3* plays a role in the neuronal development of embryos by a conserved mechanism (at least from humans to mice).

1.3.3 Roles of *MEG3* in the regulation of cell cycle and its clinical implications

The Cancer LncRNA Census (CLC) database (Carlevaro-Fita *et al*, 2020) classifies *MEG3* as a tumour suppressor gene and highlights its involvement in 16 out of the 29 cancer types included in the database. Previous studies characterising the links between *MEG3* and the PRCs hinted towards a role for *MEG3* in cancer biology through PRC-dependent cell cycle-regulation (Iyer *et al*, 2017). Several clinical studies have revealed a connection between loss of *MEG3* lncRNA levels and tumour progression, including tumours of the pancreas (Iyer *et al*, 2017), lungs (Terashima *et al*, 2017), liver (Braconi *et al*, 2011) and the pituitary gland (Chunharojrith *et al*, 2015; Li *et al*, 2015). Furthermore, experimental data have provided evidence of *MEG3* preventing uncontrolled cell proliferation when overexpressed in cancer cell lines (Zhang *et al*, 2003). Taken together, these observations characterise *MEG3* as a potential therapeutic target for the suppression of cancer proliferation.

Different molecular mechanisms have been found to connect *MEG3* with disease progression for the different types of cancer where it is involved. For instance, in pancreatic cancer, *MEG3* is not expressed due to changes in the methylation pattern surrounding the *Dlk/Gtl2* locus. Aberrant methylation of the *Dlk/Gtl2* locus was attributed to loss of menin 1 (MEN1) expression (Agarwal & Jothi, 2012). Pancreatic lineage cells derived from *Men1*^{-/-} mouse embryonic stem cells exhibit a significant loss in methylation of *Hox* gene loci (Agarwal & Jothi, 2012) suggestive of Polycomb group protein deregulation. Taken together, downregulation of Men1 leads to loss of *MEG3* expression which could in turn lead to abnormal *MEG3*-dependent PRC2 targeting (Iyer *et al*, 2017). Notably, the changes in methylation pattern at the *Dlk/Gtl2* locus are not expected to only affect the expression of *MEG3* but of all the

other co-regulated ncRNAs of the *MEG3* locus (see chapter 1.3.1 and references therein). Thus, any results linking *MEG3* and cancer via its genomic locus should be carefully reviewed to avoid missing any puzzle pieces in the form of a small co-regulated RNA for instance.

A study using mouse pancreatic cell lines and in archived human pancreatic islets, found that *MEG3* controls the regulation of the growth receptor c-Met (Modali *et al*, 2015). According to the proposed mechanism, *MEG3* controls the expression of c-Met by direct interactions with the DNA through the formation of triple helices (Iyer *et al*, 2017). According to this model *MEG3* can form a bridge between the c-Met locus and PRC2, silencing c-Met expression via chromatin modification. Essentially, the interaction between *MEG3* and the EZH subunit of the PRC2 requires full-length *MEG3*, suggesting that ribonucleoprotein complex formation requires structural integrity of *MEG3* (Iyer *et al*, 2017). However, an interaction assay with structure-retaining *MEG3* mutants or a negative control with random, linear RNA of the same length as *MEG3* has not been reported and therefore no strong conclusions can be drawn. Since there is no extensive data to support this model, it is also possible that *MEG3* might simply block access to transcriptional activators. Notably, no further studies have been performed to explore the structural features of *MEG3* and the *MEG3*-PRC2 complex or the possibility of *MEG3* interacting with other transcriptional regulator proteins. Nevertheless, the two existing models for the involvement of *MEG3* in pancreatic cancer support that the tumour suppressor function of *MEG3* is driven through PRC-dependent chromatin remodelling.

In lung cancer cell lines, *MEG3* inhibits epithelial-mesenchymal transition via downregulation of transforming growth factor- β (TGF- β)-responsive genes (Terashima *et al*, 2017). *MEG3* was proposed to interact with the PRC2 via its subunit JARID2 complex to drive repression of downstream targets of TGF- β . This is in line with the identified *Meg3*-PRC2 interactions involved in mouse embryonic development (Kaneko *et al*, 2014) but disagrees with the proposed mechanism for c-Met repression in pancreatic cancer via EZH2 (Iyer *et al*, 2017). In addition, a third team showed that *MEG3* could be co-purified with EZH2 from breast cancer cells and that *MEG3* formed DNA-RNA triple helices to direct PRC2 onto genes of the TGF- β pathway (Mondal *et al*, 2015). Importantly, although Terashima and colleagues have shown a correlation between levels of JARID2 and the level of

MEG3 on chromatin as well as the degree of repression, they have not shown any evidence supporting a direct interaction between any of the PRC2 subunits and *MEG3*.

The involvement of *MEG3* in hepatocellular carcinoma (HCC) has also been attributed to changes in the methylation status of the *MEG3* locus (Braconi *et al*, 2011). Lack of *MEG3* due to *MEG3* promoter hypermethylation is highly prevalent amongst HCC patients supporting the characterisation of *MEG3* as a genuine tumour suppressor. The involvement of miRNA-dependent dysregulation of DNA methylase enzymes has been proposed as a mechanism leading to initial *MEG3* silencing, however the exact mechanism through which *MEG3* acts in HCC remains unknown.

The roles of *MEG3* as a cell cycle regulator are not limited to the PRC-dependent pathway. *MEG3* is commonly downregulated in a variety of pituitary tumours ranging from the relatively benign non-functioning pituitary adenomas (Li *et al*, 2015) to tumours of gonadotroph lineage which are the most aggressive type of pituitary carcinomas (Zhang *et al*, 2003). Interestingly, *MEG3* is the only lncRNA listed for its involvement in pituitary cancer in the CLC database (Carlevaro-Fita *et al*, 2020). Studies by the Klibanski lab support that the tumour suppressor function of *MEG3* in the brain is via the tumour regulating master transcription factor p53 (Zhang *et al*, 2010b). This differentiates the mechanism of *MEG3* involvement in cancer progression in the pituitary compared to the pancreas and lungs without excluding the possibility of both PRC-dependent and independent pathways being involved in all three tissues. Because p53 is one of the most important cancer-related proteins, with p53 defects present in the majority of human tumours (OMIM ID #191170), the p53-*MEG3* interaction poses a highly interesting subject for further study. The results of these approaches have potential clinical value for the development of new cancer therapies.

1.3.4 Current state of research on *MEG3*

Initial studies on *MEG3*, primarily by the Klibanski lab, showed early evidence of *MEG3* structure being important for its molecular mechanism (Zhang *et al*, 2010a). In one key study a computational prediction model of the secondary structure of *MEG3* was generated (Zhang *et al*, 2010a). In this study, the researchers connected the functional effect of structural motifs from this model by coupling targeted

deletions of entire structured regions to a functional reporter assay (see chapter 2.5). This study by Zhang *et al* (2010a) set the ground for the study *MEG3* biology as it highlighted the potential connection between individual structural motifs on *MEG3* and its function. The work described in this thesis further investigates the connection of *MEG3* structure and its function in the regulation of p53-dependent gene expression.

1.4 The p53 pathway

Early evidence for the tumour suppressor function of *MEG3*, connected its activity to the p53 pathway. p53 belongs to a family of regulatory genes including p53, p63 and p73 all of which act as tumour suppressors (Chillemi *et al*, 2017). The 53 kDa protein p53 has been extensively studied in two fields (1) the regulation of the cell-cycle and (2) cell death by apoptosis.

1.4.1 The major cell-cycle regulator: p53

Transformation-related protein 53 (TRP53) (OMIM ID: 191170) more commonly known as p53, is perhaps best-known for its involvement in virtually every type of human cancer (Vousden & Lane, 2007). In healthy cells, p53 is activated upon stress conditions such as DNA damage or nutrient deprivation acting as a transcription factor for the regulation of genes involved in cell cycle control, metabolism, cellular senescence and apoptosis (Vousden & Lane, 2007). Activation of p53 can have one of three outcomes: (1) DNA repair, (2) cell cycle arrest, or (3) apoptosis upon prolonged stress. Maintenance of p53 at the inactive state is primarily achieved by the ubiquitin ligase enzyme MDM2 which targets p53 for proteasomal degradation (Vousden & Lane, 2007). The interactions between p53 and MDM2 are regulated by p53 phosphorylation. Unphosphorylated p53 is readily marked for degradation. Upon cellular stress, specific protein kinases phosphorylate p53, thereby inhibiting the MDM2-dependent ubiquitination and allowing p53 to interact with DNA, in turn activating stress-related genes (Figure 1:11). An additional mechanism of p53 activation is through its interaction with activator proteins such as p14. Interestingly, p14 has been observed to be regulated by other lncRNAs (Schmitt & Chang, 2016).

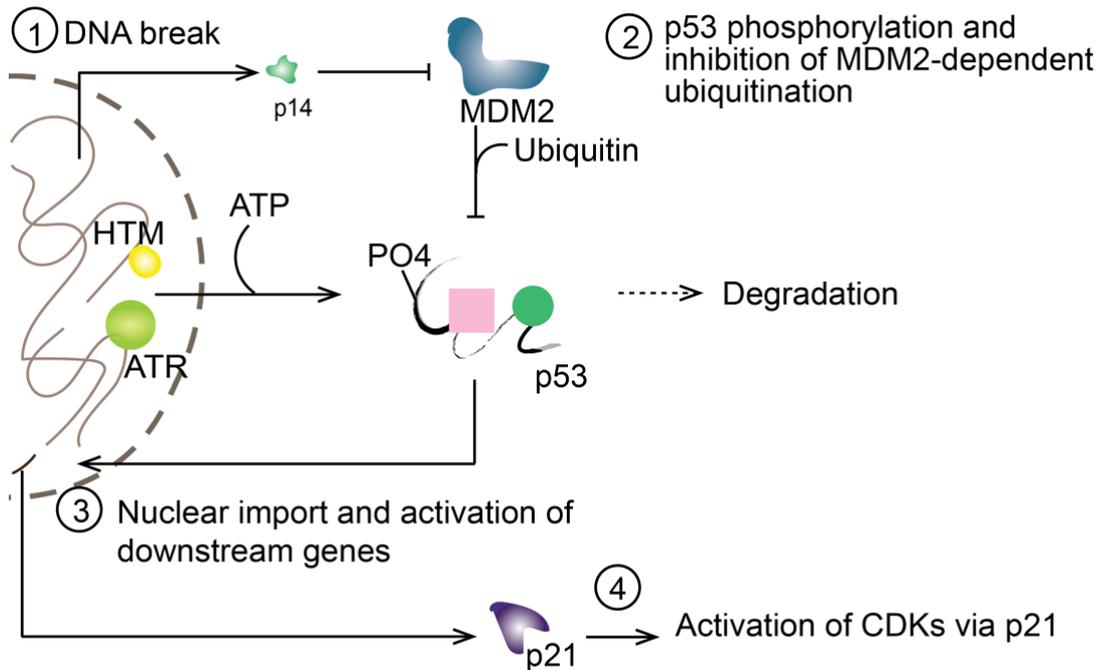


Figure 1:11 Schematic representation of the p53 pathway

The p53 pathway may be activated under cellular stress such as DNA damage. DNA damage is sensed in the nucleus by proteins such as HTM and ATR which then migrate outside of the nucleus to phosphorylate cytoplasmic p53 inhibiting its ubiquitination by MDM2. Simultaneously p14 is activated upon DNA damage and independently inhibits MDM2. Due to reduced ubiquitination, p53 is stabilised and escapes proteasomal degradation, migrates into the nucleus where it acts as a transcription factor for stress-related genes.

The human *TP53* gene can be alternatively spliced to nine isoforms (Bourdon *et al*, 2005). The active form of p53 protein is a tetramer of the 939 amino-acid isoform (Chillemi *et al*, 2017; Clore *et al*, 1995). Three regions of p53, spanning a total of ~150 amino acids, are intrinsically disordered. This makes p53 a challenging target for structural studies in its full-length, wild-type version. Nevertheless, due to its important role as a regulatory protein, a lot of research has focused on trying to understand the molecular basis of its interactions and its biochemical mechanism. Thus, through efforts of many groups we now have information on individual parts of p53 (Figure 1:12) as well as some full-length structural models of stable mutants. p53 can be separated in three domains: the N-terminal domain (NTD), the core which contains the DNA-binding domain (DBD), and the C-terminal domain (CTD). Within the NTD lies the transactivation domain of p53. The core contains the DNA-binding domain (DBD) and is linked to the termini via flexible linkers (Rajagopal, 2005). Finally, the CTD contains a homo-tetramerisation domain (Figure 1:12).

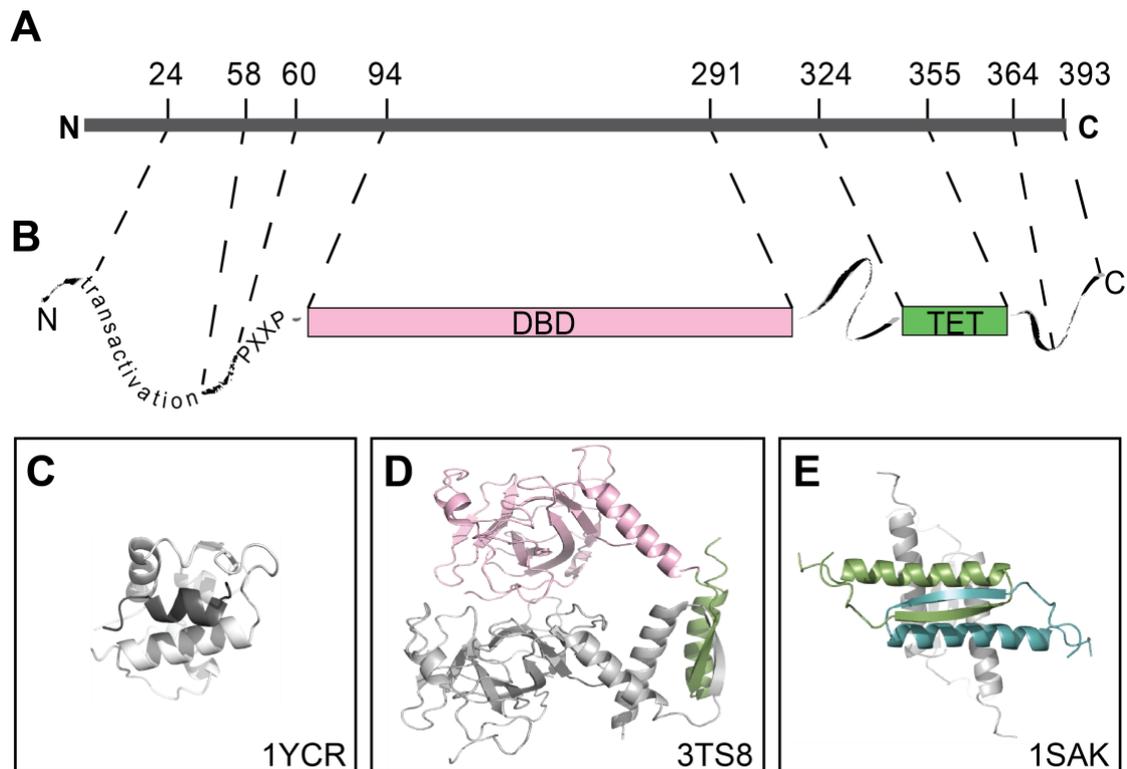


Figure 1:12 Three graphic representations of p53

(A-B) Schematic representation of the different domains of p53 and which amino acid residues they span; (B) domains shown in coloured boxes represent the domains that have more defined structures. The region marked between residues 364-393 represents the putative secondary nucleic acid binding area of p53 at its C-terminus; (C) structural model of the transactivation domain on p53 (dark grey) bound to MDM2 (white). Since this domain is intrinsically disordered, it can only be resolved when stabilized by bond formation (PDB ID: 1YCR); (D) structural model of the DBD of a p53 dimer as bound on DNA (hidden in figure). In the presented crystal structure, the DBD of one monomer (pink) was fused to the tetramerisation domain (green) removing the connecting flexible loop for stabilisation. The other copy of p53 DBD is shown in light grey (PDB ID: 3TS8); (E) four copies of the tetramerisation domain of p53. In this interaction the tetramerisation domains of two p53 monomers (green and teal) each donate one beta strand to form a short beta sheet thus stabilising each other (PDB ID: 1SAK).

1.4.2 p53 response elements

Activation of p53 can either up- or down-regulate target genes through two distinct pathways (Sullivan *et al*, 2018). Most commonly p53-dependent gene downregulation is mediated through a gene regulatory network involving the protein p21 and the DREAM complex (Sullivan *et al*, 2018). Target gene upregulation on the other hand occurs via direct binding of p53 on promoter proximal sequence (Fischer, 2019). Binding of p53 on DNA occurs via the DBD of p53 which comprises the largest part of the protein as it consists of amino acids 101 to 295 (Kern *et al*, 1991). DNA binding by the DBD is sequence-dependent (Kern *et al*, 1991). Specifically, the

interactions between p53 and the DNA occur at a consensus sequence on the DNA known as a p53 response element (RE) (Figure 1:13). The highest affinity RE consensus has been determined to consist of two decameric palindromes of the consensus 5'-RRRCWWGYYY-3' (where R = purine, C = Cytosine, W = Adenine or Thymine and Y = pyrimidine) that separated by a 0-13 base-pair spacer (El-Deiry *et al*, 1992). The palindromes can be split in quarters and hence each quarter can possibly form a higher order palindrome with up to two other quarters (Figure 1:13). Subsequently, variations of the RE consensus sequence affects the formation of those higher-order palindromic elements and in turn this has an effect on the observed mode of transcriptional regulation by p53 from the promoters coupled to each RE (Joerger & Fersht, 2008). For instance, coupling of the outermost quarter-palindromes has been loosely linked to transcription activation by p53 (Ma & Levine, 2007). The DBD can be divided into a loop-sheet-helix motif responsible for interactions with the major groove of DNA and the minor groove-interacting domain consisting of two flexible loops stabilised by a Zn²⁺ ion (Joerger & Fersht, 2008). Upon contact with an RE, the DBD of p53 undergoes a conformational change at loop L1 which is the loop contacting the major groove of the DNA (Petty *et al*, 2011) (Figure 1:13). Each quarter of RE palindromes is recognised by one p53 protein consistent with the p53 tetramer binding on the full RE. Interestingly, when comparing different structural models of the DBD bound to DNA probes that mimic REs, the fine details of the structural arrangement vary (Joerger & Fersht, 2008). Although it cannot be excluded that such variation is due to experimental deviations, it has been suggested that at least to an extent the observed variations reflect differences in the interactions between the DBD and REs with different sequence constitution. In line with this view, it has been observed that depending on the DNA probe studied, the interactions with the sugar-phosphate backbone at the minor groove may be direct or water-mediated whilst the interactions of R280 of p53 with the conserved guanine residue of the RE remains unchanged (Kitayner *et al*, 2006).

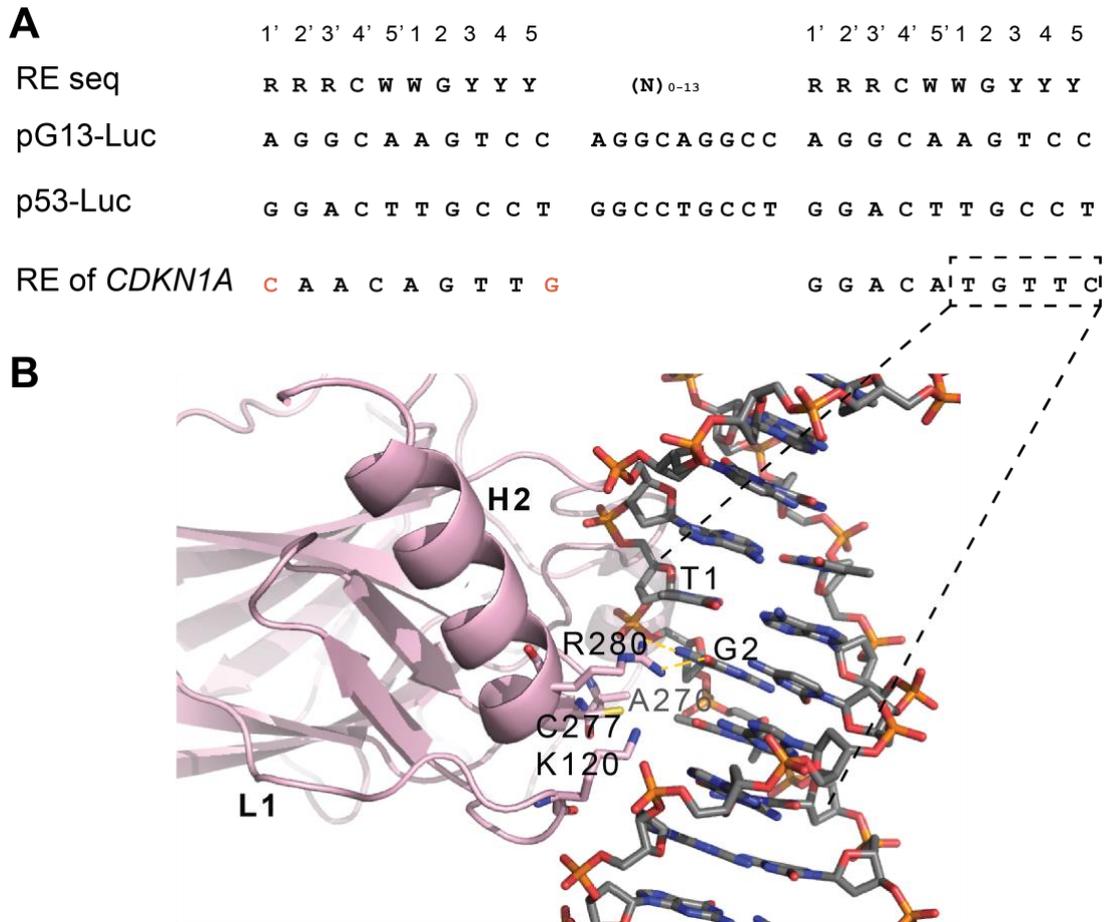


Figure 1:13 Recognition of p53 response elements

(A) The p53 response element (RE) consensus sequence, the sequences of the REs used for the experiments described in the results section and the naturally occurring RE of *CDKN1A* gene. The *CDKN1A* RE does not perfectly follow the consensus rule (nucleotides in red). (B) Graphical representation of the binding interphase of the DNA binding domain of a p53 monomer (pink) on one pentamer repeat. Loop L1 becomes inserted into the minor groove of the DNA whilst residues K120, A276, C277 and R280 form direct bonds with the bases (PDB ID: 3TS8).

1.4.3 Interactions of p53 with RNA

Interestingly the DBD is not the only part of p53 that interacts with nucleic acids. p53 has been found to associate with single- or double-stranded DNA and RNA via its very C-terminal region (residues 364-393) in a non-sequence specific manner (Riley *et al*, 2006). The functional importance of this region of p53 is not yet entirely clear however there is some evidence suggesting that nucleic-acid-binding by the CTD can activate sequence-specific binding of chromatin by the DBD via an unknown mechanism (Ayed *et al*, 2001).

Recent evidence from a yeast three-hybrid system suggests that p53 interacted with RNA in a structure-dependent manner and that p53-RNA interactions occurred via

the secondary nucleic acid binding C-terminal region of p53 (Riley *et al*, 2006). The results of the three-hybrid study were rather controversial and failed to elucidate the molecular mode of the interactions. Later efforts to characterise the molecular basis of the identified interactions demonstrated that the CTD of p53 bound RNA without undergoing any conformational changes or inducing any structural rearrangements on the RNA (Riley *et al*, 2007). To complicate this further, in the same study it was found that the CTD of p53 did not have a preference for highly structured RNA such as tRNAs (Riley *et al*, 2007). Failure to form strong interactions with tRNAs might imply selectivity of p53 for certain RNA structural elements rather than bias against structured RNAs. Finally, a different study by the same lab found that all CTD interactions that had been identified *in vitro* could not be detected in cultured human cells (Riley & Maher, 2007b). The same study suggested that post-translation modifications of p53 prevent its non-sequence specific interactions with nucleic acids (Riley *et al*, 2006).

Nevertheless, certain ncRNAs are more widely accepted as interaction partners of p53 and in fact their role in the p53 pathway has been characterised up to a certain extent. For example, the lncRNA *DINO* directly interacts with p53 upon DNA damage and promotes p53 stability upregulating genes involved in the DNA damage response (Schmitt *et al*, 2016). *DINO* binding by p53 is dependent on a specific stem loop structure of *DINO* but there is no sequence recognition involved in this interaction (Schmitt *et al*, 2016). Similarly, p53 has been found to interact with *MEG3* lncRNA (Zhu *et al*, 2015; Bauer *et al*, 2021). One group has reported binding of *MEG3* by the core domain of p53 (Zhu *et al*, 2015). Less is known about which regions of *MEG3* could bind p53. Functional studies suggest that the interaction depends more on the structure of this lncRNA, rather than its sequence (Zhang *et al*, 2010). Finally, there are examples of lncRNAs that regulate the p53 pathway without directly interacting with p53 protein. For instance, the lncRNA *PURPL* interacts with the regulatory protein MYBBP1A leading to p53 downregulation and a subsequent increase in cell proliferation rate (Li *et al*, 2017). The exact mechanism of *MEG3*-dependent regulation of the p53 pathway remains to be elucidated.

1.4.4 Conservation of p53 protein and of p53 pathway across mammals

The p53 proteins of mice and humans are very well conserved in terms of their primary sequence, domains and structures (Lion *et al*, 2015). However, the p53 gene regulatory networks of the two organisms exhibit substantial variations (Fischer, 2019). Specifically, the set of genes commonly downregulated by p53 through the p21-*DREAM* network seem to be better conserved from humans to mice than the commonly p53-upregulated genes (Fischer, 2019). This variation is greatly explain in terms of evolutionary turnover of the p53 REs regulating each one of those genes (Yue *et al*, 2014). This suggests that whilst it is possible that the regulatory mechanism between humans and mice is conserved, the gene regulatory networks vary since the DNA target sites for direct p53 binding have diverged in the course of evolution.

Chapter 2. Materials & Methods

2.1 Cell culture and treatments

2.1.1 Cell culture materials

Table 2:1 contains a summary of materials used for the methods described in this section.

Table 2:1 List of materials used in cell culture

Materials for cell based experiments		
Name	Provider	Cat number
Cell lines		
Hepa1-6	ATCC	CRL - 1830
HCT116 p53 ^{+/+}	Horizon Discovery	HD104-001
HCT116 p53 ^{-/-}	Horizon Discovery	HD104-001
Mouse Embryonic Fibroblasts (MEF)	ATCC	CRL-2991
Other materials		
Brilliant II SYBR® Green qPCR Master Mix	Agilent	600828
Dual Luciferase assay kit	Promega	E1910
Dulbeco's Modified Eagle's Media	Thermofischer	11054001
Erythrosine B	Roth	C.I. 45430
Foetal Bovine Serum	Thermofischer	10082147
L-Glutamine	Thermofischer	25030024
Lipofectamine 2000	Thermofischer	11668019
McCoy's 5A modified media	Invitrogen	16600082
Non-essential amino acids	Thermofischer	11140050
(±)-Nutlin-3	Bertin bioreagent	10004372
OptiMEM	Thermofischer	31985062
Phosphate buffered Saline	Sigma	806552
Trypsin 0.05% EDTA	Thermofischer	25300054
Software		
Prism	Graphpad	version 8.0.0
Microsoft Excel	Microsoft	version 16.42
Nanodrop 2000/2000C	Thermo Fischer	version 1.6.198

2.1.2 Cell culture

All cell culture procedures took place in the EMBL Grenoble Eukaryotic Expression Facility. Human colon carcinoma HCT116 p53^{+/+} and p53^{-/-} cells derived from a male adult were grown in McCoy's 5A modified media supplemented with 10% FBS and following standard protocols. Mouse embryonic fibroblasts (MEF) and Hepa1-6 cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS and following standard protocols.

2.1.3 RNA extraction and real time quantitative PCR (RT-qPCR)

Cells were grown in 6-well plates initially seeded with 16.6×10^4 cells per well. Cells were harvested and RNA was extracted using the RNA-easy kit (Qiagen). The quality of the extracted RNA was assessed by running a Bioanalyzer RNA 6000 Nano chip. RNA was reverse transcribed using the reverse-transcriptase SuperScript II (Thermofischer) following manufacturer's instructions and using random hexamers to primer the reaction. The RNA template was then digested by RNase H and the cDNA was amplified using the Brilliant II SYBR® Green qPCR Master Mix.

2.2 Molecular biology

2.2.1 Materials for molecular biology techniques

Table 2:2 contains a summary of materials used for the methods in this section.

Table 2:2 List of materials used for molecular cloning

Materials for molecular biology		
Name	Provider	Catalog number
Equipment		
C1000 PCR cycler	Bio Rad	-
x-tracta Gel Extraction Tool	Sigma	Z722390-100EA
Nanodrop 2000/2000C	Thermofischer	version 1.6.198
Cells		
<i>E. coli</i> Mach I competent cells	Thermofischer	C862003
Enzymes		
CutSmart Buffer	NEB	B7204S
DpnI	NEB	R0176S
EcoRI-HF	NEB	R3101S
HindIII-HF	NEB	R3104S
KpnI-HF	NEB	R3142S
NotI-HF	NEB	R3189S

Q5 hot-start polymerase	NEB	M0493S
T4 DNA ligase	NEB	M0202S
XbaI	NEB	R0145S
Other biochemistry materials		
Ampicillin	Sigma	A9393
DNA gel extraction kit	Zymoresearch	D4008
dNTP mix	NEB	N0447S
GelPilot DNA Loading Dye, 5x	Qiagen	239901
Kanamycin	Sigma	239901
LB media	EMBH GR	NA
NucleoBond Xtra maxiprep ki	Machery Nagel	740414
NucleoSpin plasmid mini-prep kit	Machery Nagel	740588.50
Orange G	Sigma	O3756-25G
Quick-Load® 2-Log DNA Ladder	NEB	N3200S
SOC media	Sigma	S1797
SYBR Safe DNA Gel Stain	Thermofischer	S33102
Ultrapure agarose	Thermofischer	16500100
Materials for analytical ultracentrifugation		
Name	Provider	Cat number
Equipment		
AN-50 Ti rotor	Beckman Coulter	363782
Beckman XL-A/XL-I centrifuge	Beckman Coulter	-
Ultracentrifugal cells	Beckman Coulter	-
Software		
GUSI	The University of Texas Southwestern Medical Center	version 1.08
GraphPad Prism	GraphPad Software Inc	version 8
Microsoft Excel	Microsoft	version 2013
Sedfit	(Schuck, 200)	version 14.6e
Unicorn	GE Healthcare	version 5.20

2.2.2 DNA and RNA Agarose electrophoresis

Gel electrophoresis was run on 1% agarose gels stained with SYBR Safe. The usual run time was 30 minutes at a voltage of 100 V. DNA samples for gel extraction were mixed with GelPilot DNA Loading Dye to 1X whilst RNA samples were stained orange with RNA dye (50% glycerol 10 mM Na-EDTA pH 8, 0.5% Orange G).

2.2.3 Sequence- and ligation-independent cloning (SLIC)

Sequence- and ligation-independent cloning (SLIC) was conducted as described in (Li & Elledge, 2012). For cloning of the pBluescript-mouse *Meg3* vector (pLA7), the sequence of mouse *Meg3v3* was acquired by gene synthesis by the Genewiz TURBO DNA synthesis service. The sequence was amplified by PCR and inserted by SLIC into the scaffold of a modified pBluescript vector immediately downstream of a T7 promoter sequence and immediately upstream of an XbaI restriction site. The modified pBluescript backbone originated from pTU1 as described in (Uroda *et al*, 2019). A list of primers used for SLIC can be found in Table 2:3.

Table 2:3: List of selected primers used for SLIC.

Selected primers for SLIC	
Name	Sequence 5' to 3'
E1_P3R	GAGGATGCTTGGCAGGAGATGGAGACACCAGCAGGGGCTTGTCTGGG GGCACTGGGAACCTTTGTGAGCGTGCTTCCACGGAGTAGAGCG
E1_P3F	GGCCTCAGGGCTATGGACAGACTCCTGTCCCATCCCAGAGACCCCT CGTGATCGTGCCCTGGCACGTGGGCCGTGGCCCCGGCTGGGTCTGG
E1_P4R	CCGACCCAGCCGGGCCACGGCCACGTGCCAGGGCAGATCACGAG GGTCTCTGGGATGGGACAGGAGTCTGTCCATAGCCCTGAGGCC
E1_P4F	CGCTCTACTCCGTGGAAGCACGCTCACAAAGGTTCCCAGTGCCCCC GACAAGCCCCTGCTGGTGTCTCCATCTCCTGCCAAGCATCCTC
E1_Oligo_E	CTGGTGTCTCCATCTCCTGCCAAGCATCCTCCAGTGCCTCCTCCTG TGGCCCTGGCCTCAGGGCTATGGACAGACTCCTGTCC

2.2.4 Quick-change polymerase chain reaction cloning

All point mutants of *MEG3* used in this thesis were constructed by Quick-change site directed mutagenesis unless otherwise stated. Q5[®] Hot start DNA Polymerase was used following the manufacturer's protocol. A list of primers used for Quick-change PCR can be found in Table 2:4.

Table 2:4 List of primers used for quick-change site-directed mutagenesis

Selected primers for quick-change PCR		
Name	Sequence 5' to 3'	Use
LA_E1#1	CGACTCACTATAGGGGCAGAGAGGGAGCGC	pTU1a generation (Uroda <i>et al</i> , 2019)
LA_E1#2	GCGCTCCCTCTCTGCCCTATAGTGAGTCG	
SMEG3Nr	GTGCGGCCGCTTTTTGTTAAGACAGGAAAC	Insertion of restriction sites (Uroda <i>et al</i> , 2019)
E5.2	GGTACCGCAGAGAGGGAGCGCGCC	
E14_29	CTTCCCACCCCTCTTGCTTGTCTGACTTGTCTTAT TTATT	Single TR mutants
E14_30	AATAAATAAGACAAGTCAGACAAGCAAGAGGGGTG GGAAG	
E14_31	CCCCTCTTGCTTGTCTTACTGGTCTTATTTATTCT CCAACA	
E14_32	TGTTGGAGAATAAATAAGACCAGTAAGACAAGCAA GAGGGG	
E14_33	TCTTGCTTGTCTTACTTGTCTGATTTATTCTCCAA CAGCACTC	
E14_34	GAGTGCTGTTGGAGAATAAATCAGACAAGTAAGAC AAGCAAGA	
E14_35	CCTCTTGCTTGTCTTACTTGTCTTATTGATTCTCC AACAGCAC	
E14_36	GTGCTGTTGGAGAATCAATAAGACAAGTAAGACAA GCAAGAGG	
E14_41	CTTCCCACCCCTCTTGCTGGTCTTACTTGTCTTAT TT	
E14_42	AAATAAGACAAGTAAGACCAGCAAGAGGGGTGGGA AG	
E14_73	GGTTTGACAGGTCAGTCCCTTCCCACCCCTCGTGC TTGTCTTACTTGTCTTATTTATTC	
E14_74	GAATAAATAAGACAAGTAAGACAAGCACGAGGGGT GGGAAGGGACTGACCTGTCAAACC	

E14_45	CTTCCCACCCCTCGTGCTTGTCTGACTTGTCTTAT TTATT	Double TR mutants
E14_46	AATAAATAAGACAAGTCAGACAAGCACGAGGGGTG GGAAG	
E14_47	CTTCCCACCCCTCTTGCTGGTCTGACTTGTCTTAT TTATT	
E14_48	AATAAATAAGACAAGTCAGACCAGCAAGAGGGGTG GGAAG	
E14_51	CCCCTCTTGCTGGTCTTACTGGTCTTATTTATTCT CCAACA	
E14_52	TGTTGGAGAATAAATAAGACCAGTAAGACCAGCAA GAGGGG	
E14_53	CCCCTCTTGCTTGTCTGACTGGTCTTATTTATTCT CCAACA	
E14_54	TGTTGGAGAATAAATAAGACCAGTCAGACAAGCAA GAGGGG	
E14_55	CAGTCCCTTCCCACCCCTCGTGCTTGTCTTACTTG TC	
E14_56	GACAAGTAAGACAAGCACGAGGGGTGGGAAGGGAC TG	
E14_57	CTTCCCACCCCTCTTGCTGGTCTTACTTGTCTGAT T	
E14_58	AATCAGACAAGTAAGACCAGCAAGAGGGGTGGGAA G	
E14_59	CTTCCCACCCCTCTTGCTTGTCTGACTTGTCTGAT T	
E14_60	AATCAGACAAGTCAGACAAGCAAGAGGGGTGGGAA G	
E14_61	CCCCTCTTGCTTGTCTTACTGGTCTGATT	
E14_62	AATCAGACCAGTAAGACAAGCAAGAGGGG	
E14_63	CTTCCCACCCCTCTTGCTTGTCTGACTTGTCTTAT TGATTCTCC	
E14_64	CTTCCCACCCCTCTTGCTTGTCTGACTTGTCTTAT TGATTCTCC	

E14_65	CCTCTTGCTTGTCTTACTGGTCTTATTGATTCTCC AACAGCAC
E14_66	GTGCTGTTGGAGAATCAATAAGACCAGTAAGACAA GCAAGAGG
E14_67	TCTTGCTTGTCTTACTTGTCTGATTGATTCTCCAA CAGCACTC
E14_68	GAGTGCTGTTGGAGAATCAATCAGACAAGTAAGAC AAGCAAGA
E14_69	GAAAAGGTTTGGACAGGTCAGTCCCTTCCCACCCCT CTGGCTTGTCTTACTTGTCTTATTTATTCTCCAAC AGCACTCCAGG
E14_70	CAAGGAGAGCCCCTGGACAAGGGCTGCCTGGAGTG CTGTTGGAGAATAAATAAGACAAGTAAGACAAGCC AGAGGGGTGGGAAGGGACTGACC
E14_71	CAGTCCCTTCCCACCCCTCTGGCTTGTCTTACTTG TC
E14_72	GACAAGTAAGACAAGCCAGAGGGGTGGGAAGGGAC TG
E14_73	GGTTTGACAGGTCAGTCCCTTCCCACCCCTCGTGC TTGTCTTACTTGTCTTATTTATTTC
E14_74	GAATAAATAAGACAAGTAAGACAAGCACGAGGGGT GGGAAGGGACTGACCTGTCAAACC
E14_75	GTCCCTTCCCACCCCTCTTGCTGGTCTTAC
E14_76	GACCAGCAAGAGGGGTGGGAAGGGACTGACCTG
E14_77	CTTCCCACCCCTCTGGCTTGTCTGACTTGTCTTAT TTATT
E14_78	AATAAATAAGACAAGTCAGACAAGCCAGAGGGGTG GGAAG
E14_87	CTTGTCTTACTGGTCTTATTTATTCTCCAACAGCA C
E14_88	GAATAAATAAGACCAGTAAGACAAGCCAGAGGGGT G
E14_89	CCCACCCCTCTGGCTGGTCTTACTTGTGTC
E14_90	GGAGAATAAATAAGACAAGTAAGACCAGCCAGAGG GG

E14_91	CCACCCCTCTGGCTTGTCTTACTTGTCTTATTGAT TCTCCAAC	
E14_92	GGAGTGCTGTTGGAGAATCAATAAGACAAGTAAGA CAAGCCAGAGGGG	
31F	GGTCAGTCCCTTCCCACCCCTCTTGCTTGTCTTAC TGGTCTGATTGATTCTCCAACAGCACTCCAGGCAG CCC	triple TR mutants
31R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATCAGA CCAGTAAGAC AAGCAAGAGGGGTGGGAAGGGACTGACC	
32F	GGTCAGTCCCTTCCCACCCCTCTTGCTTGTCTGAC TTGTCTGATTGATTCTCCAACAGCACTCCAGGCAG CCC	
32R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATCAGA CAAGTCAGACAAGCAAGAGGGGTGGGAAGGGACTG ACC	
34F	GGTCAGTCCCTTCCCACCCCTCTGGCTTGTCTTAC TTGTCTGATTGATTCTCCAACAGCACTCCAGGCAG CCC	
34R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATCAGA CAAGTAAGACAAGCCAGAGGGGTGGGAAGGGACTG ACC	
35F	GGTCAGTCCCTTCCCACCCCTCTTGCTTGTCTGAC TGGTCTTATTGATTCTCCAACAGCACTCCAGGCAG CCC	
35R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATAAGA CCAGTCAGACAAGCAAGAGGGGTGGGAAGGGACTG ACC	
36F	GGTCAGTCCCTTCCCACCCCTCTTGCTGGTCTTAC TGGTCTTATTGATTCTCCAACAGCACTCCAGGCAG CCC	
36R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATAAGA CCAGTAAGACCAGCAAGAGGGGTGGGAAGGGACTG ACC	

37F	GGTCAGTCCCTTCCCACCCCTCTGGCTTGTCTTAC TGGTCTTATTGATTCTCCAACAGCACTCCAGGCAG CCC
37R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATAAGA CCAGTAAGACAAGCCAGAGGGGTGGGAAGGGACTG ACC
38F	GGTCAGTCCCTTCCCACCCCTCTTGCTGGTCTGAC TTGTCTTATTGATTCTCCAACAGCACTCCAGGCAG CCC
38R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATAAGA CAAGTCAGACCAGCAAGAGGGGTGGGAAGGGACTG ACC
39F	GGTCAGTCCCTTCCCACCCCTCTGGCTTGTCTGAC TTGTCTTATTGATTCTCCAACAGCACTCCAGGCAG CCC
39R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATAAGA CAAGTCAGACAAGCCAGAGGGGTGGGAAGGGACTG ACC
40F	GGTCAGTCCCTTCCCACCCCTCTTGCTTGTCTGAC TGGTCTTATTGATTCTCCAACAGCACTCCAGGCAG CCC
40R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATAAGA CCAGTCAGACAAGCAAGAGGGGTGGGAAGGGACTG ACC
41F	GGTCAGTCCCTTCCCACCCCTCTTGCTGGTCTTAC TGGTCTGATTTATTCTCCAACAGCACTCCAGGCAG CCC
41R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATAAGA CCAGTCAGACAAGCAAGAGGGGTGGGAAGGGACTG ACC
42F	GGTCAGTCCCTTCCCACCCCTCTGGCTTGTCTTAC TGGTCTGATTTATTCTCCAACAGCACTCCAGGCAG CCC
42R	GGGCTGCCTGGAGTGCTGTTGGAGAATAAATCAGA CCAGTAAGACAAGCCAGAGGGGTGGGAAGGGACTG ACC

43F	GGTCAGTCCCTTCCCACCCCTCTTGCTGGTCTGAC TTGTCTGATTTATTCTCCAACAGCACTCCAGGCAG CCC	
43R	GGGCTGCCTGGAGTGCTGTTGGAGAATAAATCAGA CAAGTCAGACCAGCAAGAGGGGTGGGAAGGGACTG ACC	
44F	GGTCAGTCCCTTCCCACCCCTCTGGCTTGTCTGAC TTGTCTGATTTATTCTCCAACAGCACTCCAGGCAG CCC	
44R	GGGCTGCCTGGAGTGCTGTTGGAGAATAAATCAGA CAAGTCAGACAAGCCAGAGGGGTGGGAAGGGACTG ACC	
45F	GGTCAGTCCCTTCCCACCCCTCTTGCTGGTCTGAC TGGTCTTATTTATTCTCCAACAGCACTCCAGGCAG CCC	
45R	GGGCTGCCTGGAGTGCTGTTGGAGAATAAATAAGA CCAGTCAGACCAGCAAGAGGGGTGGGAAGGGACTG ACC	
46F	GGTCAGTCCCTTCCCACCCCTCTGGCTTGTCTGAC TGGTCTTATTTATTCTCCAACAGCACTCCAGGCAG CCC	
46R	GGGCTGCCTGGAGTGCTGTTGGAGAATAAATAAGA CCAGTCAGACAAGCCAGAGGGGTGGGAAGGGACTG ACC	
47F	GGTCAGTCCCTTCCCACCCCTCTGGCTGGTCTTAC TGGTCTTATTTATTCTCCAACAGCACTCCAGGCAG CCC	
47R	GGGCTGCCTGGAGTGCTGTTGGAGAATAAATAAGA CCAGTAAGACCAGCCAGAGGGGTGGGAAGGGACTG ACC	
48F	GGTCAGTCCCTTCCCACCCCTCTTGCTTGTCTGAC TGGTCTGATTGATTCTCCAACAGCACTCCAGGCAG CCC	Quadruple TR mutants
48R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATCAGA CCAGTCAGACAAGCAAGAGGGGTGGGAAGGGACTG ACC	

49F	GGTCAGTCCCTTCCCACCCCTCTTGCTGGTCTTAC TGGTCTGATTGATTCTCCAACAGCACTCCAGGCAG CCC	
49R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATCAGA CCAGTAAGACCAGCAAGAGGGGTGGGAAGGGACTG ACC	
50F	GGTCAGTCCCTTCCCACCCCTCTGGCTTGTCTTAC TGGTCTGATTGATTCTCCAACAGCACTCCAGGCAG CCC	
50R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATCAGA CCAGTAAGACAAGCCAGAGGGGTGGGAAGGGACTG ACC	
51F	GGTCAGTCCCTTCCCACCCCTCTTGCTGGTCTGAC TGGTCTGATTTATTCTCCAACAGCACTCCAGGCAG CCC	
51R	GGGCTGCCTGGAGTGCTGTTGGAGAATAAATCAGA CCAGTCAGACCAGCAAGAGGGGTGGGAAGGGACTG ACC	
52F	GGTCAGTCCCTTCCCACCCCTCTGGCTTGTCTGAC TGGTCTGATTTATTCTCCAACAGCACTCCAGGCAG CCC	
52R	GGGCTGCCTGGAGTGCTGTTGGAGAATAAATCAGA CCAGTCAGACAAGCCAGAGGGGTGGGAAGGGACTG ACC	
53R	GGTCAGTCCCTTCCCACCCCTCTGGCTGGTCTGAC TGGTCTTATTTATTCTCCAACAGCACTCCAGGCAG CCC	
53R	GGGCTGCCTGGAGTGCTGTTGGAGAATAAATAAGA CCAGTCAGACCAGCCAGAGGGGTGGGAAGGGACTG ACC	
62F	GGTCAGTCCCTTCCCACCCCTCTGGCTGGTCTTAC TTGTCTGATTTATTCTCCAACAGCACTCCAGGCAG CCC	Quintuple TR mutants
62R	GGGCTGCCTGGAGTGCTGTTGGAGAATAAATCAGA CAAGTAAGACCAGCCAGAGGGGTGGGAAGGGACTG ACC	

63F	GGTCAGTCCCTTCCCACCCCTCTGGCTGGTCTGAC TTGTCTTATTTATTCTCCAACAGCACTCCAGGCAG CCC	
63R	GGGCTGCCTGGAGTGCTGTTGGAGAATAAATAAGA CAAGTCAGACCAGCCAGAGGGGTGGGAAGGGACTG ACC	
64F	GGTCAGTCCCTTCCCACCCCTCTGGCTGGTCTTAC TTGTCTTATTGATTCTCCAACAGCACTCCAGGCAG CCC	
64R	GGGCTGCCTGGAGTGCTGTTGGAGAATCAATAAGA CAAGTAAGACCAGCCAGAGGGGTGGGAAGGGACTG ACC	
65F	GGTCAGTCCCTTCCCACCCCTCTTGTCTGTCTGAC TGGTCTGATTTATTCTCCAACAGCACTCCAGGCAG CCC	
65R	GGGCTGCCTGGAGTGCTGTTGGAGAATAAATCAGA CCAGTCAGACAAGCAAGAGGGGTGGGAAGGGACTG ACC	
TR1_struct_ re FW	GAGGCCTCAGGCAGGATCTGGCCTAGAGGAGGTGA TCAGCAAATGTTTGTTG	Structural rescue of TR mutants
TR1_struct_ re RV	CAACAAACATTTGCTGATCACCTCCTCTAGGCCAG ATCCTGCCTGAGGCCTC	
TR4_struct_ re FW	GAGGCCTCAGGCCGGATCTGGCATAGAGGAGGTGA TCAGCAAATGTTTGTTG	
TR4_struct_ re RV	CAACAAACATTTGCTGATCACCTCCTCTATGCCAG ATCCGGCCTGAGGCCTC	
TR1+4_struct_ re FW	GAGGCCTCAGGCCGGATCTGGCCTAGAGGAGGTGA TCAGCAAATGTTTGTTG	
TR1+4_struct_ re RV	CAACAAACATTTGCTGATCACCTCCTCTAGGCCAG ATCCGGCCTGAGGCCTC	
E16_1	GTGTTTCCTGTCTTAACAAAAAGCGGCCGCTCTAG AGTCG	Cloning and sequencing of mouse <i>Meg3</i> into pcDNA and
E16_2	GCCTGCAGGTCGACTCTAGAGCGGCCGCTTTTTGT TAAGACAGG	

E16_3	CTAATACGACTCACTATAGGGAAGCTTAGCACAGA AGACGAAGAGC	pBluescript vectors
E16_4	GCTCTTCGTCTTCTGTGCTAAGCTTCCCTATAGTG AGTCGTATTAG	
E16_seq0	CAAGGCCAGCAGAGCCGAGG	
E16_seq1	CTCGAGCAGCAGCCTTGGC	
E16_seq2	CCTCTTTGTGCGCTTGTCCTC	
E16_seq3	CTCAGGACATTGTTAGGACAGCATG	
E16_seq4	CAACAACCCTAGGAGGTTGCTCC	
E16_seq5	GTTCAATCTCATGAGTGAGTCCGGG	
E16_seq6	CGATGGCTAGGATTTTCGAGTTCGAC	
E16_seq7	CGTGTTGTGCGTGAAGTCCTC	
E16_seq8	GCGCAGTTCATCAGTCAGTAGG	
E16_seq9	GCTTTAGTTAAAACAAGAAATTTATTGAAAGCACC ATG	
E16_seq10	GCATGGGGCCTAGTGGCTCATGG	

2.2.5 Restriction cloning

Initial plasmids were digested overnight with the respective enzyme in CutSmart® Buffer at 37°C. After digestion, the reaction mixtures were mixed with GelPilot DNA Loading Dye to 1X and fragments were separated by agarose gel electrophoresis. Samples were run on 1% agarose stained with 6 µL of SYBR Safe DNA gel stain, at 100 V for 30 minutes along 10 µL of Quick-Load® Purple 2-Log DNA Ladder. The bands corresponding to the desired sizes were extracted under blue light. DNA was extracted from the gel bands using the Zymoclean Gel DNA Recovery kit and following the manufacturers protocol. Fragment ligation was using T4 DNA Ligase and following manufacturer's protocol.

2.2.6 Transformation and plasmid purification

All constructs were transformed and purified from *E. coli* Mach I chemically competent cells. Plasmid purification was performed using either the NucleoSpin plasmid mini-prep kit or the NucleoBond Xtra maxiprep kit for transfection-grade

plasmid DNA All plasmids were sequenced for verification using the Eurofins Tube-seq service.

2.3 RNA probing

2.3.1 Materials for RNA probing

Table 2:5 and Table 2:6 list essential instruments and materials respectively used for the methods described in this section, whilst Table 2:7 contains a list of all primers used for RNA probing experiments. Figure 2:1 shows a schematic of the workflow followed for RNA probing experiments and highlights the major differences between different types of RNA probing assays.

Table 2:5 List of essential instruments used in RNA probing experiments

Biochemistry instruments		
Name	Provider	Catalog number
C1000 PCR cycler	BioRad	-
AKTA system	GE Healthcare	-
Sephacryl S500 column	GE Healthcare	17-0613-01
Sephacryl S400 column	GE Healthcare	17-0609-01
Amicon® Ultra Centrifugal Filter	Millipore	UFC5100
Nanodrop 2000/2000C	Thermo Fischer	version 1.6.198
magnetic stirring bars	Dutscher	2070
magnetic stirring base	Roth	Y888.1
Glass Wool Non-Treated	Sigma	20384
Hamilton 500 µL	Dutscher	074492

Table 2:6 List of materials used in RNA probing experiments

Biochemistry materials		
Name	Provider	Catalog number
enzymes		
CutSmart Buffer	NEB	B7204S
EcoRI-HF	NEB	R3101S
Omniscript RT kit	Qiagen	205113
proteinase K	Sigma	P2308
RNaseOUT	Invitrogen	10777019
T7 RNA polymerase	EMBL HD	-
Thermo Sequenase kit	Life Technologies	785001KT
TURBO DNase	Invitrogen	AM1907

Xbal	NEB	R0145S
Chemicals		
(+)-Sodium L-ascorbate	Sigma	11140
(NH ₄)Fe(SO ₄) ₂ ·6H ₂ O	Sigma	3505
1-methyl-6-nitroisatoic anhydride (1M6)	Sigma	S888079
1-methyl-7-nitroisatoic anhydride (1M7)	EMBLHD	-
Acetate Glacial	VWR Chemicals	20104.298
Acrylamide	Roth	3030.1
Agarose, Ultrapure	Life Technologies	16500-500
APS	MP	802829
BaSO ₄	Sigma	243353-100G
Betaine	Sigma	B0300-1VL
CaCl ₂	VWR Chemicals	208291-250
Cacodylic Acid	Sigma	C0125-50G
6-Carboxy-4', 5'-dichloro-2', 7'-dimethoxyfluorescein, succinimidyl ester; (6-JOE)	Tebu-bio	AS-81011
Carboxyfluorescein, succinimidyl ester; (5-FAM)	Tebu-bio	AS-81007
deionised formamide	Sigma	47671
DEPC	Sigma	40718-25ML
DMSO	Sigma	D8418
Gel DNA recovery kit	Zymoresearch	D4002
DTT	MP	856126
EDTA	Roth	8043.1
EtOH Absolute	VWR Chemicals	20816.298
Glycerol	Euromedex	EU3550
Glycogen from mussels	Roche	10901393001
H ₂ O ₂ solution	Sigma	16911
HCl	VWR Chemicals	20252.209
HEPES	Euromedex	10-110-C
Isopropanol	VWR Chemicals	20880.290
KCl	Euromedex	P017
KOH	VWR Chemicals	26669.290
MgCl ₂ Hexahydrate	Sigma	13152-1KG
MgCl ₂ Tetrahydrate	Sigma	M3634-100G
MOPS	Sigma	M1254-100G
NaCl	Euromedex	1112-A

NaOH	VWR Chemicals	28244.295
N-methylisatoic anhydride (NMIA)	Sigma	129887
Orange G 25G	Sigma	O3756-25G
Quick-Load® 2-Log DNA Ladder	NEB	N3200S
RNase Zap	Invitrogen	AM9782
Spermidine	Sigma	S0266
SYBR Safe DNA Gel Stain	Thermo Fischer	S33102
TEMED	MP	805615
Terbium (III) Chloride hexahydrate	Sigma	212903-5G
Thiourea	Sigma	T 8656
Tris Base	Euromedex	200923-A
Triton X-100	Sigma	T8787
Urea	Euromedex	EU0014-B
Nucleotides		
ATP	GE Healthcare	27-1006-01
CTP	GE Healthcare	27-1200-04
GTP	GE Healthcare	27-2000-04
UTP	Sigma	U6750-1G
dNTP mix (10 mM)	NEB	N0447S
dNTPs kit	Roche	11969064001
ddGTP	Roche	3732738001
Other biochemistry materials		
Amersham MicroSpin G-25 Columns	Cytiva	GE27-5325-01
Zymoclean™ Gel DNA Recovery Kit	Zymoresearch	D4001
software		
GraphPad Prism	GraphPad Software Inc	version 8
Microsoft Excel	Microsoft	version 2013
QuSHAPE	(Karabiber <i>et al</i>, 2013)	version 1.0
SuperFold	(Siegfried <i>et al</i>, 2014)	version 1.0
Unicorn	GE Healthcare	version 5.20
VARNA	(Darty <i>et al</i>, 2009)	version 3-93

Table 2:7 List of primers used in RNA probing experiments

Primers for RNA probing		
Name	Sequence 5' to 3'	Use
Meg3RT1	CCTTGAAGACAAGGAGGTGG	human <i>MEG3</i> v1 (Uroda <i>et al</i> , 2019)
Meg3RT3	CCCAAAGGGATCCTTCCATTCAGG	
Meg3RT6	CCAGGAAGGAGACGAGAGGC	
Meg3RT8	CAGGAAACACATTTATTGAGAGCACAG	
Meg3RT22	CTGGCTGGTCAGTTCCGGTC	
Meg3RT42	CTATGCCAGATCCTGCCTGAGGC	
Meg3RT52	CAGGCCTTCAAGAAGCTTGGC	
Meg3RT72	CGTCCATCAGTCAGAGGGCG	
Meg3aRT2	CTGTCCATAGCCCTGAGGCC	
Meg3aRT	CTGTCCATAGCCCTGAGGCC	human <i>MEG3</i> v3 (Uroda <i>et al</i> , 2019)
E16_RT1	CTCGAGCAGCAGCCTTGGC	mouse <i>Meg3</i>
E16_RT2	CCTCTTTGTGCGTTGTCCCTC	
E16_RT3	CTCAGGACATTGTTAGGACAGCATG	
E16_RT4	CAACAACCCTAGGAGGTTGCTCC	
E16_RT5	GTTCAATCTCATGAGTGAGTCCGGG	
E16_RT6	CGATGGCTAGGATTTTCGAGTTCGAC	
E16_RT7	CGTGTTGTGCGTGAAGTCCTC	
E16_RT8	GCGCAGTTCATCAGTCAGTAGG	
E16_RT9	GCTTTAGTTAAAACAAGAAATTTATTGAAAGCACCATG	

2.3.2 Fluorescent dye coupling to primers used for RNA probing

Primers were designed and fluorescently-labelled following the guidelines described in (Chillón *et al*, 2015). Primers covering the lncRNA of interest every 200 nucleotides were designed to have melting temperature around $T_m \sim 60^\circ\text{C}$. Primers for RNA probing were ordered from Eurofins with a 5' amine group. Primers arrived lyophilised and upon arrival were resuspended in nuclease-free water. The labelling reaction (75 mM BaSO_4 pH 8.5, 2.52 $\mu\text{g}/\mu\text{L}$ 5-Carboxyfluorescein, succinimidyl ester; (5-FAM)) or 6-Carboxy-4', 5'-dichloro-2', 7'-dimethoxyfluorescein, succinimidyl ester; (6-JOE) dye respectively, 500 ng primer) was left overnight at 22°C shaking at 1100

rpm and covered to prevent photobleaching. Labelled primers were pelleted by the addition of ethanol to remove all uncoupled dye. Pellets were briefly washed with 70% ethanol solution and resuspended in 100 μ L nuclease-free water followed by 100 μ L deionised formamide. Primers were run on a 15% Urea-polyacrylamide-TBE gel at 20 W for >3 hours to separate the labelled from the unlabelled primer species. The bands corresponding to the labelled primers were excised under UV light. The labelled primers were extracted from the gel pieces by the “crush-soak method” in gel extraction buffer (10 mM Tris-HCl pH 8.0, 300 mM NaCl) overnight at 4°C and then pelleted in 100% ethanol. The labelled primers were resuspended in nuclease-free water. Primer concentration was measured using a nanodrop correcting for the respective fluorescent dye using the microarray mode. The concentration of primers was adjusted to 20 μ M stock and 2 μ M working aliquots respectively. A list of primers used for RNA probing assays can be found in Table 2:7

2.3.3 Generation of sequencing ladders

Sequencing ladders with primers labelled with 5-Carboxyfluorescein, succinimidyl ester; (5-FAM) were generated. For human *MEG3* variants, ladders detecting cytosines were generated and for mouse *Meg3*, detecting guanines. Sequencing ladder generation was conducted using Thermo Sequenase as described in (Chillón *et al*, 2015). Briefly, ladder generation reaction mix (template plasmid 1 ng/ μ L, 1X Thermo Sequenase buffer, dNTPs:ddNTP mix, 0.25 U/ μ L Thermo Sequenase) was PCR-amplified with the cyclor conditions shown in Table 2:8. Finally, samples were ethanol-purified and resuspended sequentially in 2 μ L and 38 μ L deionised formamide.

Table 2:8 PCR cyclor program for ladder generation reaction

Number of cycles	Temperature (°C)	Time
1	96	1 min
30	96	20 s
	55	20 s
	72	1 min
1	72	5 min

2.3.4 Preparation of NTPs

ATP, GTP CTP and UTP were purchased in powder form and reconstituted in nuclease free water. The pH of the solutions was adjusted to 6.8 by the addition of 2 M NaOH. The concentration of the NTP solutions was measured via nanodrop measure absorbance at 260 and 320 nm and applying the Beer-Lambert law. Tris-HCl pH 7.5 was added to a final concentration of 10 mM and finally the NTP solution concentration was adjusted to 100 mM by the addition of nuclease-free water.

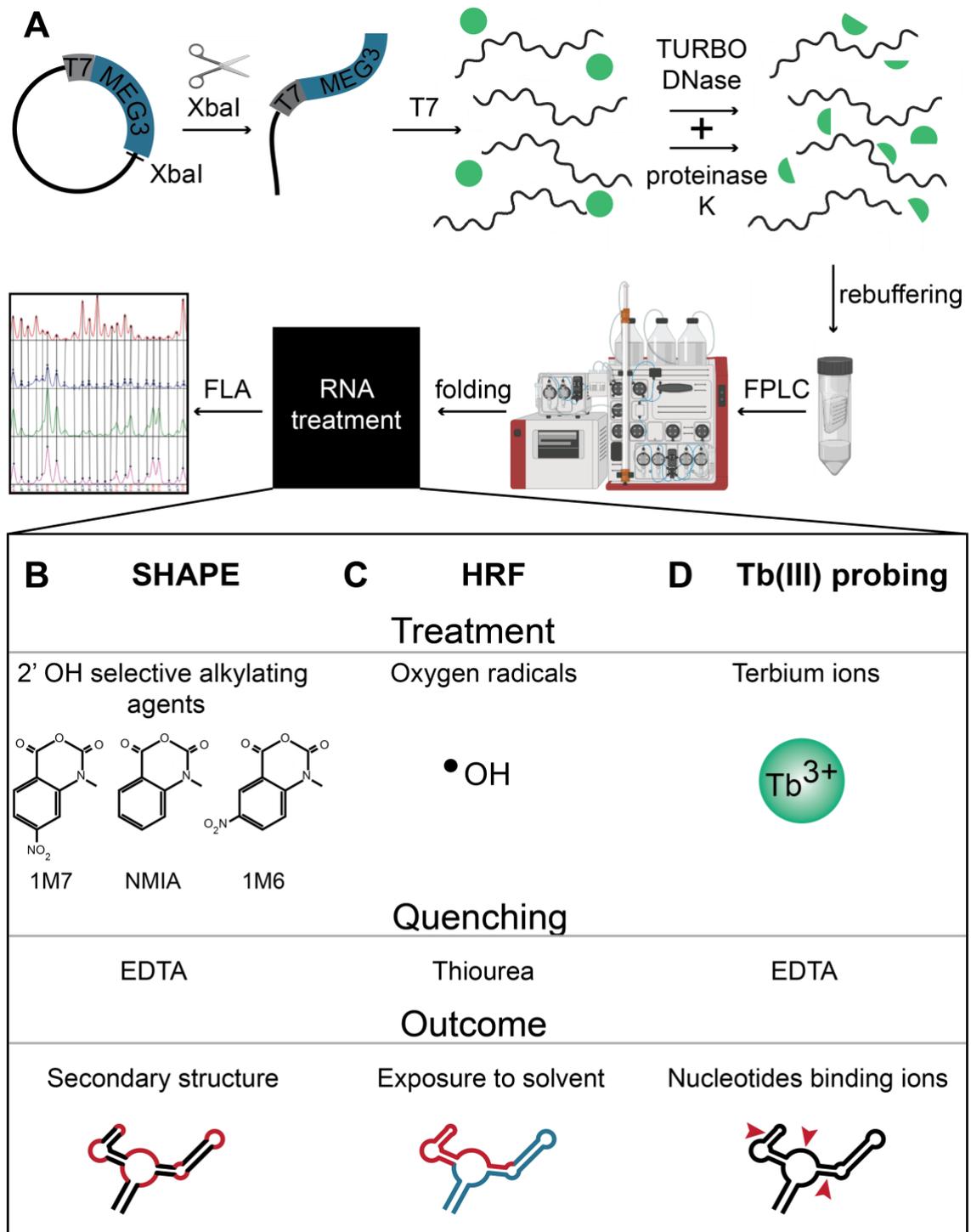


Figure 2:1 Workflow of RNA probing assays

(A) *MEG3* is encoded in a plasmid after a T7 promoter and followed by a restriction site. After restriction-digestion, RNA is produced by T7 Pol. RNA is isolated by sequential digestion with DNase and proteinase K. The RNA is rebuffered and concentrated in an Amicon column and then purified by liquid chromatography at room temperature. The purified RNA is folded and then treated with a “black-box” of reagents. Finally, the treated RNA is reverse-transcribed, and the cDNA reads are analysed by fragment length analysis. (B) Details of the treatment step in SHAPE experiments; RNA is treated with selective 2' OH alkylation agents

1M7, 1M6 or NMIA which alkylate the 2' OH group of ribose only on non-base-paired nucleotides. The reaction is quenched by EDTA and the assay reveals thermodynamically secondary structures. (C) Details of the treatment step in HRF experiments; RNA is treated with $\cdot\text{OH}$ radicals which break the nucleotide chain of RNA with at solvent-exposed nucleotides. The reaction is quenched by thiourea addition and the assay reveals which structural modules of the folded RNA are solvent-exposed. (D) Details of the treatment step in Terbium probing; RNA is treated with Tb(III) ions at neutral pH cleaving the nucleotide chain of RNA on the position of nucleotides chelating Mg^{2+} ions. The reaction is quenched by EDTA and the assay reveals the nucleotides coordinating Mg^{2+} ions. Figure adapted from (Chillón *et al*, 2015) and (Karabiber *et al*, 2013). Figure was partially created with Biorender.

2.3.5 *In vitro* RNA transcription

In vitro RNA transcription was performed as described in (Chillón *et al*, 2015). Briefly, 50 ng of linearised and desalinated template plasmid were mixed with 40 μL of each NTP (see section 2.3.4), 10 μL of DTT (1 M), 100 μL of transcription buffer (100 mM MgCl_2 , 400 mM Tris-HCl pH 8.0, 20 mM spermidine, 0.1% Triton X-100), 2 μL of RNaseOUT recombinant RNase inhibitor and 60 μL recombinant T7 RNA polymerase (EMBL Heidelberg protein purification facility) in a final reaction volume of 1 mL. Transcription was allowed to proceed for 2 hours at 37°C and was then terminated by centrifugation at 14,000 rpm for 5 min at 18°C and template digestion by TURBO DNase for 30 min at 37°C followed by protein digestion by 50 μL proteinase K [3 mg/ μL] in proteinase K buffer (10 mM Tris-HCl pH 7.5, 1 mM CaCl_2 , 40% glycerol) for another 30 min at 37°C. EDTA was added at equimolar concentration to the amount of divalent ions in the transcription buffer. Samples were briefly vortexed and used for RNA purification as describe in the next section.

2.3.6 Native purification of *in vitro* transcribed RNA

The transcription reaction product was transferred to two Amicon® Ultra Centrifugal Filter with a 100 kDa molecular weight cut-off and centrifuged at 5,000 rpm for 5 min at room temperature. Flow-through was discarded and the volume was restored to 0.5 mL per tube by the addition of gel filtration buffer. The process was repeated a total of 5 times. The sample was then purified by gel filtration chromatography at room temperature using a fractionator collecting fractions every 0.5 mL as described in (Chillón *et al*, 2015). Fractions with highest absorbance at 260 nm were collected and their RNA concertation was determined by absorbance at 260. A 10 μL sample of the fraction was tested for homogeneity by native agarose gel electrophoresis.

For SHAPE experiments K-MOPS gel filtration buffer (8 mM MOPS pH 6.5, 100 mM KCl, 0.1 mM EDTA pH 8.5) was used. For HRF and Terbium (III) probing experiments K-cacodylate gel filtration buffer (10 mM K-cacodylate pH 7.0, 0.1 mM Na-EDTA, pH 8.0, 150 mM KCl) was used.

Human *MEG3* and mouse *Meg3* full-length constructs were purified in a Sephacryl™ S500 column whereas *MEG3* Domains 2-3 constructs were purified in a Sephacryl™ S400 column.

2.3.7 Selective Hydroxyl-alkylation and primer extension (SHAPE) using fragment length analysis (FLA)

SHAPE was conducted following the protocol presented in [\(Chillón *et al*, 2015\)](#). Briefly, RNA was purified as described in section 2.3.6. 20 pmol of RNA from the peak fraction were allowed to fold to their native structure during the folding reaction (20 pmol RNA, 1X monovalent ion mixture, 15 mM MgCl₂) in a total volume of 490 µL and incubated at 37°C for 45 min. Four identical folding reactions were prepared in parallel. The monovalent ion mixture was previously prepared at 5X concentration (1 M KCl, 0.25 M HEPES pH 7.4, 0.5 mM Na-EDTA pH 8.5). Fresh dilutions of 1M7, 1M6 and NMIA reagents were prepared in DMSO to reach 10 mM. After folding 54.4 µL of one of each diluted reagent or DMSO in the control sample respectively were added to the samples. The samples were briefly vortexed and incubated for 5 min at 37°C. Then the SHAPE reactions were quenched by the addition of 40.5 µL SHAPE quenching solution (1.5 M Na-Acetate pH 5.2, 0.05 M Na-EDTA pH 8, 0.8 mg/mL glycogen). The RNA was purified by isopropanol precipitation and the pelleted RNA was re-suspended in 50 µL RNA storage buffer (1 M K-MOPS pH 6.5, 0.1 M Na-EDTA pH 8.5) and RNA was stored at -80°C until reverse transcription. Figure 2:2 shows the chemistry behind the SHAPE reaction and a schematic of its downstream analysis.

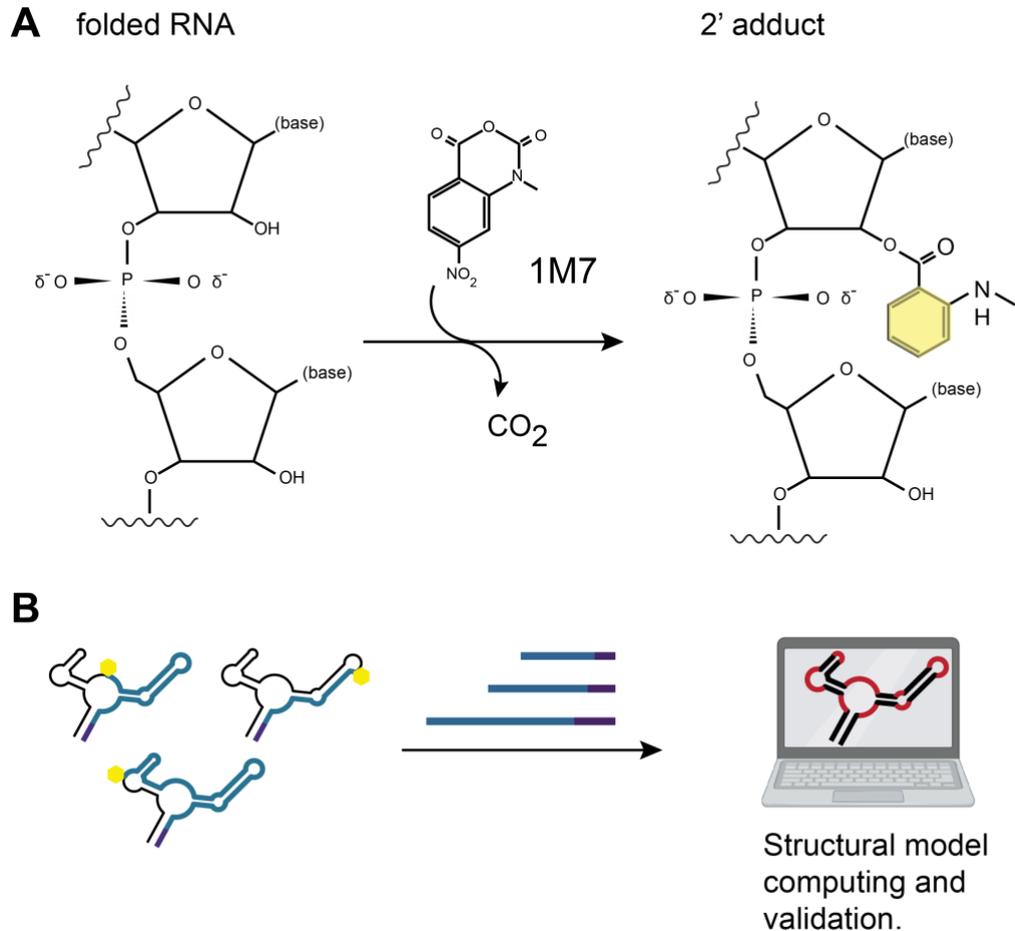


Figure 2:2 SHAPE chemistry and downstream analysis

(A) SHAPE reaction adapted from (Velema & Kool, 2020) (B) downstream analysis of SHAPE data. The deposited modifications cause the reverse-transcriptase to drop off the RNA template. Analysis of the fragments generated by reverse-transcription of treated RNA allows the generation of thermodynamically stable models of secondary structure. Figure was partially created with Biorender.

2.3.8 Hydroxyl-radical foot-printing (HRF)

HRF was performed as described in (Uroda *et al*, 2019). Briefly, RNA was purified as described in section 2.3.6. 20 pmol of RNA from the peak fraction were folded at 37°C for 45 min in a total volume of 94 μ L in the absence (10 pmol RNA 1X K-cacodylate buffer) or presence of divalent ions (10 pmol RNA 1X K-cacodylate buffer, 17.5 mM MgCl₂). Two of each MgCl₂ (+) and (-) folding reactions were prepared in parallel. One of each MgCl₂ (+) and (-) samples were treated with 9 μ L nuclease free water as control samples whereas the remaining two were treated with 3 μ L of 0.03% H₂O₂ solution, 3 μ L 50 mM Na-Ascorbate solution and 3 μ L of 4 mM (NH₄)Fe(SO₄)₂•6H₂O / 4.4 mM Na-EDTA (pH 8.0). The reagents or water were

added to the walls of the sample tube and the samples were vortexed to ensure all reagents begin to react with the RNA simultaneously. Figure 2:3 shows the chemistry behind the HRF reaction. Samples were incubated at 37°C for 15 seconds and then the reaction was stopped by the addition of 30 µL of HRF quenching solution (100 mM thiourea 200 mM EDTA pH 8.0) and transferring on ice. RNA was then precipitated by isopropanol precipitation and pellets re-suspended in 50 µL RNA storage buffer (1 M K-MOPS pH 6.5, 0.1 M Na-EDTA pH 8.5) stored at -80°C until reverse transcription. All reagents were freshly prepared and kept on ice.

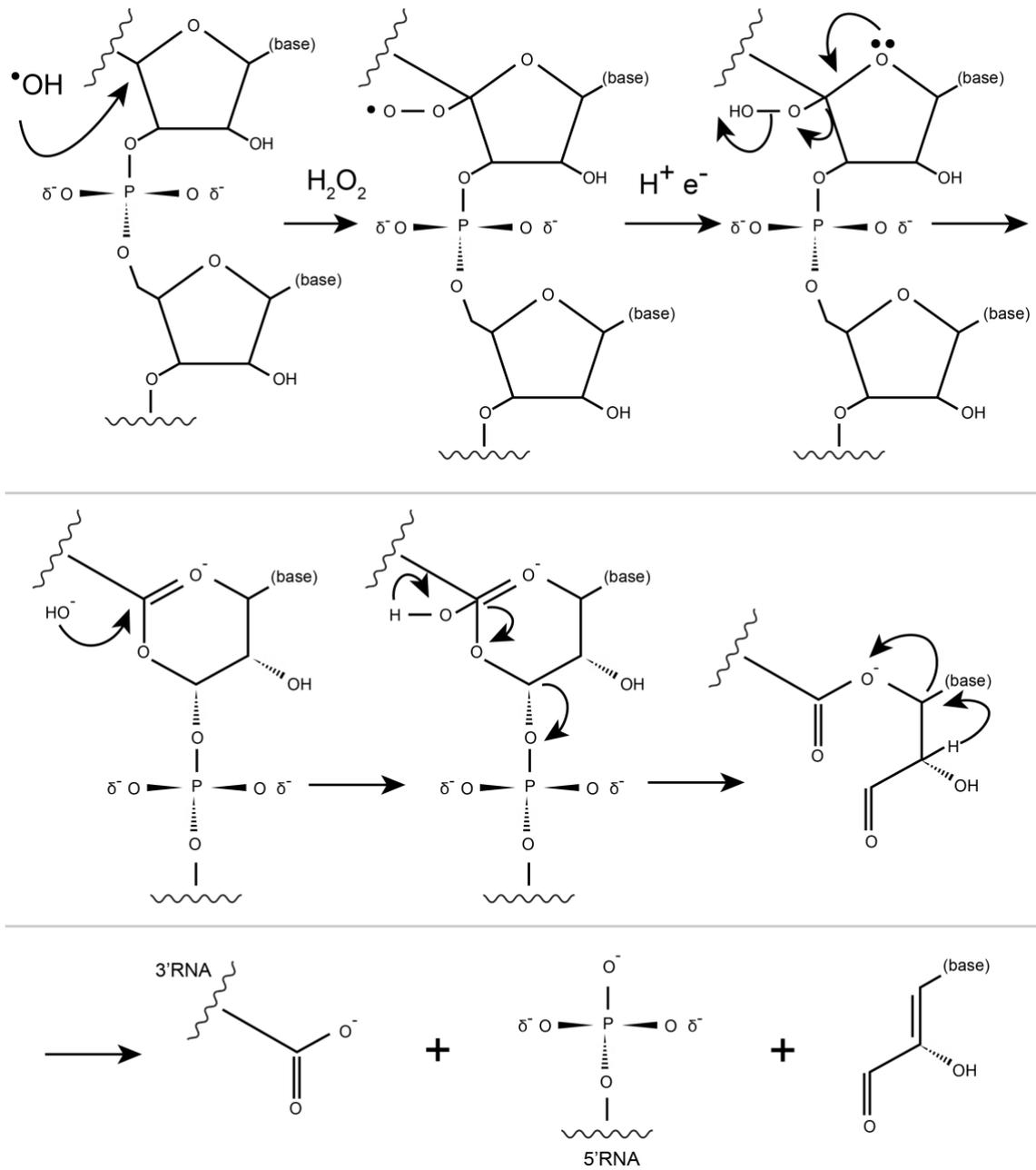


Figure 2:3 HRF chemistry

Reactive hydroxyl radicals cause cleavage of the ribonucleotide chain through a series of nucleophilic attacks. Adapted from (MacMillan, 2001).

2.3.9 Terbium (III) RNA probing

RNA for Terbium (III) RNA probing was produced and purified as for HRF. RNA from the peak fraction of the gel filtration step was folded in 1X K-cacodylate gel filtration buffer, supplemented with varying $MgCl_2$ concentrations ranging from 0 to 10 mM. After 45 minutes of folding at 37°C samples were treated with varying, freshly-prepared Terbium (III) solutions at final concentrations ranging from 0 to 10 μM and the samples were incubated for a further 45 minutes at 37°C. The reaction (Figure 2:4) was quenched by addition of the quenching solution (1.3 M Na-Acetate pH 5.2, 55 mM Na-EDTA pH 8.5, glycogen 20 $\mu g/\mu L$) and finally the RNA was purified by isopropanol precipitation. The treated RNA was stored in RNA storage buffer (Chillón et al, 2015) and was reverse-transcribed as described in the next section.

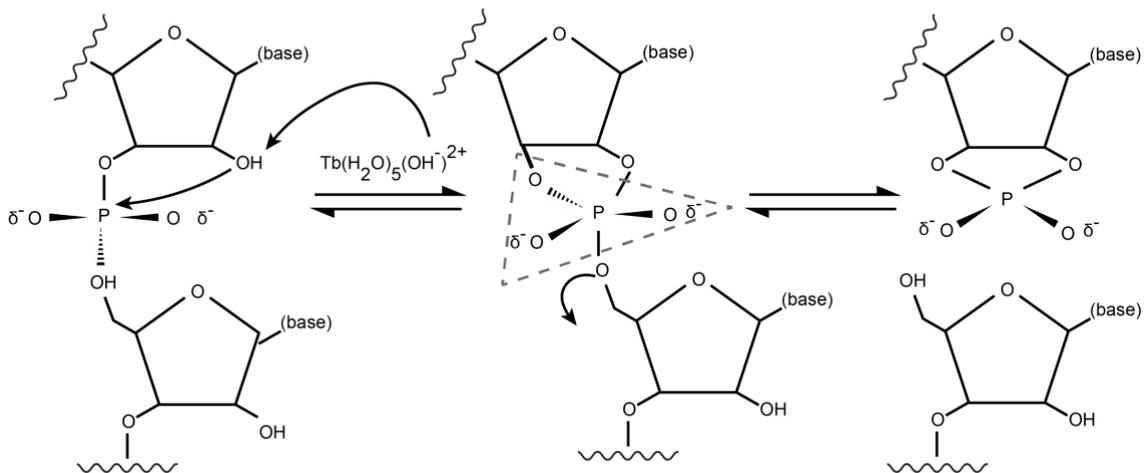


Figure 2:4 Tb(III) probing chemistry

As the first step Tb^{3+} ions replace Mg^{2+} within the folded RNA molecule since Tb^{3+} ions have a very similar ionic radius to this of Mg^{2+} ions. In a second step a series of nucleophilic attacks will lead to cleavage of the phosphodiester bond keeping the RNA chain connected. Figure adapted from (Harris et al, 2014).

2.3.10 Reverse transcription of RNA probing assays

RNA from all types of probing assays was reverse transcribed as described in Uroda et al 2015. Briefly, The treated RNA was thawed on ice and allowed to anneal with 6-JOE-coupled primers in the annealing reaction (1 pmol RNA, 0.2 μM primer, 0.4 M betaine) in a final volume of 12 μL at 65°C for 5 minutes and then the reactions were transferred on ice for a further 5 minutes. Reverse transcription was conducted using

the Omniscript RT kit. Briefly, 8 μL of reverse transcription reaction mix (2.5X Omniscript reaction buffer, 1.5 M betaine, 1.25 M Omniscript dNTP mix, 10 U RNaseOUT™, 0.25 U Omniscript reverse-transcriptase) were added to the annealing mix and the reaction was allowed to run for 1 hour at 37°C. The reaction was quenched by addition of the RT-quenching solution (1.3 M Na-Acetate pH 5.2, 11 mM Na-EDTA pH 8.5, 110 $\mu\text{g}/\mu\text{L}$ glycogen) and the RNA was precipitated by isopropanol precipitation. The pelleted RNA was resuspended in 2 μL nuclease free water followed by 38 μL deionised formamide. Samples were kept at -20°C.

2.3.11 Sequencing of RNA fragments from RNA probing assays

Samples were sequenced in parallel with their respective sequencing ladder. Fragment sequencing was done by the applied genetics department of Eurofins Genomics Germany GmbH, Ebersberg, Germany.

2.3.12 Fragment length analysis (FLA) of probing assays

The raw fragment length analysis (FLA) data were analysed using QuSHAPE (Karabiber *et al*, 2013). Values from QuSHAPE for successive ~400 nucleotide fragments were saved in text format and assembled to a continuous set of values corresponding to each nucleotide in the sequence of the RNA of interest. Values from three independent experimental replicas were collected and their average was used for downstream modelling. Standard error and Spearman's correlation were calculated using GraphPad Prism v.6.05 (GraphPad Software Inc) and outliers were removed. Reactivity values from all three replicates were normalised to their corresponding control sample using the "simple2boxplot.py" python script (Rice *et al*, 2014).

2.3.13 RNA secondary structure modelling

The *in vitro* SHAPE models were built as described in (Uroda *et al*, 2019). Briefly, the 1M6 reactivity values were first normalised and then subtracted from the NMIA reactivity values by running the python script "differenceByWindow.py" (Rice *et al*, 2014). Normalised 1M7 reactivity values were classified in 3 groups where: 0-0.40 non-reactive (suggesting that they are most probably base-paired), 0.40-0.85 moderately reactive and > 0.85 very reactive (suggesting that they are most probably flexible/ single stranded). The software SuperFold (Siegfried *et al*, 2014) with default

settings was used to generate secondary structure maps. Finally, the Java applet VARNA ([Darty et al, 2009](#)) was used to create graphical representations of the the secondary structure maps.

2.4 Biochemical and biophysical methods

2.4.1 Sedimentation velocity analytical ultracentrifugation (AUC)

Table 2:9 List of materials used for analytical ultracentrifugation

Materials for analytical ultracentrifugation		
Name	Provider	Cat number
Equipment		
AN-50 Ti rot	Beckman Coulter	-
Beckman XL-A/XL-I centrifuge	Beckman Coulter	-
Software		
GUSSI	The University of Texas Southwestern Medical Center	version 1.08
GraphPad Prism	GraphPad Software Inc	version 8
Microsof Excel	Microsoft	version 2013
Sedfit	(Schuck, 2000)	version 14.6e
Unicorn	GE Healthcare	version 5.20

Sedimentation velocity analytical ultracentrifugation (AUC) experiments were performed as described in ([Chillón et al, 2015](#)). A list of materials used for this experiment can be found in Table 2:9, Briefly, *Meg3* was purified as described in section 2.3 Purified *Meg3* was allowed to fold in AUC folding buffer (100 mM HEPES pH 7.4, 400 mM KCl, 0.2 mM Na-EDTA pH 8.5) supplemented with varying concentrations of MgCl₂ ranging from 0.1 mM to 25 mM. Samples were analysed using Beckman XL-A/XL-I centrifuge with AN-50 Ti rotor (Beckman Coulter). All experiments were performed at 20°C at 25,000 rpm overnight. Experiments were conducted at the Institute de Biologie Structurale (IBS) in Grenoble. AUC data were analysed with Sedfit ([Schuck, 2000](#)) using continuous c(s) distribution model. Data from three independent replicates were analysed and the standard error was calculated using GraphPad Prism.

2.5 Dual Luciferase Assay

The luciferase reporter assay was used as a functional assay for different *MEG3* constructs as described in (Zhou *et al*, 2007) and in (Uroda *et al*, 2019). The materials and cell lines used for the experiments are listed in Table 2:10

Table 2:10 List of materials used for luciferase reporter assays

Materials for cell-based experiments		
Name	Provider	Cat number
Cell lines		
Hepa1-6	ATCC	CRL - 1830
HCT116 p53 ^{+/+}	Horizon Discovery	HD104-001
HCT116 p53 ^{-/-}	Horizon Discovery	HD104-001
Mouse Embryonic Fibroblasts (MEF)	ATCC	CRL-2991
Other materials		
Dual Luciferase assay kit	Promega	E1910
Dulbecco's Modified Eagle's Media	Thermofischer	11054001
Erythrosine B	Roth	C.I. 45430
Foetal Bovine Serum	Thermofischer	10082147
L-Glutamine	Thermofischer	25030024
Lipofectamine 2000	Thermofischer	11668019
McCoy's 5A modified media	Invitrogen	16600082
Non-essential amino acids	Thermofischer	11140050
(±)-Nutlin-3	Bertin bioreagent	10004372
OptiMEM	Thermofischer	31985062
Phosphate buffered Saline	Sigma	806552
Trypsin 0.05% EDTA	Thermofischer	25300054
Software		
Prism	Graphpad	version 8.0.0
Microsoft Excel	Microsoft	version 16.42
Nanodrop 2000/2000C	Thermo Fischer	version 1.6.198

2.5.1 Cell seeding

Cells were detached from T75 flasks using 2 mL trypsin 0.05% and incubated at room temperature for 1 min. 8 mL of media were added and the cell suspension was transferred to 50 mL Falcon tubes. Cells were pelleted by centrifugation at 1,200 rpm for 5 minutes and the trypsin-containing media were replaced with fresh media. A 15 µL aliquot was mixed with equal volume of erythrosine B and two 15 µL aliquots

of the mix were independently measured using the Countess automated cell counter (Invitrogen). The average of the two measurements was calculated and used to determine the dilutions required for seeding.

2.5.2 Transfections

24 h post-seeding, cells were transfected with 115.96 fmol of pcDNA3 vector containing the respective *MEG3* constructs, 50 ng of p53-Luc (Zhou *et al*, 2007) and 5 ng of pRL *Renilla* luciferase Control Reporter Vector (Promega) using Lipofectamine 2000 (Life Technologies), according to manufacturer's instructions. Transfected cells were incubated for 48 h.

2.5.3 Luciferase Assay

Cells were seeded as described in section 2.5.1 to a final concentration of 8.3×10^4 cells per well of a 12-well plate and final volume 1 mL per well. 24 hours post seeding cells were transfected as described in section 2.5.2. 48 h post transfection media were removed by aspiration and cells were lysed by 1 x Passive Lysis Buffer (Promega E1910) for 30 min shaking. The cell lysate suspension was moved to Eppendorf tubes and centrifuged at 15000 rpm for 1 min. 10 μ L of the lysate were transferred to a well of a white flat-bottomed 96-well plate and the relative *Renilla* luciferase amount was determined using the Dual-Luciferase® Reporter (DLR™) Assay System (Promega E1910) following manufacturer's protocol. A CLARIOstar plus microplate reader was used for the readout.

2.5.4 Luciferase assay signal normalisation

The readout from the plate reader was analysed using Microsoft Office Excel. The relative *Renilla* luciferase signal was normalised over this of *Firefly* luciferase. The activity signal was expressed as a percentage of this exhibited either from the control vector or from the human *MEG3v1* condition depending on the requirements of respective experiments. Standard error of three independent replicates was calculated and experiments with higher than 50% were repeated.

2.6 Computational alignments

Table 2:11 lists the software used for sequence and structural alignments.

Table 2:11 Software used for structure and sequence-based alignments

Computational alignments materials	
Name	Reference
Software	
Clustal Omega	(Li <i>et al</i> , 2015)
Infernal	(Nawrocki <i>et al</i> , 2009)
Jalview	(Waterhouse <i>et al</i> , 2009)
R2R	(Weinberg & Breaker, 2011)

2.6.1 RNA sequence alignments

BLAT (Kent, 2002) was used to identify the homologous sequences of each human *MEG3* exon in the genomes of other mammals. The identified sequences were aligned using Clustal Omega (Li *et al.*, 2015) as described in (Uroda *et al*, 2019). Results were visualised using the JAVA applet Jalview (Waterhouse *et al*, 2009).

2.6.2 RNA structural alignments

Secondary structure-based alignments and co-variation analysis were done using Infernal (Nawrocki *et al*, 2009) as described in (Uroda *et al*, 2019). Briefly, mammalian sequences homologous to human *MEG3* exon E3 were aligned using Clusal Omega (Li *et al*, 2015). The seed sequences were pre-selected so that each order of mammals was equally represented. Infernal (Nawrocki *et al*, 2009) was used to generate and calibrate a structure-based alignment, assess co-variation and explore the genomic DNA of *Bos taurus* for potential *MEG3* analogues. Results were illustrated by the programme R2R (Weinberg & Breaker, 2011) and the JAVA applet Jalview (Waterhouse *et al*, 2009).

2.7 Crystallisation trials

Table 2:12 lists the software used for sequence and structural alignments.

Table 2:12 Materials used for crystallisation trials

Materials for crystallisation trials		
Name	Provider	Catalog number
Equipment		
Bioanalyzer	Agilent	-
Bioanalyzer RNA 6000 Nano Kit	Agilent	5067-1511
Crystallisation trial plates		
Index	Hampton research	HR2-144
The Natrix	Hampton research	HR2-116
The Nucleix suite	Qiagen	130719
HELIX	MD	(Viladoms & Parkinson, 2014)

The crystallisation constructs were purified by size exclusion chromatography as described in section 2.3.6 and then concentrated to 5 mg/mL using an Amicon column with 100 kDa molecular weight cut-off. Sample purity was tested using a Bioanalyzer RNA 6000 Nano chip following manufacturer's protocol. Samples were submitted to the high throughput crystallisation (HTX) facility at EMBL Grenoble and several nucleic acid specific crystallisation trial plates were set (see Table 2:12).

Chapter 3. Results

3.1 Identification of the active core of the lncRNA *MEG3*

3.1.1 Résumé en français

Le gène humain *MEG3* code pour 12 exons (E1 à E12) entrecoupés d'introns de différentes longueurs (Figure 1:1). Dans les tissus où *MEG3* est exprimé, la base de données du NCBI contient au moins 15 isoformes d'épissage de *MEG3* humain (Figure 1:1). Toutes les isoformes documentées de *MEG3* ont en commun l'inclusion des exons flanquants E1-E3 et E10-E12. La variation provient de l'épissage alternatif des exons centraux E4-E9 ainsi que du choix de certains sites d'épissage alternatifs dans les exons E1 et E3. Les premiers efforts pour caractériser fonctionnellement les effets des différentes isoformes d'épissage de *MEG3* ont montré que différents variants de *MEG3* peuvent avoir des efficacités différentes dans la conduite de l'expression *en trans* via la protéine p53 (Zhang *et al*, 2010).

Sur la base de cette observation, notre laboratoire a sélectionné 3 isoformes d'épissage de *MEG3* humain pour des études détaillées. Nous avons choisi les variants v1, v3 et v9 (Figure 1:1). *MEG3v1* est le variant le plus abondant dans les cellules et est donc le plus pertinent d'un point de vue biologique. J'utiliserai la *MEG3v1* comme référence pour toutes les études décrites dans cette thèse. De plus, nous avons choisi d'étudier les variantes v3 et v9 de *MEG3* qui montrent une différence minimale dans la composition de leurs exons par rapport à *MEG3v1*. Plus précisément, *MEG3v3* a un exon supplémentaire par rapport à v1 tandis que *MEG3v9* a un exon de moins que v1 (Figure 1:1). Néanmoins, l'inclusion d'un seul exon supplémentaire (v3) ou de moins (v9) semble être suffisante pour entraîner de manière reproductible des différences significatives en termes d'efficacité de ces variantes à entraîner l'expression génique dépendant de p53 (Zhang *et al*, 2010; Zhou *et al*, 2007) (Figure 1:2). De cette observation découle la première question à laquelle je cherche à répondre dans cette thèse : comment la composition des exons affecte-t-elle l'architecture moléculaire de *MEG3* et comment l'architecture moléculaire se traduit-elle en fonction? Dans ce chapitre, je décris l'approche expérimentale pour répondre à cette question et j'explique ce que les résultats nous disent.

3.1.2 Chapter Introduction

As described in chapter 1.3, the lncRNA *MEG* is encoded in several mammals and is involved in the regulation of development during embryonic stages and of the cell cycle in adults. Whilst some research has tried to cast light on the molecular mechanisms involved in these processes, there are still a lot of unknowns. For example, previous research has shown that *MEG3* can act on the p53 pathway (see chapter 1.3.3) but, how does the sequence of *MEG3* dictate its downstream effect on cell cycle regulation? As part of my PhD I set out to explore this question by analysis naturally occurring splicing isoforms of *MEG3* in humans and comparing their structural and functional similarities.

The human *MEG3* gene encodes 12 exons (E1 to E12) interspersed with introns of various lengths (Figure 3:1) In the tissues where *MEG3* is actively expressed, the NCBI nucleotide database contains at least 15 splicing isoforms of human *MEG3* lncRNA (Figure 3:1). Common to all documented splicing isoforms of *MEG3* is the inclusion of flanking exons E1-E3 and E10-E12. Variation arises from alternative splicing of the middle exons E4-E9 as well as the choice of some alternative splice sites within exons E1 and E3. Some initial efforts to functionally characterise the effects of the different splicing isoforms of *MEG3* have shown that different *MEG3* variants can have different efficiencies in driving p53-dependent gene expression *in trans* (Zhang *et al*, 2010a).

Based on this initial observation, our lab singled out 3 splicing isoforms of human *MEG3* for detailed studies. We chose variants v1, v3 and v9 (Figure 3:1). *MEG3v1* is the most abundant variant in cells and is therefore the most biologically relevant one. I shall use *MEG3v1* as a reference for all studies described in this thesis. Additionally, we chose to study *MEG3* variants v3 and v9 which show minimal difference in their exon composition compared to *MEG3v1*. Specifically, *MEG3v3* has one extra exon compared to v1 (exon E7) whilst *MEG3v9* has one less exon than v1 (exon E5) (Figure 3:1). Nevertheless, the inclusion of a single extra (v3) or less (v9) exon seems to be sufficient to reproducibly drive significant differences in terms of the efficiency of those variants in driving p53-dependent gene expression (Zhang *et al*, 2010a; Zhou *et al*, 2007) (Figure 3:2). From this observation, arises the first question which I aim to answer in this thesis: How does exon composition affect the molecular architecture of *MEG3* and how does molecular architecture in turn

translate to function? In this chapter, I describe the experimental approach to answering these questions and explain what the results tell us.

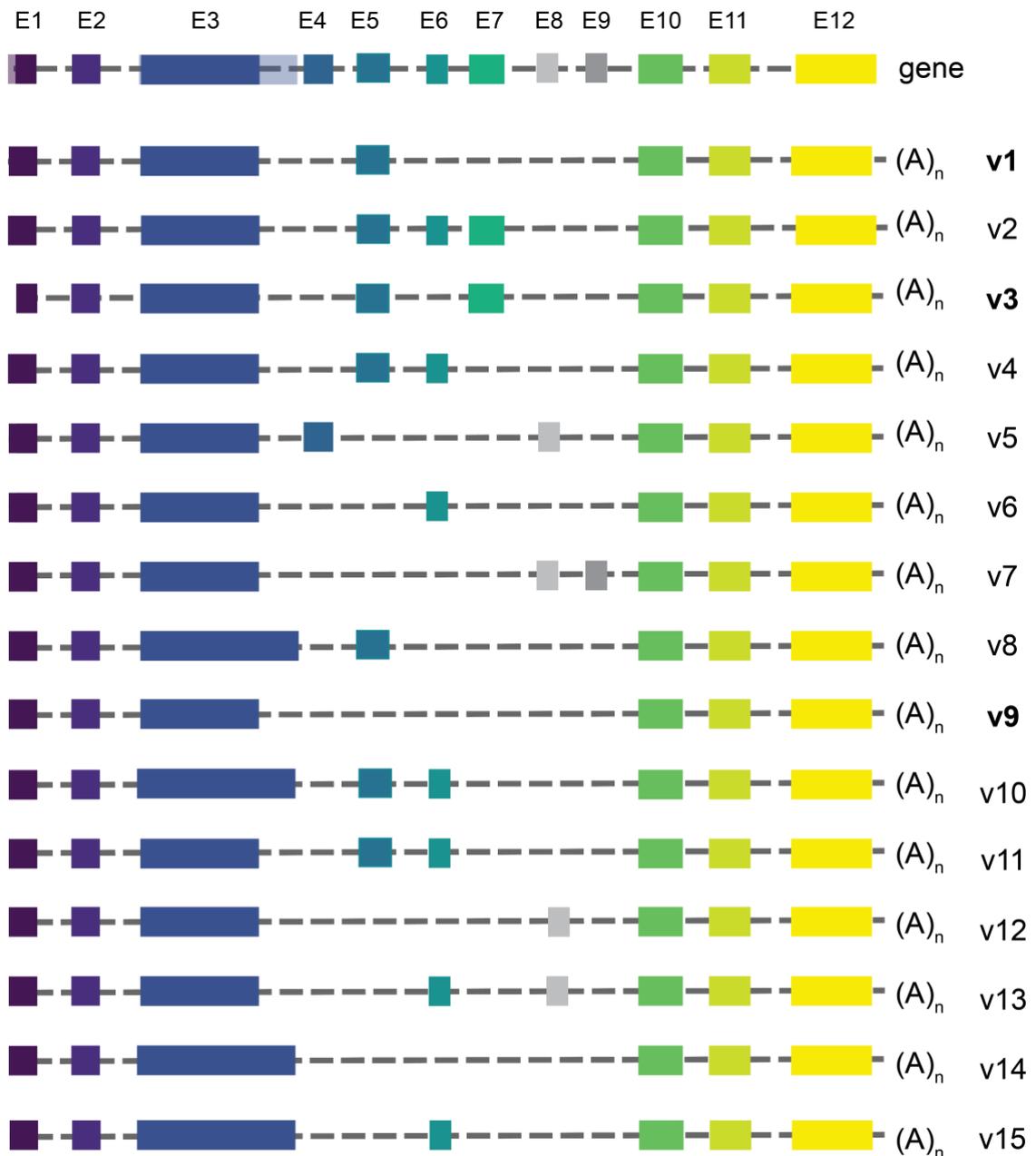


Figure 3:1 *MEG3* gene structure and known splicing isoforms. Graphic representation of the *MEG3* gene exon structure (top) and 15 different splicing isoforms of *MEG3* found in human cells. Isoforms are numbered v1 to v15. Exons are represented by coloured boxes and are labelled E1-E12, the opaque boxes in exons E1 and E3 represent an alternative transcription start site and an alternative 5' splice site respectively, polyadenine tails are denoted by $(A)_n$. Variants highlighted in bold (v1, v3 and v9) are those which will be described in full in this section of my thesis. The NCBI identification numbers of all human *MEG3* isoforms can be found in appendix 5.1.

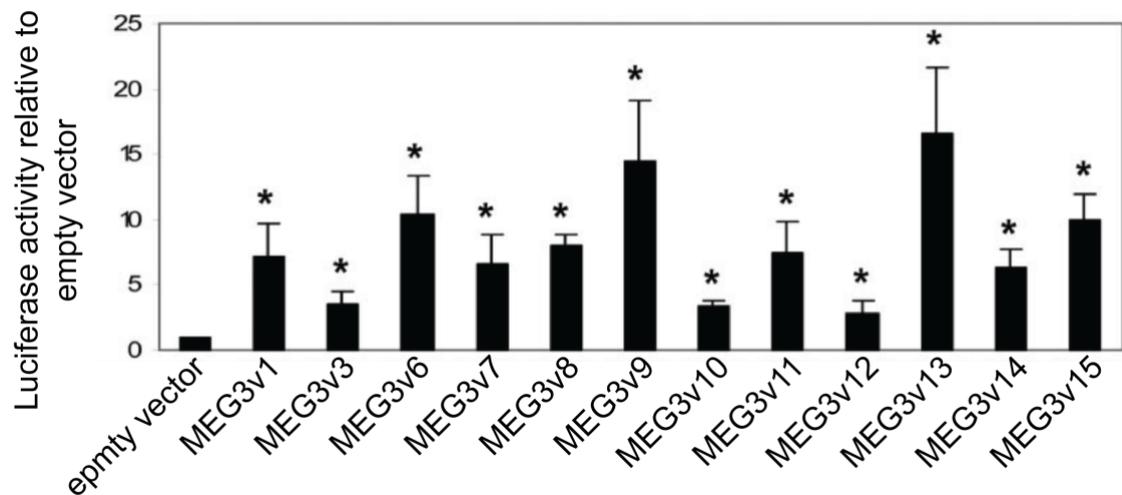
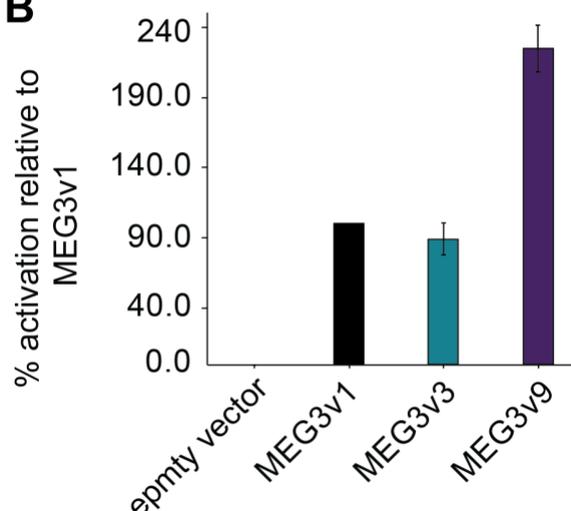
Aadapted from (Zhang *et al*, 2010)**B**

Figure 3:2 Functional characterisation of human *MEG3* splicing isoforms in the activation of p53-dependent gene expression

(A) Activation efficiency of selected naturally occurring human *MEG3* splicing isoforms identified by the Klibanski lab. Figure has been adapted from (Zhang *et al*, 2010). For a conversion table explaining the isoform numbering in the original paper and the updated numbering used in this thesis please refer to (Appendix 5.1). (B) Reproduction of functional assay showing agreement of the results in my hands.

3.1.3 *In vitro* SHAPE probing of *MEG3* variants v1, v3 and v9

How does exon composition affect the molecular architecture of *MEG3*? To answer this question, I purified *MEG3v3* (Figure 3:3) and mapped its secondary structures using *in vitro* SHAPE. In parallel, my colleague Tina Uroda followed the same

protocol for the study of *MEG3v1* and *v9*. We probed the three *MEG3* variants in triplicates (Figure 3:4) using three different SHAPE reagents 1M7, 1M6 and NMIA (see 2.3). The SHAPE reactivity values for all nucleotides across the three *MEG3* variants were consistent as shown by Spearman correlation values ranging from 0.6774 to 0.8662 (Figure 3:4). Our combined results revealed that *MEG3* organised in distinct domains, regardless of the variant (Figure 3:5). We could identify 5 domains in *MEG3v1*, hereafter denoted as D1-D5. Domain D1 comprised of the combined exons E1-E2. Domains D2 and D3 comprised entirely of the 5' and 3' parts of exon E3, respectively. Finally, domains D4 and D5 contained segments of exons E10-E11 and E11-E12, respectively. *MEG3v3* also contained five domains with similar exon compositions as *MEG3v1*. On the other hand, *MEG3v9* organised in only four domains, with exons E10-E12 having re-arranged into a single terminal domain. Overall, differences in exon composition of the three *MEG3* splicing isoforms translated to domain re-organisation.

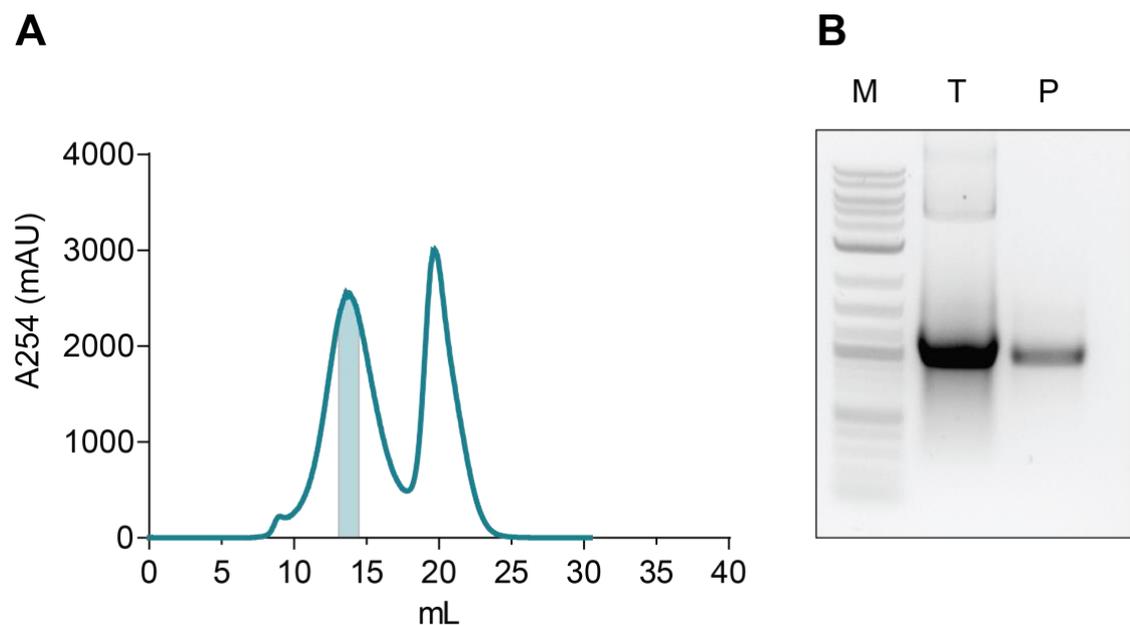


Figure 3:3 Purification of *MEG3v3*

(A) Chromatogram from the size exclusion chromatography step of *MEG3v3* purification. Shaded area represents the chosen peak fraction used for downstream experiments. (B) Agarose gel showing *MEG3v3* purification steps. M = DNA marker, T = transcription reaction, P = chromatogram peak. *MEG3 v1* and *v9* were purified by Tina Uroda and corresponding chromatogram and gel images can be found in (Uroda *et al*, 2019).

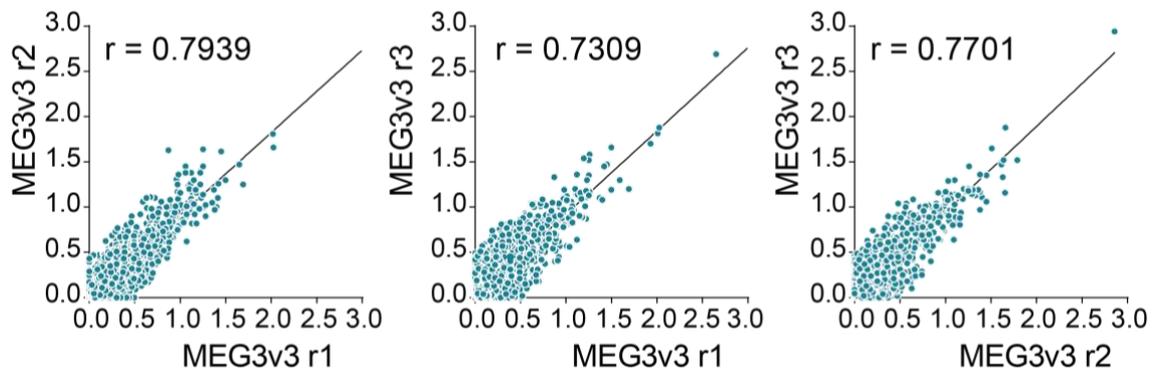


Figure 3:4 Correlation of SHAPE reactivity values across *MEG3v3* between n = 3 independent replicates

Graph shows individual nucleotide SHAPE 1M7 reactivity values across *MEG3v3*. Linear regression was calculated using GraphPad Prism, r values for each correlation analysis are shown in each graph. Corresponding graphs for *MEG3v1* and *v9* can be found in (Uroda *et al*, 2019).

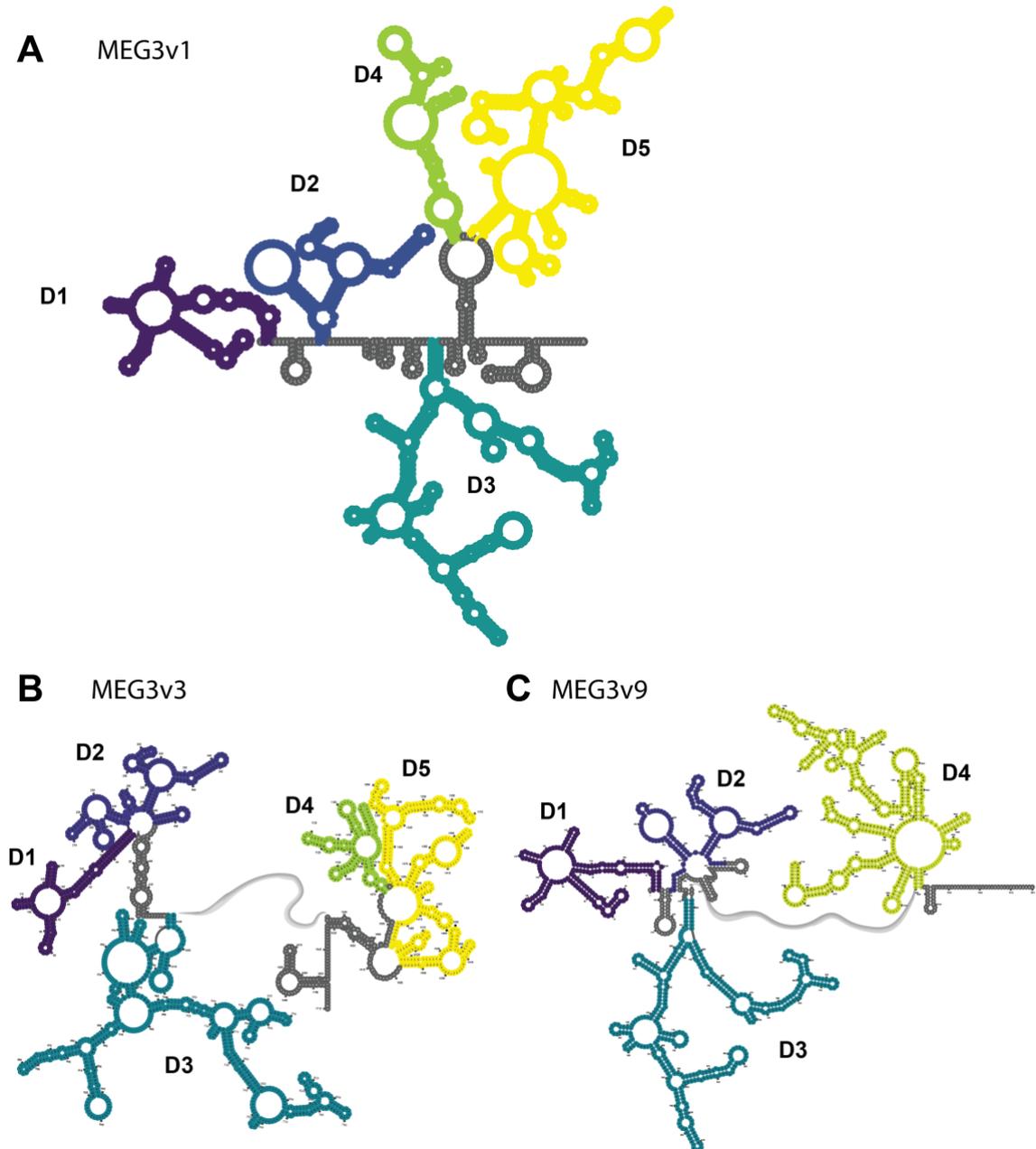
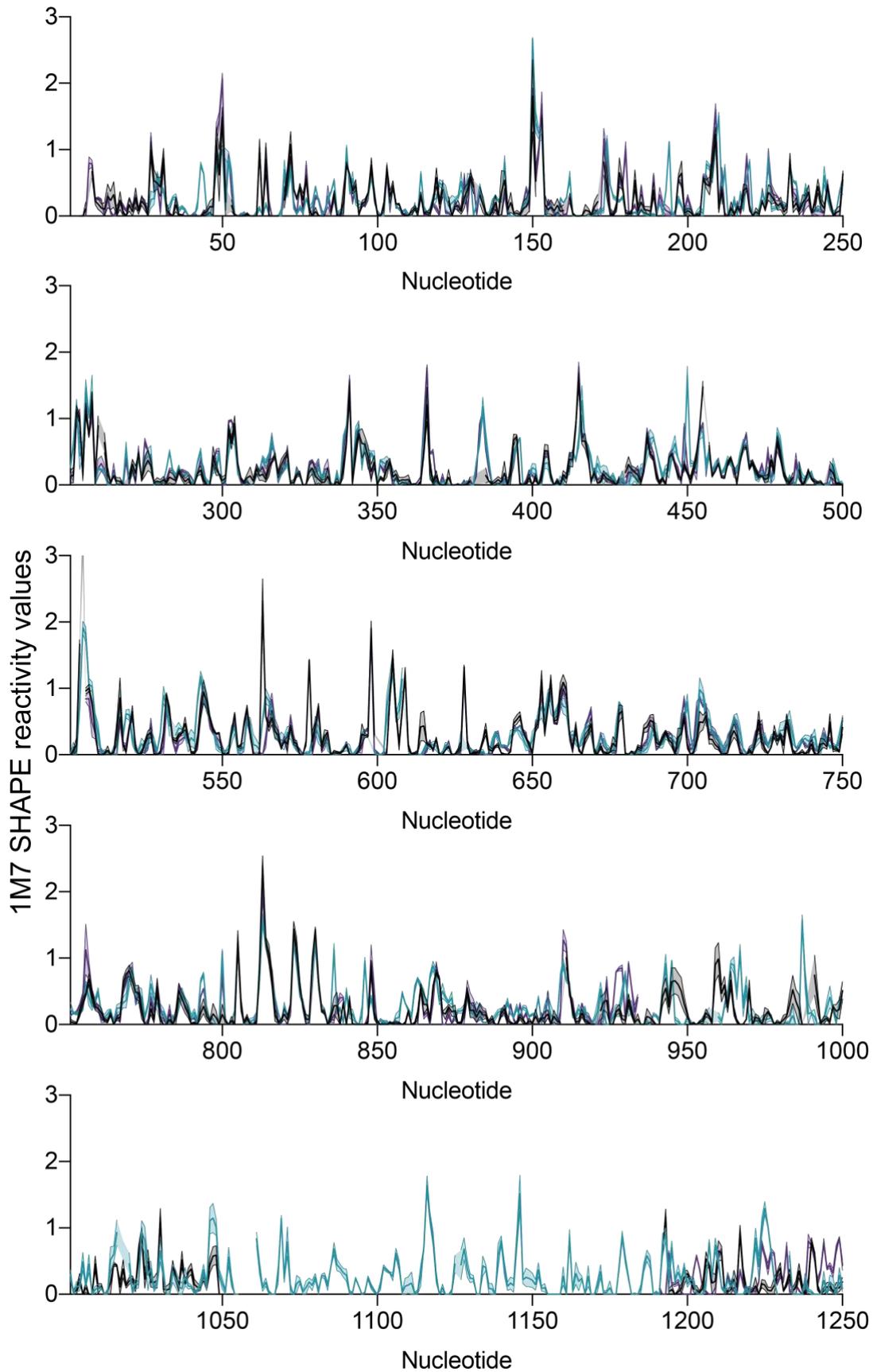


Figure 3:5 Schematic representation of the secondary structure maps of human *MEG3* variants v1, v3 and v9
 (A) *MEG3v1*, (B) *MEG3v3*, (C) *MEG3v9*. The secondary structure maps are coloured to highlight domain organisation. Same colours represent corresponding domains between different *MEG3* variants. Domain D1 is dark purple, domain D2 is blue, domain D3 is teal, domain D4 is green and domain D5 is yellow. In *MEG3v9* where there is no evident distinction between domains D4 and D5 the 3'-end domain (D4) is coloured in lime to designate the merging of domains D4 and D5 as they appear in *MEG3v1*. The maps for *MEG3v1* and v9 were produced by Tina Uroda.

Given the domain modularity of *MEG3*, the functional role of each domain was tested using a cell-based reporter assay (see 2.5) (Zhou *et al*, 2007; Uroda *et al*, 2019). Using this reporter system, my colleague Tina Uroda found that none of the individual domains of *MEG3* was able to drive p53-dependent gene activation in isolation (Uroda *et al*, 2019). However, in all cases where domains D2 and D3 were present, a minimal amount of gene activation was observed (Uroda *et al*, 2019). This observation was also true for the condition where the two domains were encoded on different plasmids, resulting in the two domains being formed by two different transcripts and therefore suggesting that the two domains fold independently of one-another and act co-operatively. Through those experiments, the active core of *MEG3* was mapped on domains D2-D3.

Even though there was undoubtable variation in the domain organisations and molecular architectures of the three *MEG3* isoforms, there were also similarities. Most distinctively, domain D2 of all isoforms contains a structural element corresponding to helix H11 in *MEG3v1* which is structurally identical in all isoforms (Figure 3:5). In *MEG3v1* helix H11 comprised residues C357 to G382. Analysis of the SHAPE reactivity values of the individual nucleotides comprising H11 confirmed this observation (Figure 3:6). Furthermore, in every replicate of all three probed *MEG3* isoforms there was an apparent irregularity in the SHAPE reactivity: the nucleotides of the loop region of helix H11 showed low SHAPE reactivity values (Figure 3:7). Low SHAPE reactivity values suggest residues involved in base-pairing interactions implying that the nucleotides are part of helices and not loops.



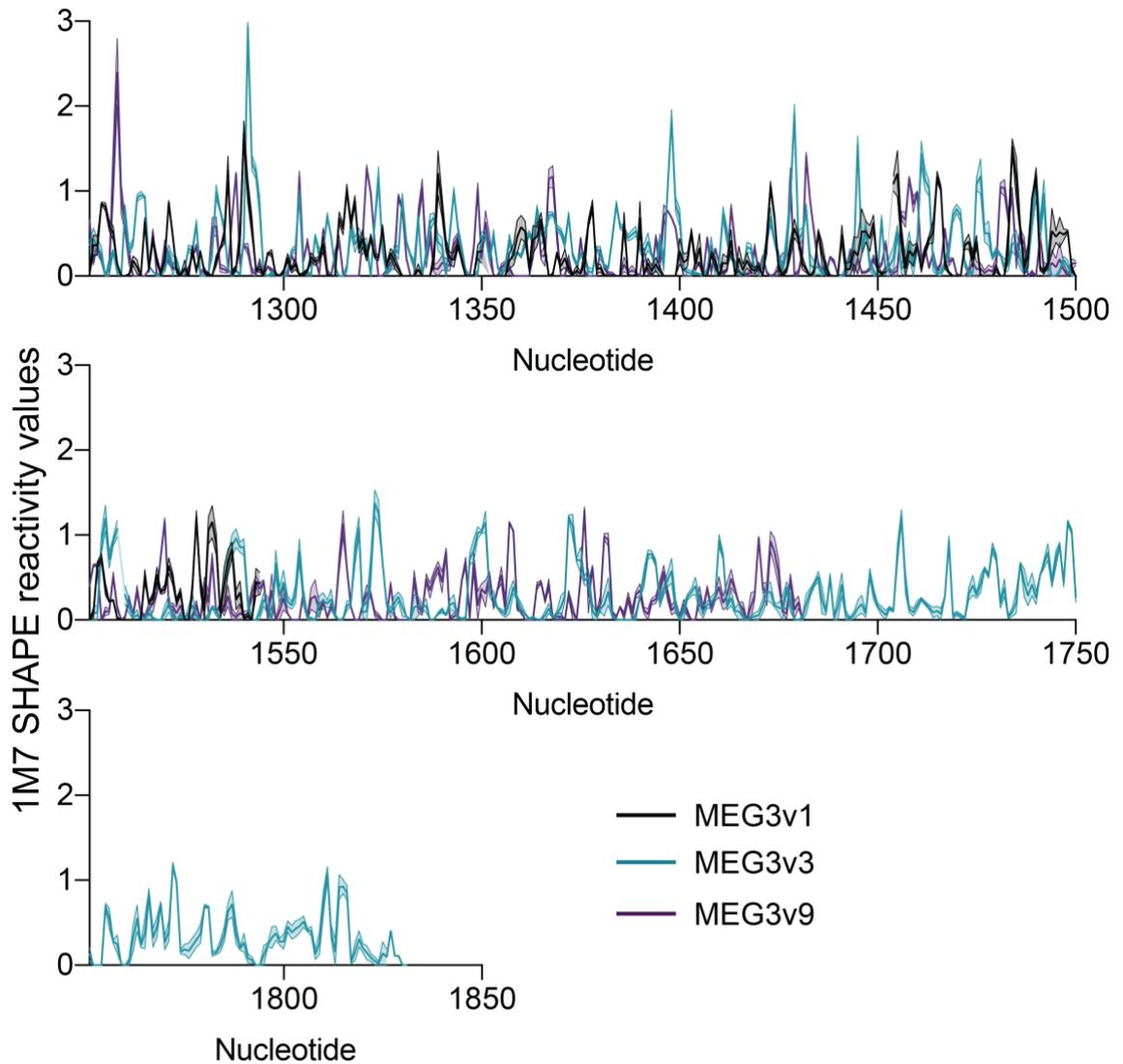


Figure 3:6 1M7 SHAPE reactivities of individual nucleotides in *MEG3v1*, *v3* and *v9*

Graph shows the 1M7 SHAPE reactivities of individual nucleotides across the aligned sequences of *MEG3* variants v1 (black) v3 (teal) and v3 (purple). The solid line in the middle of each trace represents the mean reactivity and the opaque frame is the standard error of the mean from $n = 3$ independent replicates. The SHAPE data for *MEG3v1* and v9 were produced by Tina Uroda.

Taken together, the finding that domains D2 and D3 are essential for the function of *MEG3* in driving p53-dependent gene expression, the structural conservation of helix H11 in all three splicing isoforms and the consistently irregular SHAPE reactivity in the region of H11 was intriguing. We therefore decided to study this region further.

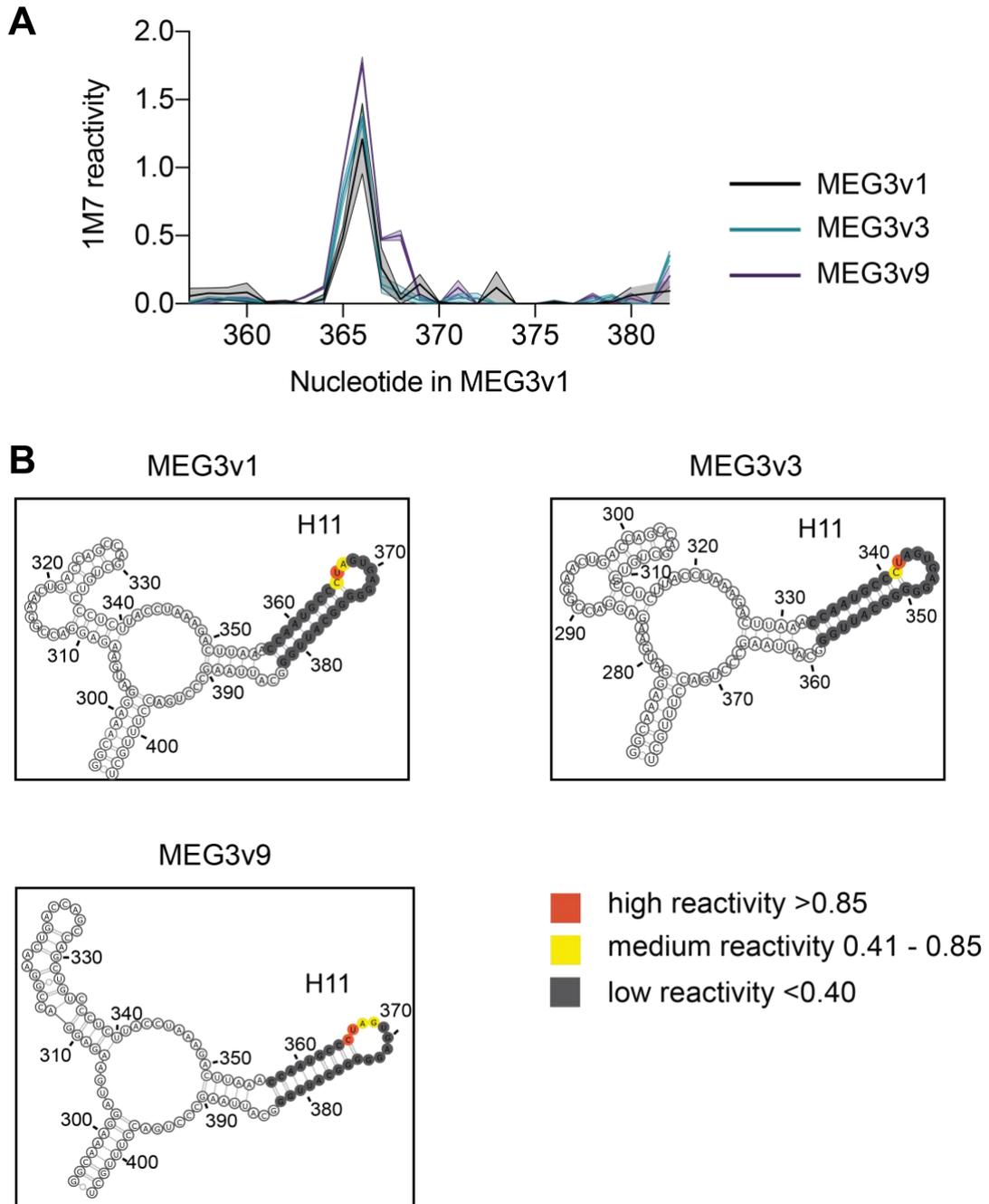


Figure 3:7 Close up on 1M7 reactivities of individual nucleotides in *MEG3*v1, v3 and v9 at the region helix H11

(A) Graph as in Figure 3:6 zoomed in to the region of helix H11. (B) Schematic representation of the secondary structural maps of *MEG3* variants v1, v3 and v9 depicting the region of helix H11 and colour-coded according to individual nucleotide SHAPE reactivity. Grey is SHAPE low reactivity, yellow is medium and red is high reactivity suggesting high, medium and low probability for base-pairing, respectively. The SHAPE data for *MEG3*v1 and v9 were produced by Tina Uroda.

3.1.4 Cell-based functional assays reveal the *MEG3* active core

To investigate the role of helix H11 in more detail, I together with my colleague Tina Uroda, used our cell-based reporter assay (see chapter 2.5) to screen for different mutations targeting H11 on the function of *MEG3* as a regulator of the p53 pathway. Briefly, cells were co-transfected with (1) expression vectors for the *MEG3* construct of interest, (2) *Renilla* luciferase which acts as a means to normalise for transfection efficiency, and (3) a p53 responsive *Firefly* luciferase (Figure 3:8). In the nucleus of the cells, the *MEG3* construct of interest imposes its regulatory effect on the p53-dependent gene expression of the *Firefly* luciferase reporter gene. The luminescence emitted by *Firefly* luciferase upon treatment of the cell lysate with luciferin can be measured and quantified giving a relative numerical value to the relative effect of each *MEG3* construct of interest on p53-dependent gene expression. In the (Uroda *et al*, 2019) paper we sometimes refer to this value as the relative activity of *MEG3*. Tina Uroda showed that deleting H11 (Δ H11) reduced the ability of *MEG3v1* to stimulate p53-dependent gene expression and I could confirm this finding for *MEG3* splicing isoforms v3 and v9 (Figure 3:9). Tina Uroda observed the same when either the structure of helix H11 was scrambled or when the sequence identity of its loop residues was mutated in *MEG3v1* (Uroda *et al*, 2019). Intriguingly, restoring the structure of H11 with complementary mutations but without preserving the sequence identity of the base-pairs rescued the deleterious effect of the scramble mutation (Uroda *et al*, 2019). Thus, this mutagenesis screen identified H11 as an essential active motif in the functional mechanism of *MEG3*-driven p53-dependent gene expression.

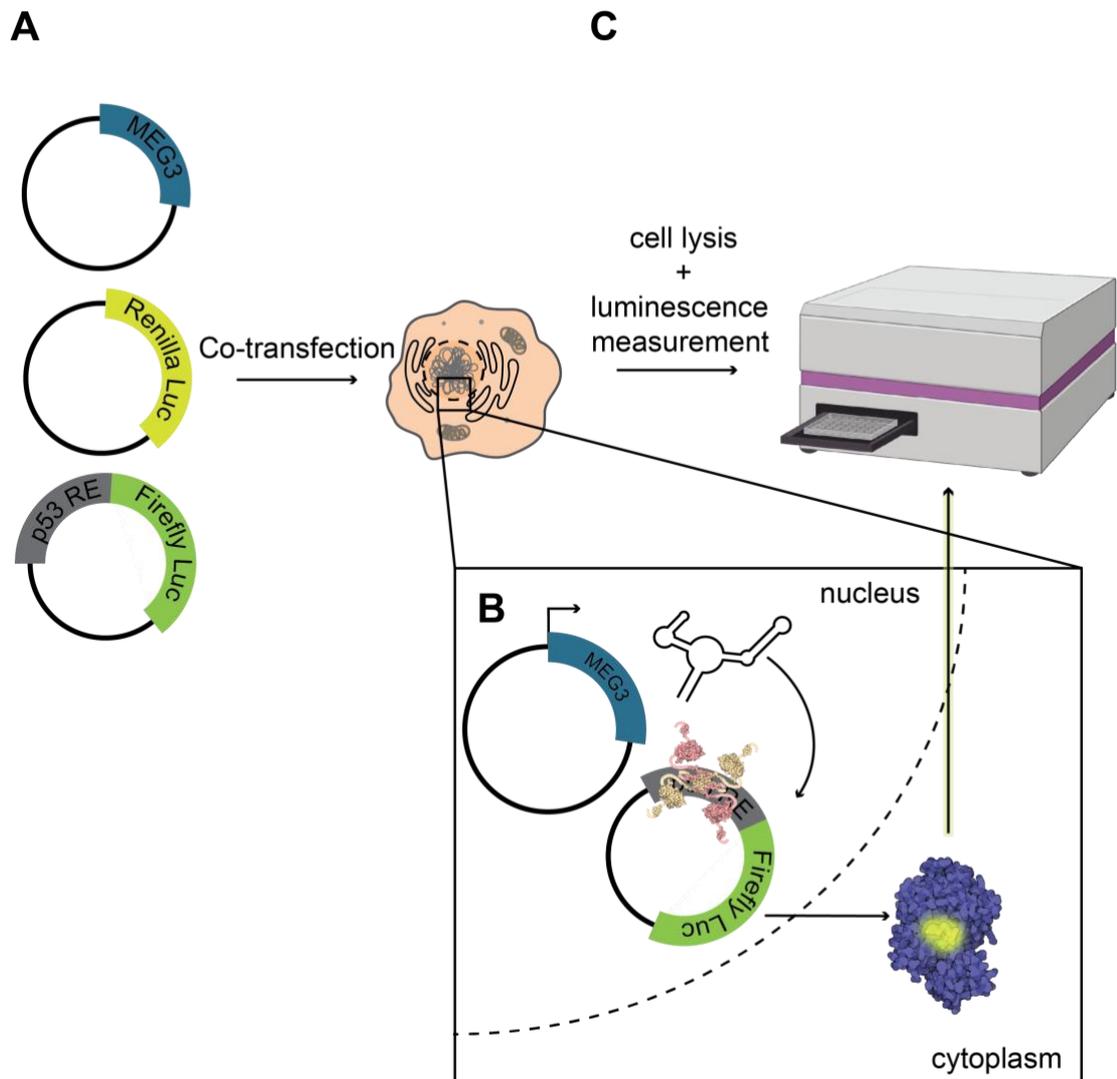


Figure 3:8 Luciferase reporter assay

(A) Cells are transfected with expression vectors for the *MEG3* construct of interest (or empty vector as control), *Renilla* luciferase and a p53 responsive *Firefly* luciferase. (B) Post-transfection, cells are incubated for 48 h to allow production of *Firefly* luciferase. (C) The luminescence emitted by *Firefly* luciferase upon treatment of the cell lysate with luciferin is measured in a plate reader and the amount of *Firefly* luciferase is quantified and normalised to the signal from *Renilla* luciferase. Figure was partly created with Biorender. *Firefly* luciferase and p53 protein cartoons are taken from PDB molecule of the month gallery, PDB IDs 2D1S (luciferase) and 1OLG (p53 tetramerization domain), 1TUP (p53 DNA-binding domain) and 1YCQ (p53 transactivation domain).

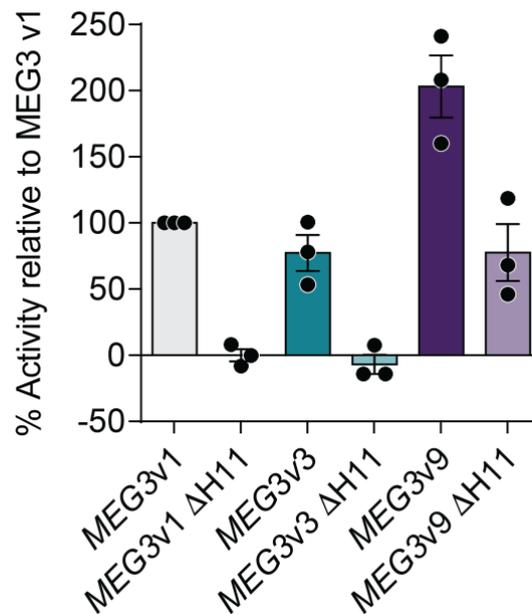


Figure 3:9 Activity assay for human *MEG3* isoforms v1, v3 and v9 and their mutants targeting helix H11

Luciferase reporter assay of human *MEG3* v1 (grey), v3 (teal), v9 (purple) and their mutants lacking helix H11 (in semi-opaque hues). Error bars show standard mean error of $n = 3$ biological replicates.

Given the irregular SHAPE reactivity pattern of the H11 loop residues and the importance of sequence identity in those residues for the function of *MEG3*, my colleague Tina Uroda studied the potential of those residues forming intramolecular long-range interactions. She identified 6 sequence patches located in domain D3, spanning helices H25-H27, that could base-pair with the loop residues of helix H11 forming a pseudoknot motif (Uroda *et al*, 2019). Because those helices were found one right after the other, they were named “*tandem repeats*” (TRs). Upon close investigation of the base-pairing interactions between the loop of H11 and either one of the TRs, it was observed that G370 of H11 was always in the middle of the base-pairs holding the pseudoknot together (Figure 3:10). The G370C mutation which significantly weakens the loop H11-TR interaction was shown to also abolish the gene upregulation effect of *MEG3* when tested using our reporter assay (Figure 3:11). Conversely, complementary mutation of the TRs rescued the mutant phenotype up to a minimal activation level (Figure 3:11). Finally, we found that complementary mutation of each one of the TRs varied in its ability to rescue the functionally deleterious effect of the G370C mutation (Figure 3:11) (Uroda *et al*, 2019).

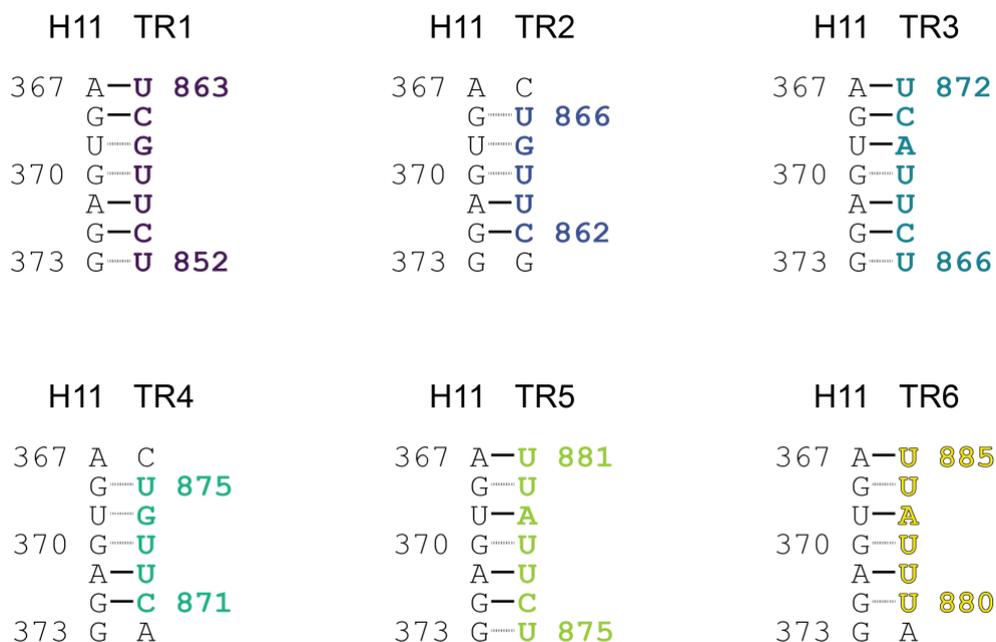


Figure 3:10 Illustration showing the potential interactions forming between each one of the tandem repeats and helix H11

The schematic shows the mutually exclusive interactions that can potentially form between each one of the tandem repeats (TRs) and the loop residues of helix H11. Each TR is colour-coded as follows: TR1 is purple, TR2 is dark blue, TR3 is teal, TR4 is dark green, TR5 is lime and TR6 is yellow. All potential interactions require the central nucleotide G370 of the loop of helix H11.

To study the role of each one of the TRs in driving the functional mechanism of *MEG3* via the formation of the H11-TR pseudoknot motif, I designed a screen where the loop of helix H11 was left intact and point-mutations were performed on the TRs. Mutations on the TRs were designed to weaken the base-pairing interactions holding together the potential pseudoknot with each one of the TRs respectively, thus forcing the pseudoknot to form with a pre-defined set of TRs. Sequential mutations affecting one, two, three, four, five or all six of the TRs were tested with the luciferase reporter system (Figure 3:12). Predictably, the more TRs were compromised by mutation, the less efficient *MEG3* became in activating p53-dependent gene expression (Figure 3:12). By compromising five of the six TRs, I could show that not all TRs had the same efficiency in driving *MEG3*-directed p53-dependent gene expression (Figure 3:12). This finding was in line with the previous observation from the complementary mutant screen targeting G370 of helix H11. Finally, *MEG3* was still capable of driving some gene activation even when all TRs were mutated, suggesting that either there is another active motif aiding the mechanism of *MEG3* in the p53 pathway

independent of the H11-TR pseudoknot or that in the absence of a perfect TR donor, helix H11 may be able to form weak base pairs with the mutated TRs that retain some functionality (Figure 3:12). Taken together these results suggest that the activity of *MEG3* depends on the formation of the pseudoknot interaction between the loop residues of H11 in domain D2 and one of the TRs located in domain D3. Thus, domains D2 and D3 are held together forming the functional core of *MEG3*. Redundancy of the TRs ensures that the lncRNA will retain its functionality in the event of mutation or failure to form the pseudoknot with one particular TR.

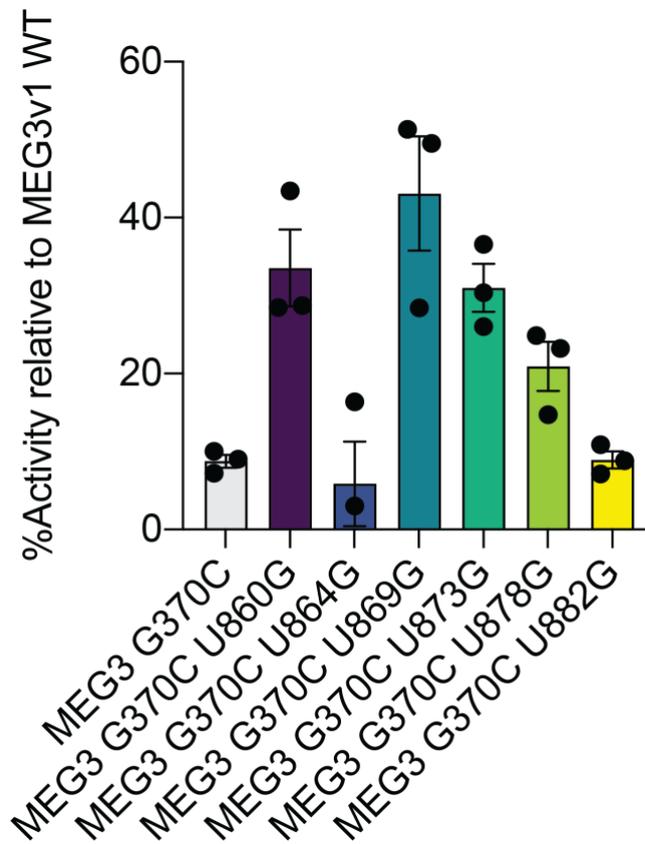
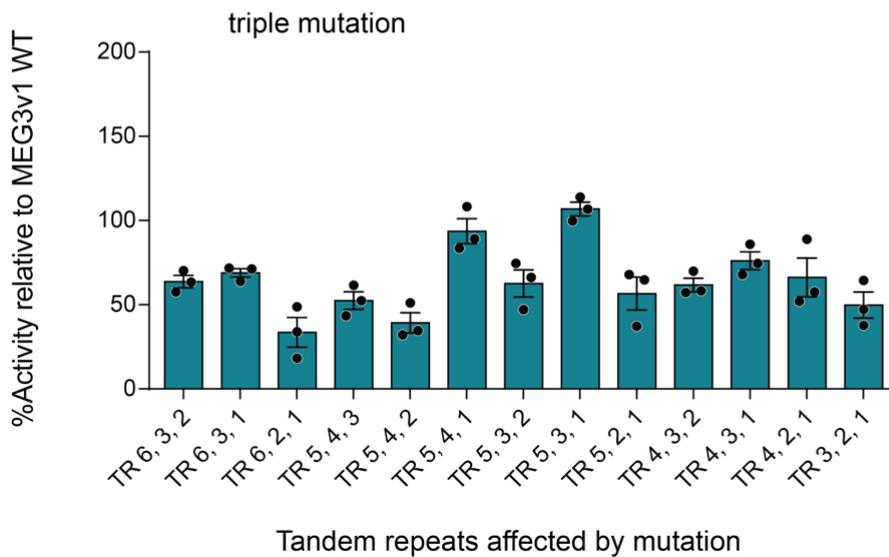
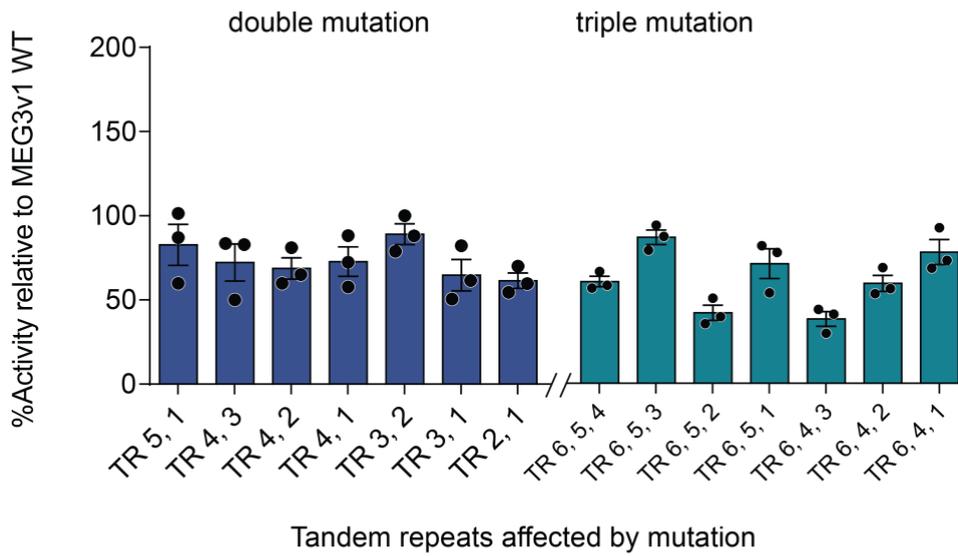
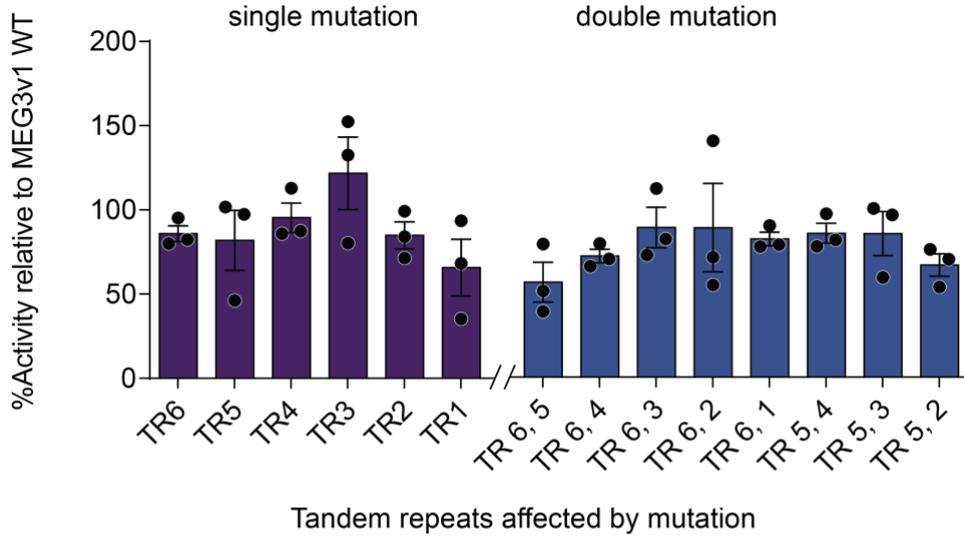


Figure 3:11 Activity assays with *MEG3* mutants targeting the H11-TR pseudoknot interaction

(A) G370C mutation on the loop of helix H11 and corresponding rescue mutations on each one of the tandem repeats (TRs). This experiment was performed by Tina Uroda. Error bars represent standard error of the mean from n=3 independent experiments.

Previously, our group and others had shown that *MEG3* drives p53-dependent gene expression (Zhou *et al*, 2007; Uroda *et al*, 2019). In light of this observation, I decided to study whether choice of the pseudoknot-forming TR in *MEG3* could affect target gene specificity. Thus, I screened the quintuple TR mutants with the luciferase reporter system using a collection of reporter vectors carrying the luciferase gene

coupled to different p53-response elements (Figure 3:13). The results of this screen were analysed by a 2-way ANOVA test (and correcting for multiple comparisons), to assess the significance of the differences in the luciferase signal between the TR mutants and *MEG3v1* WT (Figure 3:13). For the p53Luc and pGL13Luc promoters which are laboratory designed reporter promoters containing multiple copies of optimised p53 response element repeats, TR redundancy was important for the overall potency of *MEG3* in driving p53-dependent gene expression. However, the choice of pseudoknot-forming TR did not greatly affect the specificity of the lncRNA towards different promoters (Figure 3:13). The rest of the tested promoters contain naturally-occurring p53-response elements cloned from genomic DNA (see Appendix 5.2). Those reporters revealed no significant variation in the activator effect of *MEG3* depending on the pseudoknot-forming TR. For the majority of those promoters, the presence of *MEG3* regardless whether this was WT or any of the TR mutants, did not promote p53-dependent gene activation above the control level (Figure 3:13). This observation suggested that the specificity mechanism of *MEG3*-driven p53-dependent gene expression is control by a region of *MEG3* other than the H11-TR pseudoknot. For the 1138 and *MDM2* p53 response elements the choice of pseudoknot-forming TR played a slight role in the level of *MEG3*-driven p53-dependent gene up-regulation. Specifically, when the reporter gene was coupled with the 1138 promoter, formation of the H11-TR pseudoknot via the first three TRs (from 5' to 3') increased gene activation compared to the other three which showed no significant difference when compared to *MEG3v1* WT, which is in conformational equilibrium between alternative TR-donors. *MEG3* preferentially drove gene upregulation from the *MDM2* promoter. When the pseudoknot formed via TR3, gene activation was increased to 4 levels of significance (Figure 3:13). Formation of the pseudoknot by TR4 or TR5 which are very closely located to TR3 showed a slight increase in activation efficiency as well but only up to 1 significant figure. The 5'-most TRs (TR1 and TR2) did not exert any significant difference compared to *MEG3v1* WT while the 3' most (TR6) showed a slight decrease in activation capacity.



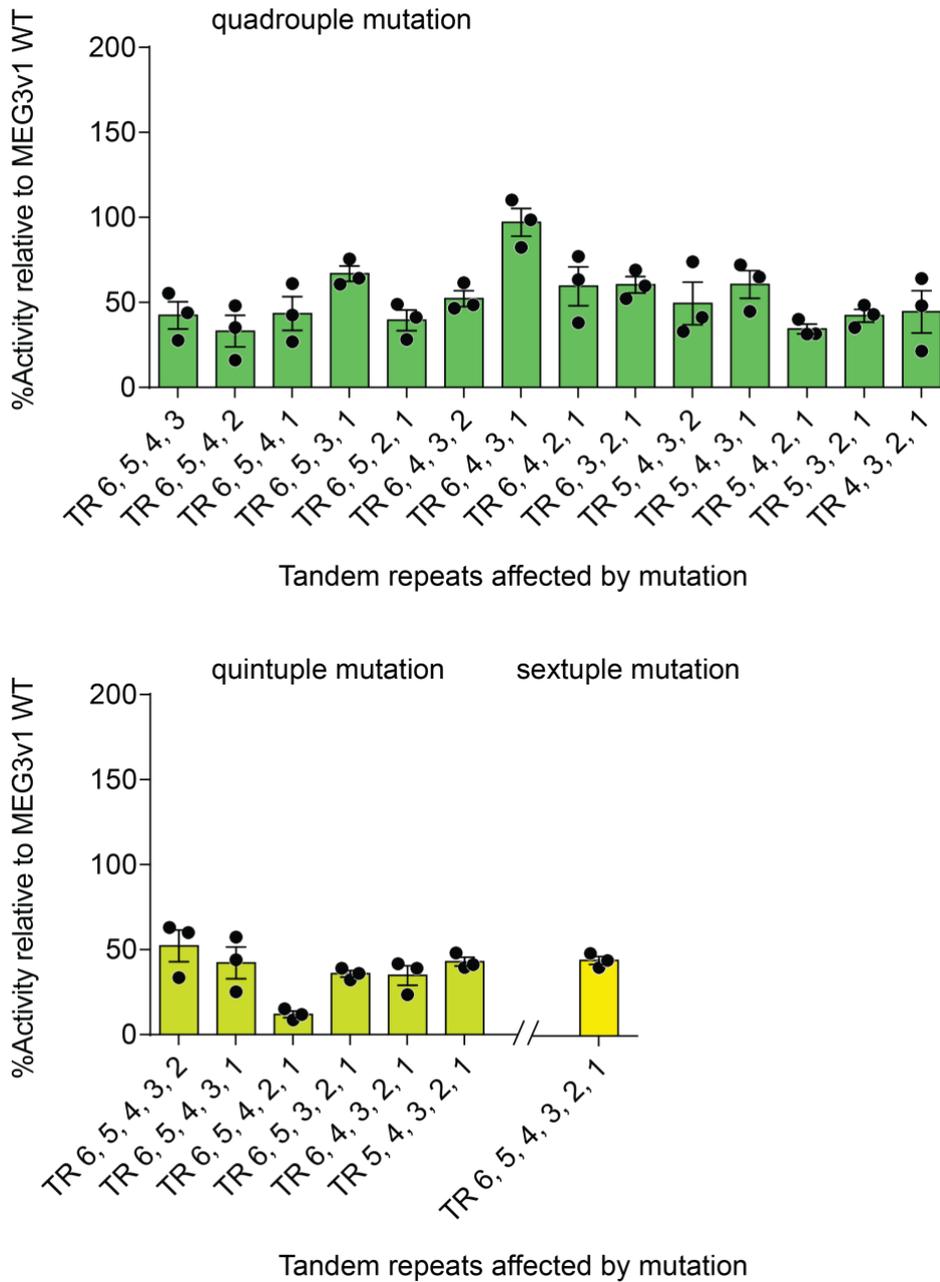


Figure 3:12 Activity assays on *MEG3v1* mutants targeting the H11-TRs interaction by altering the TRs sequence
 Activity of each mutant is expressed as a percentage of the activity of *MEG3v1* wildtype. Error bars represent standard error of the mean from n = 3 independent experiments.

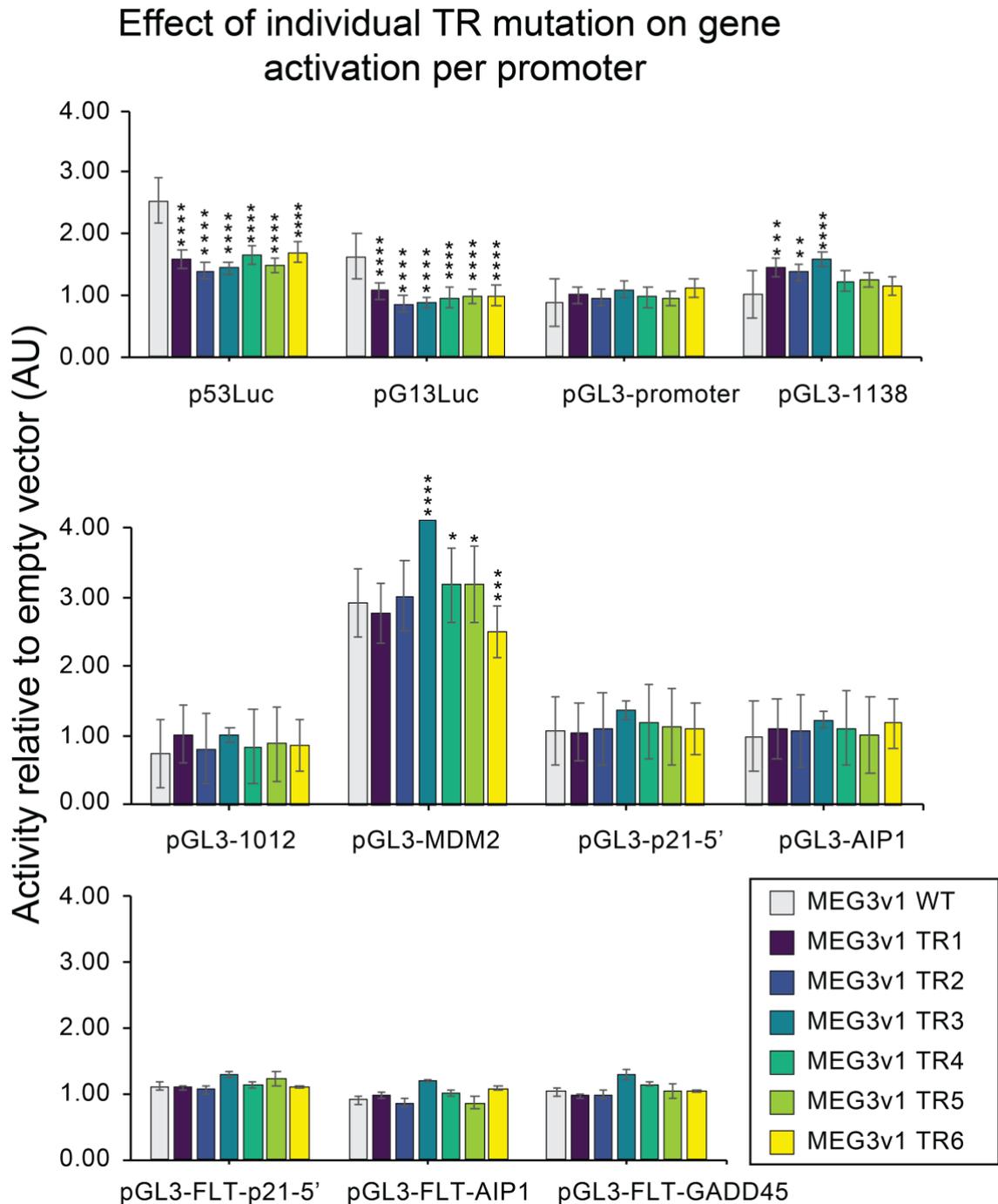


Figure 3:13 Activity assay screening the effect of pseudoknot formation by each tandem repeat on different p53-responsive promoters

Results are represented as relative activation values normalised to the effect of empty vector (no *MEG3* construct) transfection. Results are grouped by the promoter coupled to the reporter gene. Bars are coloured to represent the pseudoknot-forming TR according to the colour legend on the bottom right of the figure. Error bars represent standard error of the mean for $n = 3$ independent experiments. Asterisks represent the p-values from a 2-way ANOVA used to assess significance between the signal of each mutant compared to *MEG3v1* wild-type per promoter ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.

To test whether this apparent difference stems from variation in transcription efficiencies of the transfected vectors, I determined the relative amounts of *MEG3* lncRNA constructs inside cells 48 h post transfection by real-time quantitative PCR (RT-qPCR) (Figure 3:14). This experiment showed that there was no significant variation in the RNA levels produced by each of the vectors carrying the mutant *MEG3* constructs. In conclusion, differences in p53-dependent gene activation observed by the functional reporter assay are therefore not due to transcription efficiency or RNA stability and can be attributed to the molecular properties of the distinct mutant *MEG3* constructs.

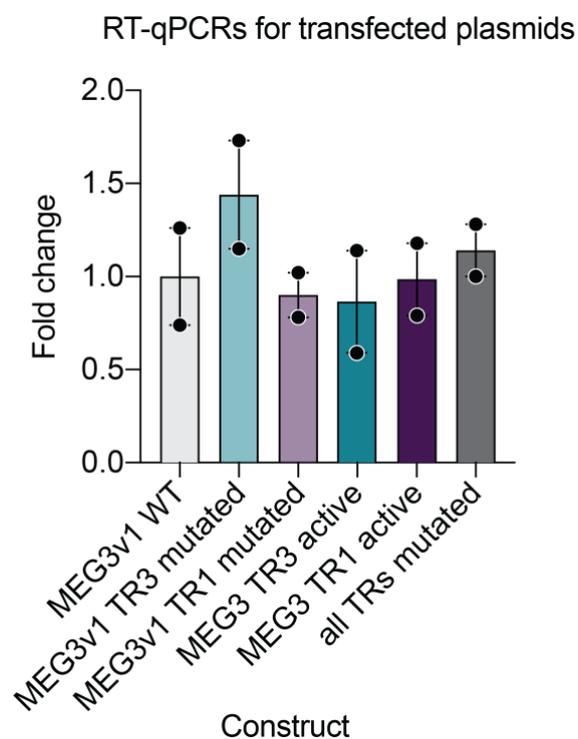


Figure 3:14 RT-qPCR quantification of transfected *MEG3* mutant-expressing plasmids compared to *MEG3v1* wildtype control

Results are plotted as fold change in relative expression levels 48h post transfection compared to the transfected *MEG3v1* wildtype (WT) control. Error bars represent standard error of the mean from $n = 2$ independent experiments.

3.1.5 Chapter Discussion

In summary, the *in vitro* secondary structures of human *MEG3* splicing isoforms v1, v3 and v9 were mapped by SHAPE. Secondary structural probing revealed that *MEG3* organised into four (v9) or five (v1 and v3) distinct domains and exon composition affected the overall arrangement. Careful comparison of the secondary structural maps of the three studied *MEG3* isoforms and coupling of the structural findings with functional characterisation led to identification of the active core of human *MEG3*. The active core is common to all isoforms and is formed by the interaction of the loop residues of helix H11, within domain D2, and one of the six TRs spanning helices H25-H27, within domain D3, forming a pseudoknot motif. A different study on *MEG3v1* suggested a different domain organisation where domain D1 is formed by residues of exons E1 and E12 (Sherpa *et al*, 2018). This finding could be likely an artefact resulting from the computational structure prediction algorithm since long and alternatively spliced RNAs most commonly fold co-transcriptionally. Nonetheless, the structural map by Shepra *et al* (2018) includes a structure for helix H11 identical to ours even though the surrounding structural elements are very different. Moreover, the corresponding SHAPE reactivity values from that study follows the trend of very low SHAPE reactivity values across the loop residues on helix H11. Thus, data from different laboratories support the formation of long-range interactions by the loop residues of helix H11. We showed that formation of the pseudoknot is both necessary and sufficient for the activity of *MEG3* in gene expression regulation via p53. Finally, through experiments performed by my colleagues in the Marcia lab and our collaborators in Alberto Inga's lab, it was shown that helix H11 is not a target site for any RNA-binding proteins, suggesting that the functional core of *MEG3* is formed by the lncRNA only (Uroda *et al*, 2019).

Our results support that the functional core of *MEG3* existed in a conformational equilibrium where the H11-TR pseudoknot interaction could form via any one of the six TRs. One more experiment that could potentially shed light on whether there is bias towards the use of one TR over the others would be psoralen analysis of RNA interactions and structures (PARIS) (Lu *et al*, 2016). PARIS utilises the psoralen derivative 4'-aminomethyltrioxalen which is a potent RNA-RNA crosslinker to identify *in cellulo* RNA tertiary structural elements such as long-range structures and pseudoknots (Lu *et al*, 2016). Notably, because the TRs are so close to each other, this

method may not be sensitive enough to distinguish between adjacent TRs and therefore we did not include this experiment in our initial set of tools for the characterisation of *MEG3*. Recently, the Weeks lab proposed a brand new method for the identification of long range interactions (e.g. pseudoknots) termed selective 2'-hydroxyl acylation analysed by primer extension and juxtaposed merged pairs (SHAPE-JuMP) (Christy *et al*, 2021). SHAPE-JuMP is a variation of SHAPE where the chemical modifier used to probe the RNA crosslinks base-paired nucleotides creating a “bridge” across long range pairs. SHAPE-JuMP utilises a specially engineered reverse transcriptase that can “jump” over the crosslinked bridges, connecting long range pairs. Hence, long range pairs appear as continuous sequences in the downstream Illumina sequencing library. This method could be used to provide further evidence for the formation of the H11-TR pseudoknot and most importantly to identify additional potential long-range interactions which we could then characterise functionally by the luciferase reporter system. Furthermore, SHAPE-JuMP could be used to identify long-range interactions unique to specific *MEG3* splicing isoforms which would help elucidate the source of the functional differences between various *MEG3* isoforms.

The choice of TR for formation of the pseudoknot was found to moderately regulate the overall activity level of *MEG3*. Notably, when the activity of *MEG3* was tested against a collection of different p53-responsive promoters, it showed preferential activation of some promoters over others which was in line with previous research suggesting that *MEG3* regulated p53 target gene specificity (Zhou *et al*, 2007; Uroda *et al*, 2019) (Figure 3:13). For some promoters the choice of pseudoknot-forming TR slightly affected the efficiency of *MEG3*-driven gene activation observed. Taken together, these results suggest that in the regulatory mechanism of p53-dependent gene activation via *MEG3*, the choice of the pseudoknot-forming TR could adjust the level of gene activation achieved. Perhaps presence or absence of alternative exons in the isoforms of *MEG3* could bias domain organisation favouring a certain set of TRs for pseudoknot formation thus leading to differences in the functions of the isoforms. Naturally, this is only a speculation and more experiments exploring the role of each TR in the functional profiles of other *MEG3* variants need to be performed before a solid conclusion can be drawn.

In this screen, the reporter coupled to the MDM2 response element was amongst the highest-activating reporters. MDM2 is a ubiquitin ligase enzyme targeting p53 protein for proteasomal degradation. Previously, it was suggested that *MEG3* acts as a tumour suppressor via downregulating MDM2 (Zhou *et al*, 2007). My findings presented here as well as those previously published by our group (Uroda *et al*, 2019) contradicted this hypothesis as our assay showed a high level of gene activation from the *MDM2* promoter in the presence of *MEG3*. This finding might hint towards a regulatory loop where excess p53 activity guided by *MEG3* becomes down-regulated by consequential accumulation of MDM2 protein and subsequent marking of p53 for degradation. There are several points to be addressed in this hypothesis. My experiment showed that *MEG3*-driven p53 activation had a preference for the *MDM2* promoter in a system where only the *MDM2* promoter was coupled to the reporter gene and transfected into cells in excess. In nature however, all p53 target genes are present in the genome simultaneously and the target choice for gene activation is not limited by availability or promoted by excessive presence of the reporter as was the case in my assay. Hence, even though my assay was useful in delineating the role of each pseudoknot-forming TR in the mechanism of *MEG3*-driven gene targeting for p53-dependent activation, its results should not be overgeneralised.

Despite the potential roles of individual TRs in fine tuning target gene preferences, all six TRs were confirmed as potential base-pairing partners for the loop of H11 and the six alternative conformers were all functional since compensatory mutations complementing the G370C mutation in either one of the TRs could rescue the G370C phenotype up to an extent. This apparent redundancy mechanism can be explained by an evolutionary perspective. Non-coding genes are under lower evolutionary pressure than protein-coding genes (Ulitsky, 2016). Thus, there is a higher probability that over time one of the TRs becomes mutated. If there was only one sequence patch that can form the functional core of *MEG3* by the pseudoknot interaction and that patch were mutated, then the entire lncRNA would lose its function. However, if there are multiple copies of potential pseudoknot donors clustered together in tandem, then the position of the active core would likely shift slightly but the molecule would retain its function. Shifting of the active core via the choice of pseudoknot-forming TR did not dramatically affect the specificity

mechanism of *MEG3* towards its target gene promoters. Hence, the active core of *MEG3* has a robust mechanism with the potential to persist random mutation, maintaining its function and specificity through redundancy of the TRs.

Altogether, my own and Tina Uroda's work elucidated the secondary structural organisation of 3 human *MEG3* isoforms and presented clear links between the structure and the function of the lncRNA within the p53 pathway. Our collective structural work led to the identification of the functional core of *MEG3* formed by domains D2-D3. Tina Uroda's work provided to the core of the data that led to the identification of our suggested *MEG3* mechanism depending on the interaction of the loop residues of helix H11 with the TRs. These results were published in (Uroda *et al*, 2019) where I am second author. For my PhD I also dwelled further into the details of the H11-TR interaction dissecting the potential role of each pseudoknot-forming TR in the regulation of translation from different p53 REs. These results have not yet been made public.

3.2 Linking the active core of *MEG3* to the tertiary structural organisation of the lncRNA

3.2.1 Résumé en français

Dans le chapitre précédent, j'ai décrit comment le noyau actif de *MEG3* a été identifié. Le noyau actif est formé par l'interaction pseudo-note entre l'hélice H11 et les TRs. Comme mentionné précédemment, des expériences complémentaires réalisées par ma collègue Tina Uroda et nos collaborateurs du laboratoire d'Alberto Inga en Italie, ont montré que H11 n'est pas impliqué dans les interactions protéine-ARN. Ainsi, la question en suspens dans l'étude du mécanisme fonctionnel de *MEG3* est «comment le motif pseudo-note H11-TR conduit-il l'activité de la *MEG3*?» Dans ce chapitre, je m'attarde sur l'exploration des détails structuraux de l'activité de *MEG3*. Je décris comment j'ai mis en place une approche d'empreinte hydroxyle-radicalaire (HRF) pour étudier le repliement tertiaire global de *MEG3* à une résolution nucléotidique individuelle. Les résultats des expériences décrites ici suggèrent que *MEG3* est un lncRNA globulaire en présence d'ions divalents à une concentration physiologique et qu'un repliement correct dans sa forme globulaire est essentiel pour son rôle fonctionnel dans la régulation de l'expression génétique.

3.2.2 Chapter introduction

In the previous chapter I described how the active core of *MEG3* was identified. The active core is formed by the pseudoknot interaction between helix H11 and the TRs. As mentioned earlier, complementary experiments performed by my colleague Tina Uroda and our collaborators in Alberto Inga's lab in Italy, showed that H11 is not involved in protein-RNA interactions. Thus, the outstanding question in the study of the functional mechanism of *MEG3* is "*how does the H11-TR pseudoknot motif drive MEG3 activity?*" In this chapter I dwell further into exploring the structural details driving *MEG3* activity. I describe how I set up a hydroxyl-radical foot-printing (HRF) approach to study the overall tertiary fold of *MEG3* at individual nucleotide resolution. The findings of the experiments described herein, suggest that *MEG3* is a globular lncRNA in the presence of divalent ions at physiological concentration and that correct folding to its globular shape is essential for its functional role in gene expression regulation.

3.2.3 The H11-TR pseudoknot affects the overall folding profile of *MEG3*

To study the effect of the formation of the H11-TR pseudoknot on the overall tertiary structure of *MEG3*, I used hydroxyl-radical foot-printing (HRF) (see 2.3.8). HRF is a technique that has been used before to study the fold of structured RNAs such as the *Tetrahymena* ribozyme (Chance *et al*, 1997) and the hairpin ribozyme (Hampel & Burke, 2001). Briefly, the method depends on treatment of the RNA with highly reactive oxygen radicals ($\bullet\text{OH}$) which are generated *in vitro* through the Fenton reaction of Fe(III)-EDTA ions with hydrogen peroxide (H_2O_2) (Tullius & Dombroski, 1986). Upon contact with RNA, $\bullet\text{OH}$ radicals attack the ribose sugar at the C4' position, decomposing the sugar and breaking the phosphodiester bond connecting two consecutive nucleotides in an RNA chain (Hampel & Burke, 2001). Thus, when the generated RNA fragments are reverse-transcribed and the reverse transcription products mapped on the full-length sequence of the RNA, the contact sites between the RNA and the $\bullet\text{OH}$ can be inferred. Provided that the reaction with the $\bullet\text{OH}$ is limited to a very short period of time and that the RNA of interest folds consistently into a stable globular shape, the $\bullet\text{OH}$ will preferably react with the nucleotides found on the periphery, solvent-exposed surface of the globularly folded RNA whilst those nucleotides that are buried in its core will be protected from $\bullet\text{OH}$ -mediated degradation. Thus, fragment ends in the reverse transcription fragment library correspond to nucleotides found on the periphery of the globularly folded RNA.

HRF had never been used to study a lncRNA before, presenting my work with a significant challenge. At the start of my PhD, the RNAs that had been characterised by HRF were significantly shorter than *MEG3*. For instance, the 50 nt-long hairpin ribozyme is more than 30 times smaller than *MEG3v1* (1,595 nt) (Hampel & Burke, 2001). In the example of the hairpin ribozyme, $\bullet\text{OH}$ treatment was performed on radioactively labelled RNA and subsequent fragment length analysis was done by running the labelled RNA fragments directly on an acrylamide gel along Sanger sequencing ladders and visualising the gel on film (Hampel & Burke, 2001). The length of *MEG3* combined with the potential health hazards of working with radioactive isotopes, render this approach nearly impossible for the study of lncRNAs. Hence, I set off to establish a modification of the existing protocols that utilises capillary electrophoresis to analyse fluorescently labelled reverse transcription reaction products. I then re-purposed the pre-existing programme

QuSHAPE (Karabiber *et al*, 2013) to analyse the frequency of •OH-mediated cleavage across the full-length of *MEG3*. QuSHAPE was originally used for the analysis of SHAPE fragments separated by capillary electrophoresis. Essentially QuSHAPE performs comparative fragment length analysis between a control and an experimental sample and therefore could be re-purposed for the analysis of any kind of RNA probing experiment coupled to fragment separation by capillary electrophoresis.

HRF probing of *MEG3v1* in triplicates produced highly reproducible results (Figure 3:15) suggesting that *MEG3v1* folded in a consistent tertiary structure where some nucleotides were preferably found on the periphery of the molecule whilst some others were buried in its core. Focusing on H11, it was found that the 5' part of the helical stem was constantly exposed to the solvent whilst the loop and the 3' side of the stem were protected from reaction with the •OH radicals (Figure 3:17 panel A). The patched appearance of H11 in this context is compatible with the hypothesis that the H11-TR pseudoknot interaction locks the entire lncRNA in a conformation where only the 5' part of the helical stem is part of the exterior shell of the lncRNA while the site of the interaction, namely the loop of helix H11 and the downstream 3' side of the helical stem are rotated and thus buried in its core.

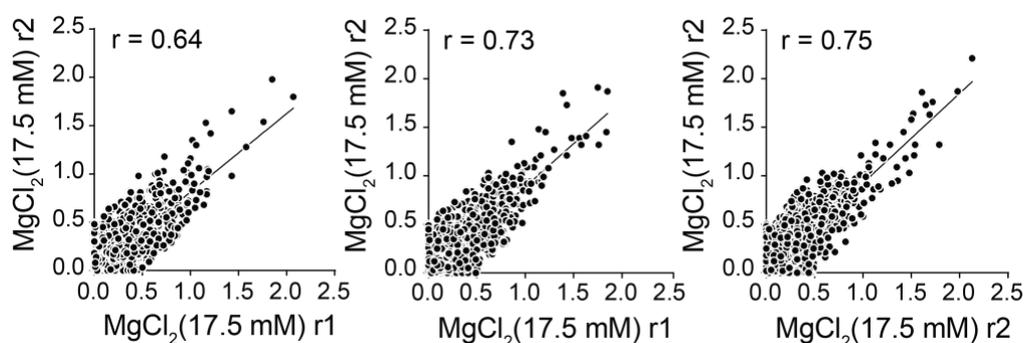


Figure 3:15 Correlation of 3 replicates of *MEG3v1* hydroxyl-radical probing at 17.5 mM $MgCl_2$

Correlation of HRF reactivity values was assessed for each individual nucleotide after normalisation of the values (see methods) using the spearman test. R values are shown on the top left corner for each graph.

By comparison of the HRF traces of the helix H11 loop poly(A) mutant and of *MEG3v1* wild-type (WT), I analysed the change in solvent accessibility of each individual nucleotide along the length of *MEG3v1* between the functional

(pseudoknot formed) and dysfunctional (no pseudoknot) states of the lncRNA (Figure 3:16). I thus, identified several continuous stretches of *MEG3*, across its entire length which gained solvent accessibility upon breakage of the H11-TR pseudoknot. This finding supported the hypothesis of the H11-TR pseudoknot tying together the global structure of *MEG3*.

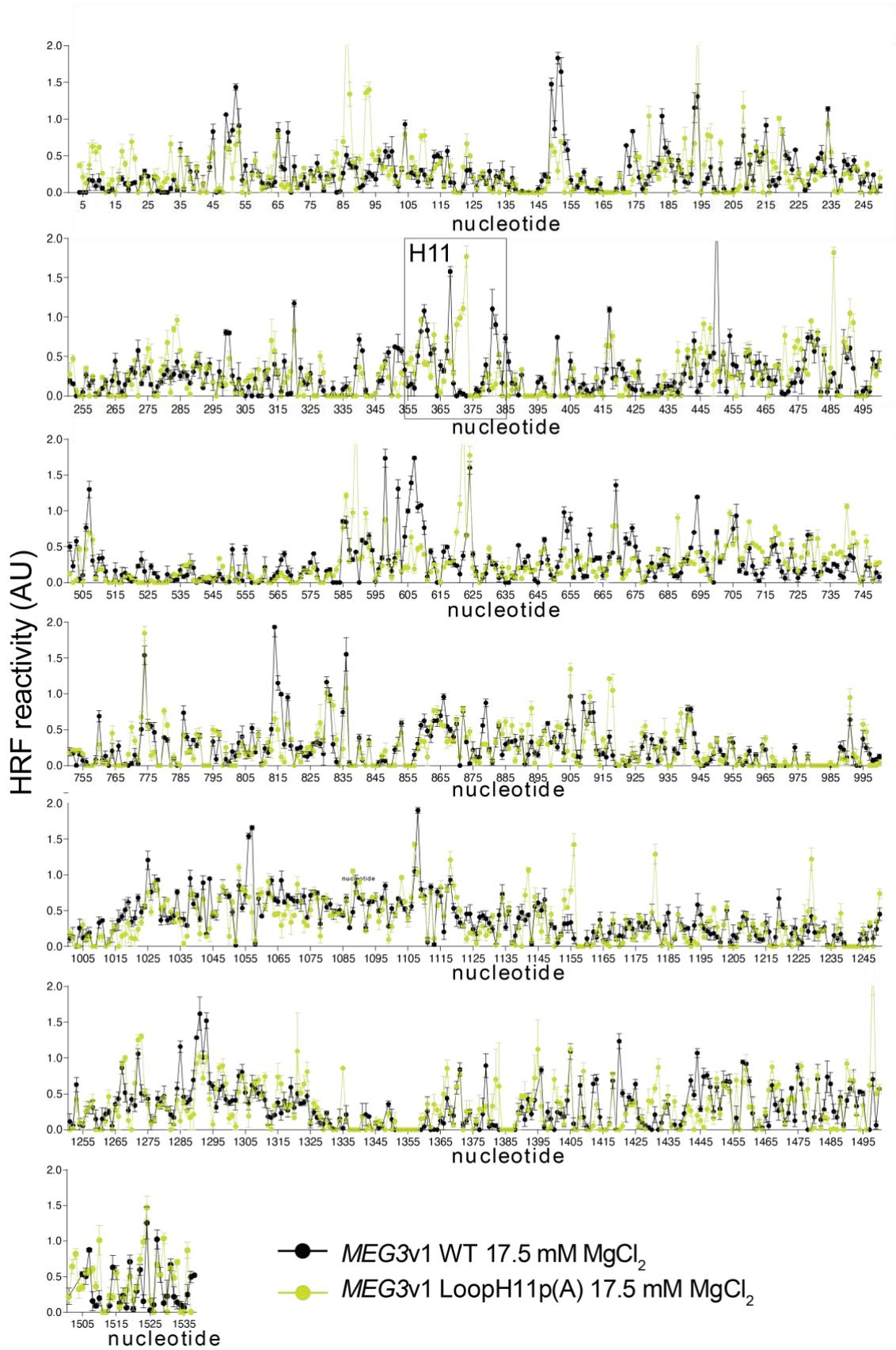


Figure 3:16 Hydroxyl-radical foot-print of individual nucleotides across wild-type *MEG3v1* and the helix H11 loop poly(A) mutant

Graph depicting HRF reactivities of individual nucleotides across *MEG3v1* (black) and the helix H11 loop poly(A) mutant (lime green). Error bars represent standard error of the mean from n = 3 individual replicas.

Upon mutation and subsequent failure to form the H11-TR pseudoknot, the loop of helix H11 as well as the 3' side of the helical stem gained solvent exposure (Figure 3:17 panel B). This finding provided strong evidence supporting the previous hypothesis about the role of the H11-TR pseudoknot in holding the tertiary conformation of the entire lncRNA in place. In parallel, my colleague Tina Uroda studied *MEG3v1* WT and the helix H11 loop poly(A) mutant by atomic force microscopy (Uroda *et al.*, 2019, 2020). Her finding provided visual evidence of the dependency of *MEG3* tertiary structure on the H11-TR interaction. Taken together, our collective results showed that formation of the H11-TR pseudoknot motif was essential not only for the function of *MEG3* but also for correct tertiary fold of the lncRNA. As mentioned earlier, neither helix H11 nor any of the TRs were target-sites for RBPs (Uroda *et al.*, 2019). Thus, for the first time, we established a clear connection between a pure RNA tertiary structural element in a lncRNA and its biological function.

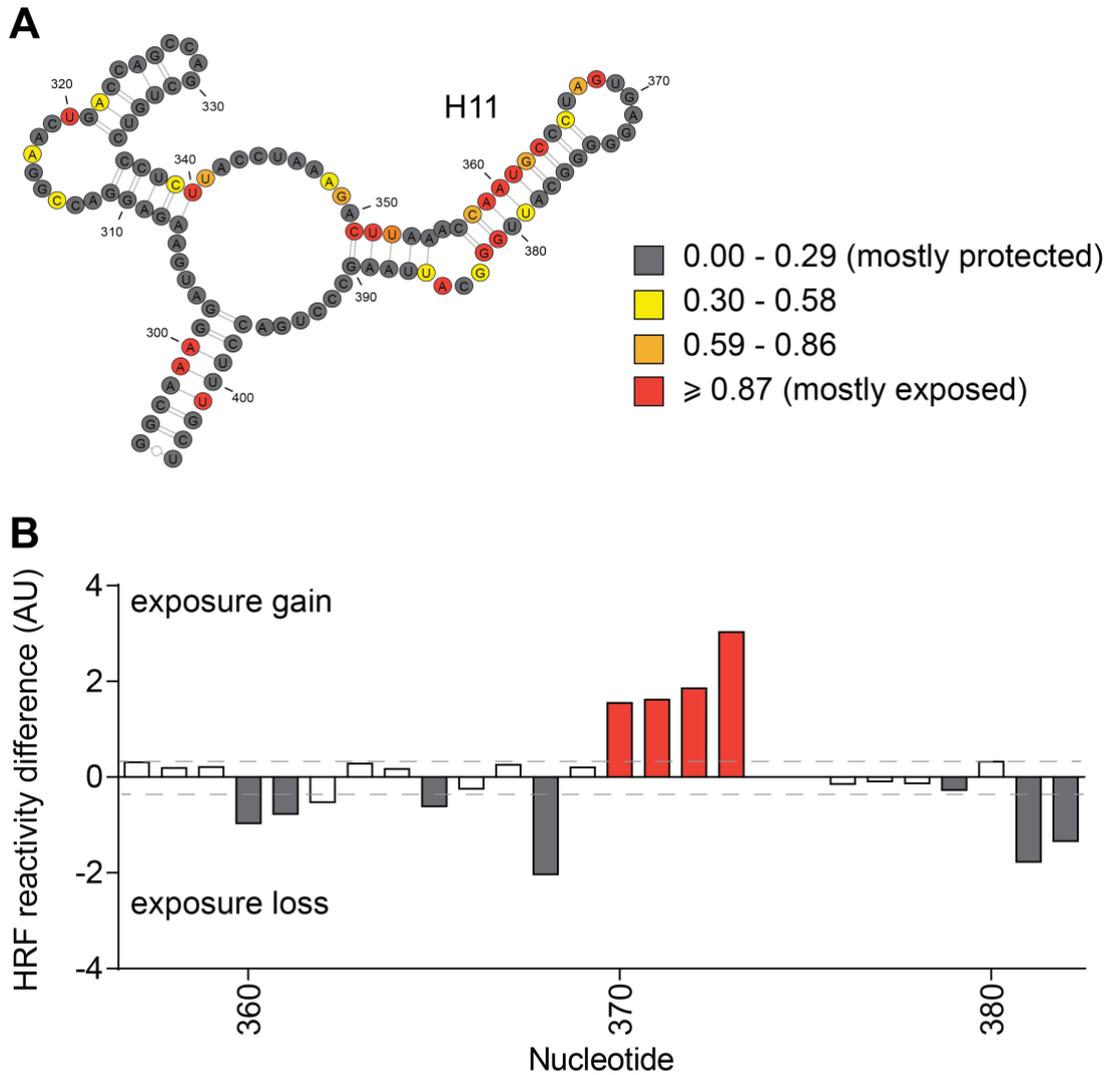


Figure 3:17 Hydroxyl-radical reactivity values of helix H11

(A) Individual nucleotide HRF reactivity values in the region around helix H11. Solvent exposure of the nucleotides was determined according to its HRF reactivity where nucleotides with reactivity values from 0.00 – 0.29 are considered protected (grey), 0.30 – 0.58 are slightly exposed (yellow), 0.59 – 0.86 are moderately exposed (orange) and values greater than 0.87 are highly exposed. (B) Detailed graph showing the difference in HRF reactivity values between the helix H11 poly(A) mutant and wild-type *MEG3v1*. Positive values in red represent nucleotides that have gained solvent exposure upon mutation, negative values in grey represent nucleotides that became more protected from exposure upon mutation. The dashed lines represent the margins for insignificant difference in solvent accessibility between the helix H11 poly(A) mutant and wildtype *MEG3v1* and are represented by bars that are left white.

3.2.4 Chapter discussion

In this chapter I described how I established a protocol for HRF for the identification of solvent accessible nucleotides in fully folded *MEG3*. With this technique, I showed that the relative 3-dimensional organisation of domain D2 depends on establishment of the pseudoknot motif between helix H11 and the TRs which are located on domain D3. With this finding, I showed a clear connection between the secondary and tertiary structures of *MEG3* since the arrangement of domain D2 depended on its interaction with another domain, namely domain D3. Furthermore, since non-pseudoknot-forming mutant of *MEG3* also fails to upregulate p53-dependent gene expression compared to WT *MEG3v1*, I could connect the tertiary structural organisation of *MEG3* with its functional mechanism.

Another method, light activated structural examination of RNA by high throughput sequencing (LASER-seq), for the identification of solvent accessible nucleotides in RNA has been developed (Feng *et al*, 2018; Zinshteyn *et al*, 2019). LASER-seq uses light activated aroyl azides to probe the RNA. Aroyl azides react with the electron-rich purines of RNA (adenosine and guanosine), forming adducts that induce mutation incorporation upon reverse transcription and thus sequencing results from LASER-seq can be analysed by mutagenesis analysis (Feng *et al*, 2018). As a next step we could use LASER-seq to characterise the solvent accessibility of *MEG3* inside cells. Following an approach similar to Δ SHAPE, where *in vivo* SHAPE traces are subtracted from *ex vivo* traces to assess changes in RNA base-pairing in the presence and absence of proteins (Smola *et al*, 2015), a predicted protein-binding site map could be constructed. Such an approach would yield very useful information about functional mechanism of *MEG3*.

Notably, LASER-seq is not without its limitations. Aroyl azides selectively react with purines, this means that this method will inevitably be biased towards purine-rich solvent-exposed regions of the RNA (Feng *et al*, 2018; Zinshteyn *et al*, 2019). The HRF map described in this chapter provides an unbiased trace of solvent-exposed residues in terms of base identity. Comparing *in vitro* methods for assessing solvent accessibility, HRF is more advantageous than LASER-seq because it is not base-biased. HRF can also be applied in living cells; however this requires applying synchrotron radiation to the sample in a very complex experimental set-up and has so far only been done on bacteria because the bacterial cell wall allows the cells to

sustain the radiation shock for the duration of the experiment (Hulscher *et al*, 2016). Hence, despite its base-identity bias, LASER-seq would be a better approach for the *in vivo* characterisation of solvent accessible regions in *MEG3* or other lncRNAs but for the *in vitro* approach that I chose here, HRF provides valuable unbiased information. Having established HRF for the study of lncRNAs, a next step in the study of *MEG3* could be to perform the HRF experiment with *MEG3* variants v3 and v9. It would be interesting to focus on the differences between the three variants especially in the regions comprising domains D4 and D5 which appear re-organised in terms of their secondary structures. This approach could indicate potential alternatively exposed regions which could act as protein docking sites explaining the differences in the functional profiles between the three variants.

3.3 Towards a high-resolution structure of the *MEG3* core

3.3.1 Résumé en français

Jusqu'à présent, dans cette thèse, j'ai fourni des preuves solides soutenant que *MEG3* est un ARN structuré avec un lien clair entre sa forme 3D et son effet fonctionnel dans la cellule. Ma collègue, Tina Uroda, a fait un premier pas dans cette direction en étudiant *MEG3v1* par chromatographie d'exclusion de taille et diffusion de lumière multi-angle (SEC-MALLS), diffusion de rayons X aux petits angles (SAXS) et microscopie à force atomique (AFM) (Uroda et al, 2019). Un autre collègue, Manikandan Karuppasamy, a fait les premiers efforts pour étudier *MEG3v1* par cryo-microscopie électronique (cryo-EM). Grâce à leurs efforts collectifs, mes collègues ont jeté les bases d'autres études structurales de *MEG3*, que j'ai eu pour objectif d'approfondir au cours de mes études de doctorat. Dans ce chapitre, je décris ma contribution à la compréhension des règles guidant la formation de la structure tertiaire de *MEG3*, afin de faire un pas de plus vers l'exploration de *MEG3* à haute résolution. Je décris d'abord mon travail de révélation des positions des ions divalents dans *MEG3*, qui sont importantes pour la formation de la structure tertiaire. Je décris ensuite mes efforts pour produire des constructions biologiquement pertinentes de *MEG3* pour des études cristallographiques et je discute des défis impliqués dans l'acquisition d'une structure à haute résolution de lncRNAs tels que *MEG3*.

3.3.2 Chapter introduction

Thus far in this thesis, I have provided strong evidence supporting that *MEG3* is a structured RNA with a clear link between its 3D shape and its functional effect in the cell. My colleague Tina Uroda made a first step towards this direction by studying *MEG3v1* by size exclusion chromatography and multi-angle light scattering (SEC-MALLS), small angle X-ray scattering (SAXS) and atomic force microscopy (AFM) (Uroda *et al*, 2019). Another colleague, Manikandan Karuppasamy, made initial efforts to study *MEG3v1* by cryo-electron microscopy (cryo-EM). Through their collective efforts, my colleagues lay the ground for further structural studies of *MEG3* which I aimed to further during my PhD studies. In this chapter, I describe my contribution towards understanding the rules guiding tertiary structure formation of *MEG3* aiming to get one step closer to exploring *MEG3* at high resolution. I first describe my work revealing the positions of divalent ions in *MEG3*, which are important for tertiary structure formation. I then describe my efforts to produce biologically relevant constructs of *MEG3* for crystallographic studies and discuss the challenges involved in acquiring a high-resolution structure of lncRNAs such as *MEG3*.

3.3.3 Terbium foot-printing reveals divalent ion binding sites essential for tertiary structure formation.

Divalent ions are essential for the tertiary shape of RNAs (Pyle, 2002). As we previously showed by analytical ultra-centrifugation, *MEG3* compaction correlates with the concentration of Mg^{2+} ions in the buffer (Uroda *et al*, 2019). In addition, as we visualised by atomic force microscopy, *MEG3* does not adopt its compacted globular structure in the absence of divalent ions (Uroda *et al*, 2020). As a further step in the exploration of the role of divalent ions in *MEG3* 3D-folding, I used Terbium (III) probing (see 2.3.9) to determine the nucleotides involved in coordinating those structural ions. The ionic radius of Tb^{3+} (0.92 Å) is very similar to Mg^{2+} (0.72 Å) (Saito & Suga, 2002). Both ions coordinate oxygen ligands in a (1, 23, 24) geometry. However, Tb^{3+} is more electrophilic than Mg^{2+} and thus readily replaces Mg^{2+} in a folded RNA when added to a sample of the RNA (Harris *et al*, 2004). Finally, the pK_a of lanthanide (II) ions, including Tb^{3+} , is very close to pH 7 and thus Tb^{3+} will readily cleave nucleic acids at the site of ion coordination under physiological conditions (Pyle *et al*, 2000). Thus, Tb(III) probing has been used for the identification of divalent ion coordination sites in globularly folded RNAs such as the aiy group II intron (Pyle *et al*, 2000) or the Hepatitis Delta Virus ribozyme (Harris *et al*, 2004). As it was the case with HRF, Tb(III) probing had been used in the past to study RNAs that are a lot smaller than *MEG3* and therefore the entire RNA was radioactively labelled and the generated fragments were directly run on a sequencing gel (Pyle *et al*, 2000; Harris *et al*, 2004). In order to adapt Tb(III) probing to my target (*MEG3v1*), I once again re-purposed the programme QuSHAPE (Karabiber *et al*, 2013) to assess Tb(III)-mediated cleavage.

In order to identify true ion coordinating nucleotides in folded *MEG3v1*, I produced the Tb(III) probing maps of *MEG3* folded in the presence of Mg^{2+} at 17.5 mM ((+) $MgCl_2$ sample) concentration and in the absence of any divalent ions ((-) $MgCl_2$ sample). Both the (+) $MgCl_2$ the (-) $MgCl_2$ samples were treated with 1.25 μM Tb (reaction sample) or with water (control sample). The Tb(III) footprint of each folding state of *MEG3v1* was determined after extracting the control sample peaks from the reaction sample profile using QuSHAPE. The footprint obtained from the (-) $MgCl_2$ sample represented the background, non-specific pattern of reverse transcription reaction arrest. Hence, the final Tb(III) probing map of human *MEG3v1* was

generated by extracting the (+) MgCl_2 sample footprint from this of the (-) MgCl_2 sample. The final map of Mg^{2+} ion binding was calculated as the average of three independent replicas (Figure 3:18). Notably the overall correlation of the Terbium probing footprint was not suggestive of consistent results. However, the correlation of highly reactive nucleotides after background deduction suggested that the identified ion-binding nucleotides were true hits (Figure 3:18). The discrepancy amongst the low-reacting nucleotides across individual replicas is likely due to the highly reactive nature of Tb^{2+} ions an aqueous solution. A total of 31 nucleotides across the entire length of *MEG3v1* were identified as ion-coordinating nucleotides (Table 3:1).

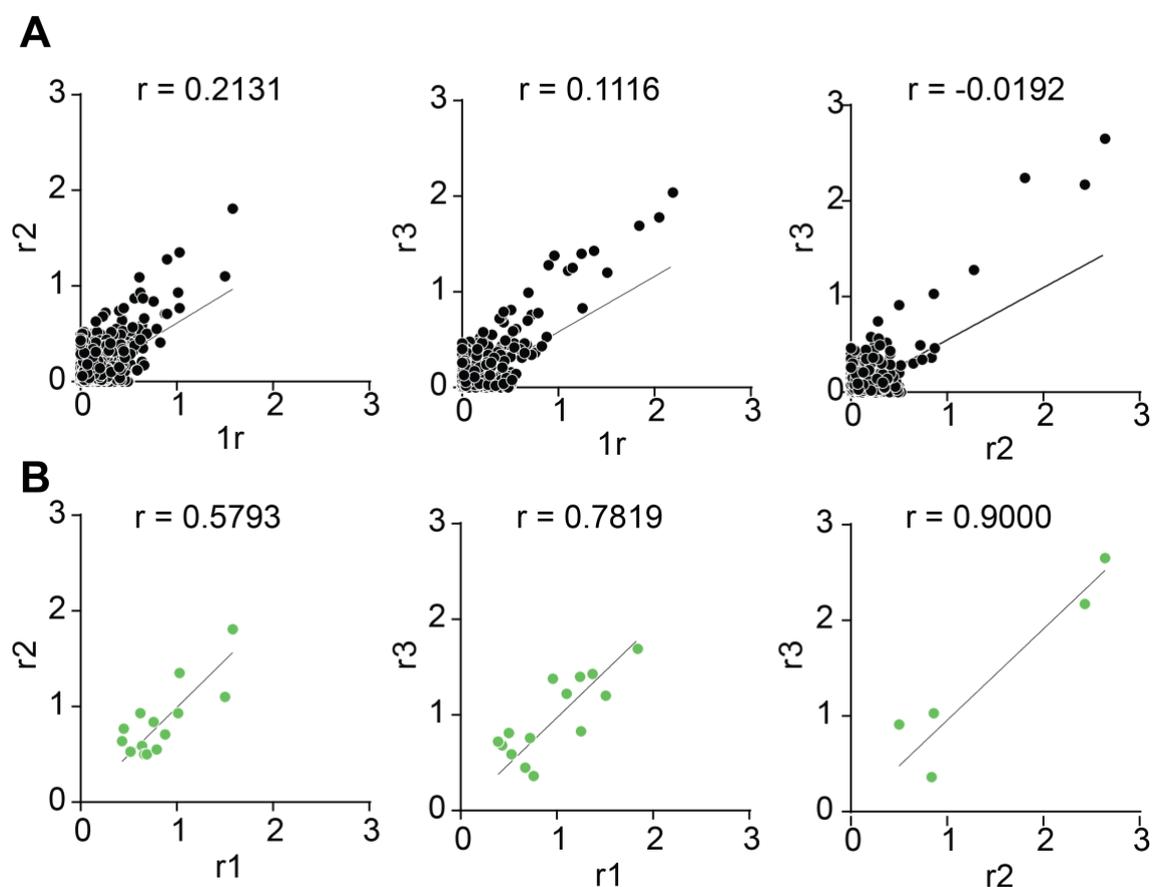


Figure 3:18 Correlation of 3 replicas of the Terbium probing experiment

Correlation of reactivity to Tb(III) ions was assessed for each individual nucleotide after normalisation of the values (see methods) using the spearman test. R values are shown on top of each graph. Panel (A) shows the correlation of replicas across the entire *MEG3v1* and panel (B) shows the correlation of the high-scoring nucleotides only.

Table 3:1 Nucleotides involved in coordinating divalent ions

	Position	Base	Structural element	Domain
1	51	C	stem	D1
2	67	C	stem	
3	68	U	stem	
4	166	U	stem	
5	169	U	stem	
6	192	G	stem	
7	266	A	loop	D2
8	270	A	loop	
9	334	U	stem	
10	373	G	loop	
11	374	G	stem	
12	393	C	junction	D3
13	415	U	linker	
14	497	U	stem	
15	530	C	stem	
16	531	U	stem	
17	533	A	junction	
18	543	G	loop	
19	588	C	stem	
20	626	C	stem	
21	639	U	stem	
22	800	G	stem	
23	814	A	loop	
24	855	C	stem	
25	921	C	stem	
26	970	U	bulge	
27	980	U	bulge	
28	981	C	stem	D5
29	1158	G	loop	
30	1235	C	loop	
31	1357	C	loop	

3.3.4 Manipulating the H11-TR pseudoknot motif can tweak the activity profile of *MEG3* in the p53 pathway

From Manikandan Karuppasamy's studies of *MEG3v1* by cryo-EM it became apparent that *MEG3* contains flexible regions preventing 2D classification and averaging steps required for solving the cryo-EM map of *MEG3* (Uroda *et al*, 2020). In addition, the flexible nature of *MEG3* makes it an unlikely candidate for crystallisation. Considering that the active core of *MEG3* forms by the interaction of domains D2 and D3, studying the structure of any domain in isolation would not produce any physiologically relevant information. Hence, I set off to design truncated *MEG3* constructs with minimised alternative conformers but maintaining functional activity. I designed constructs for crystallisation trials of *MEG3* using *MEG3v1* domains D2-D3 as base and used our lab's luciferase reporter system to assess the activity of the constructs.

I thus purified domains D2-D3 of *MEG3* and set up crystallisation trials using the high throughput crystallisation facility at EMBL Grenoble (Figure 3:19). This effort however, was unsuccessful most likely due to the high flexibility and structural heterogeneity of the construct. Hence, I designed a construct domains D2-D3 with deleted helix H8 (D2-D3- Δ H8) since previously Tina Uroda had shown that helix H8 is dispensable for the activity of *MEG3* and because it carries a large unstructured stem-loop, I expected it to hinder crystal formation (Figure 3:19). Unfortunately, no crystals formed suggesting that further mutations to stabilise the molecule were required. As a proof of principle, I tested the homogeneity of the samples by running a Bioanalyzer RNA 6000 Nano chip which provides significantly higher resolution than agarose gel electrophoresis (Figure 3:19). Electrophoresis by the Bioanalyzer revealed several bands suggesting the simultaneous presence of at least three distinct structural isoforms.

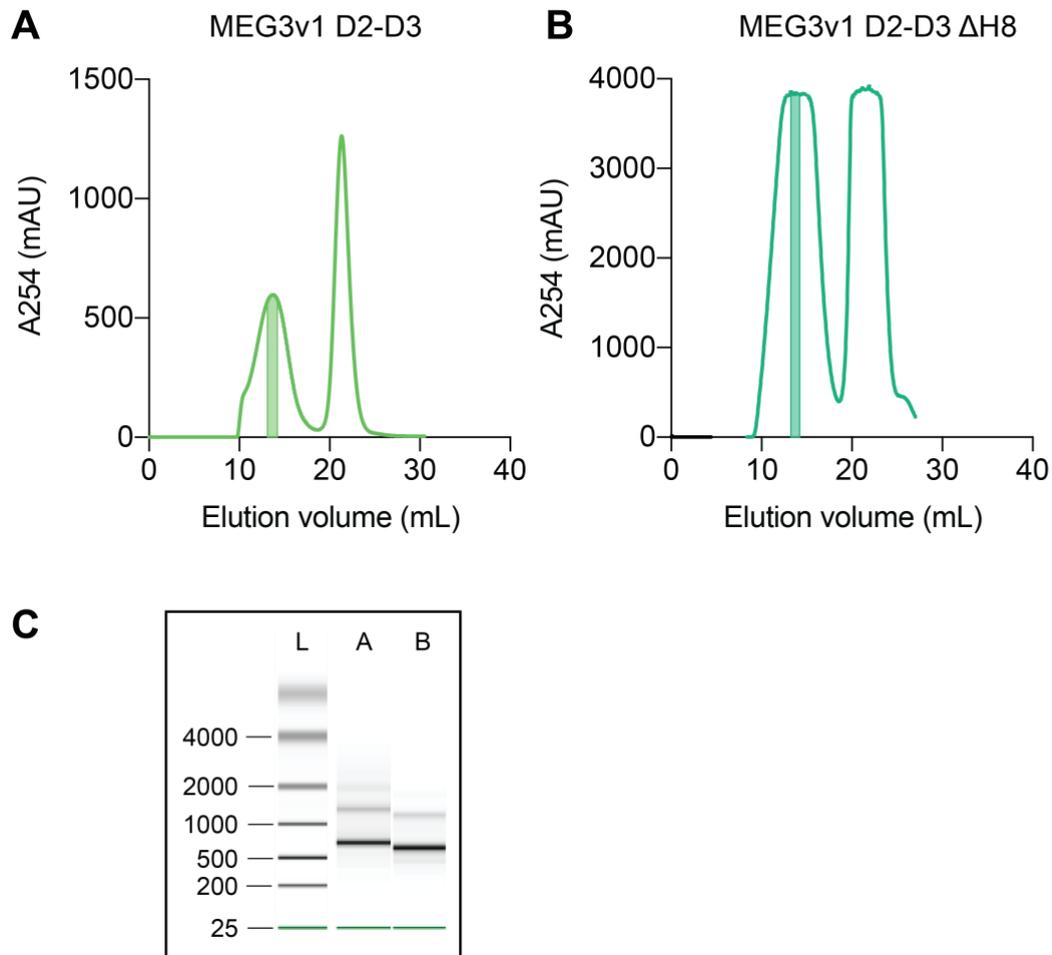


Figure 3:19 Purification of domains D2-D3 and D2-D3 with Δ H8

(A) Chromatogram showing the purification of domains D2-D3 (B) Chromatogram showing the purification of domains D2-D3 with deleted H8 (Δ H8) (C) Bioanalyzer RNA 6000 Nano chip L = Ladder; A = sample domains D2-D3 B = sample domains D2-D3 with Δ H8.

Because there are multiple TRs, each of which can act the pseudoknot receptor, folded *MEG3* likely exists at an equilibrium of tertiary structural conformers whose organisation depends on the choice of the pseudoknot-forming TR. Thus, the redundancy mechanism ensuring mutational resistance of the functional mechanism of *MEG3*, becomes a problem in the context of crystallising *MEG3* as structural heterogeneity prevents crystal lattice formation. To overcome this problem, I mutated five of TRs aiming to weaken their potential interactions with the loop of helix H11 and thus forcing the pseudoknot motif to form with the remaining, intact TR (Figure 3:13). From these experiments it was shown that all *MEG3* conformers regardless of the pseudoknot-forming TR are able to drive p53-dependent gene activation and are therefore functional. However, when the gene activation efficiency of those mutants expressed as a percentage of the activation efficiency of WT *MEG3v1*, the conformer

with TR1 as the pseudoknot receptor appears to have the greatest activation efficiency compared to the other mutants (Figure 3:11). Based on this result, I purified constructs of the *MEG3* active core with a forced preference to form the pseudoknot via TR1 (Figure 3:20). I first increased the concentration of this sample to 5 mg/mL by a centrifugal column concentrator and then I tested the homogeneity of the sample by running a Bioanalyzer RNA 6000 Nano chip (Figure 3:20). The Bioanalyzer chip revealed a highly pure and homogenous RNA sample and therefore I proceeded with setting up a crystallisation trial for this construct using the high throughput crystallisation facility at EMBL, Grenoble. Unfortunately, mutating the TRs was not sufficient to stabilise *MEG3* for crystals to form.

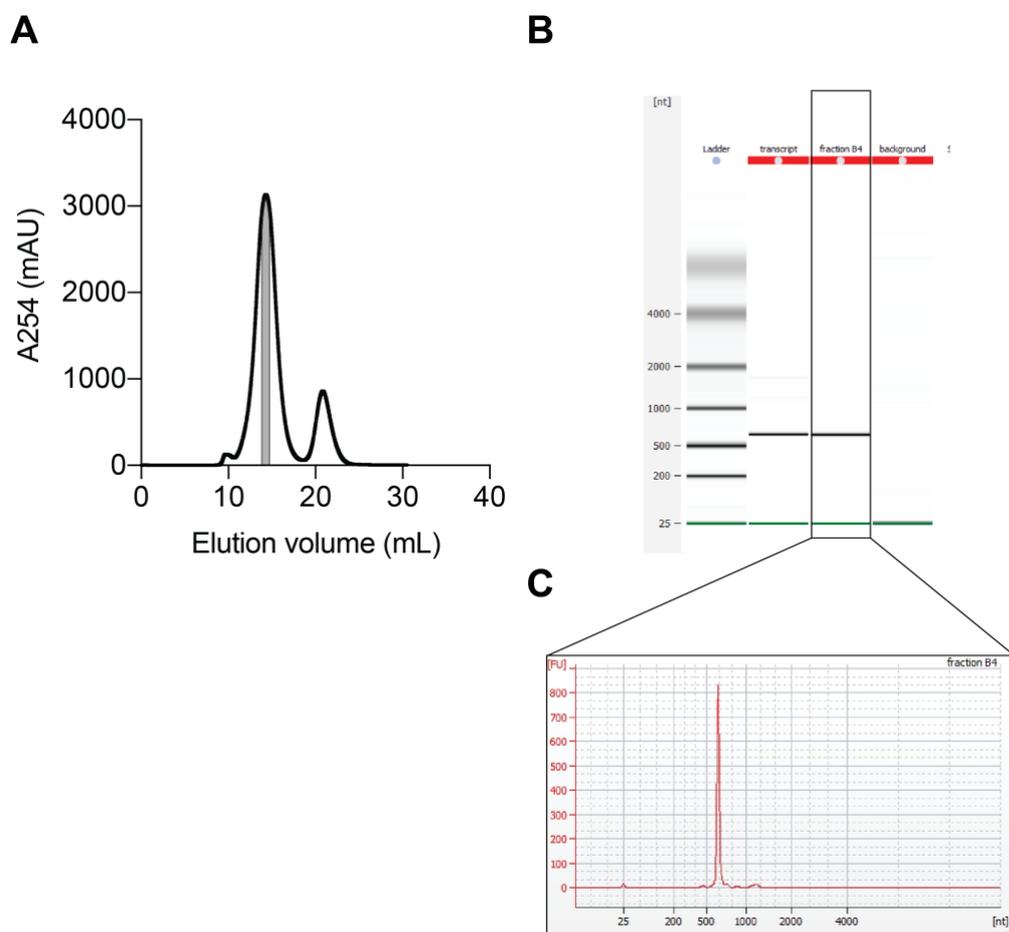


Figure 3:20 Purification of *MEG3* domains D2-D3 Δ H8 with forced pseudoknot formation on TR1 receptor

(A) Characteristic chromatogram from size exclusion chromatography purification of *in vitro* transcribed *MEG3* domains D2-D3 with deleted helix H8 and forced pseudoknot formation via TR1. Shaded area represents the peak fraction (fraction B4 in panel B). (B) Bioanalyzer RNA 6000 Nano chip image showing the *in vitro* transcription product and the purified peak fraction along an RNA size standard. (C) Electrophoretic data corresponding to purified sample from the Bioanalyzer chip.

3.3.5 Chapter discussion

In this chapter I described my efforts to characterise the tertiary structure of *MEG3* in molecular detail. Firstly, *MEG3* was studied by Tb(III) probing which revealed the nucleotide positions involved in coordinating divalent ions in 3D-folded *MEG3* and thus contributing to keeping the structure together. Given our previous evidence supporting that *MEG3* folding is dependent on Mg^{2+} concentration, Terbium (III) probing is a very informative technique as it provides experimental evidence for the detailed molecular mechanism of *MEG3* folding. One major drawback of this approach is its difficulty to validate in the absence of a high-resolution structure. Since Mg^{2+} and Tb^{3+} ions alike are coordinated by the sugar moiety of nucleotides, mutating the identified nucleotides would not necessarily change the positioning of the ions in the folded RNA, unless the mutation affects intramolecular base-pairing and the formation of secondary and tertiary RNA structures in which case the positions of ions across the entire RNA would likely be affected. Moreover, deleting individual nucleotides would not affect ion coordination since the neighbouring nucleotide would replace the deleted one. Thus, our cell-based functional assay could not be used as a means to connect individual ion-coordinating nucleotides underlying 3D structure with the functional mechanism of *MEG3* in gene regulation at this stage. Previously, Tina Uroda deleted entire structured regions such as helices from *MEG3v1* and tested those constructs using our functional reporter assay (Uroda, 2019). From those studies, we can deduce which of the identified ion-binding nucleotides may have functional importance compared to others. For instance, deleting helix H10 which is where U334 is mapped, reduced the potency of *MEG3*-driven p53-dependent gene activation by almost 50% compared to *MEG3v1* WT (Uroda, 2019). Similar results were observed when the entire region including helices H16 to H24 were deleted which contain C530, U531 and G543. In other cases, such as deleting the 3' end part of helix H27 along with helices H28 and H29 and the junction between them (construct *MEG3* Δ H27'-H29 in (Uroda, 2019)), had no effect on the function of *MEG3* on p53-dependent gene regulation. This region contains the ion-binding nucleotides G800, A814 and C855. Hence, even though we cannot use our reporter system to draw definite conclusions about the functional roles of individual ion-binding nucleotides within *MEG3*, by assessing those nucleotides in the context of their surrounding structural elements, we can deduce which ion-

binding sites may have some functional role and prioritise their functional characterisation in the future.

In order to approach characterising the 3D structure of *MEG3* at high-resolution, I experimented with crystallising the active core of *MEG3* for further studies by X-ray crystallography. Unfortunately, *MEG3* proved too flexible for crystal contacts to form and therefore this part of my PhD work was not completed. This was not an unexpected result since flexible regions which could not be assigned to a defined domain were obvious from the secondary structural model. Moreover, studies of *MEG3v1* by small angle X-ray scattering (SAXS) supported that *MEG3* was a partially folded molecule (Uroda *et al*, 2019). Notably, SAXS had not been performed on the domain D2-D3 minimal construct and due to technical limitations *MEG3v1* could be study by SAXS in the presence of monovalent (K^+) but not divalent (Mg^{2+}) ions. Crystallography may not be a suitable technique for the study of *MEG3* and in extrapolation of lncRNAs in general. Performing more mutations and deleting more linker and loop regions may potentially stabilise the molecule further however it would also increase the risk of losing or altering function as would be expected by accumulating mutations. Hence, any further crystallisation construct design should be coupled to functional characterisation ensuring the biological relevance of the constructs following the methodology described in this chapter.

Traditionally, flexible samples such as RNA have been studied by nuclear magnetic resonance spectroscopy (NMR). Briefly, in NMR spectroscopy, a magnetic field is applied to the sample causing the nuclei of certain atoms to resonate at characteristic, measurable frequencies. The resonance frequency coming from a given atomic nucleus depends on the identity and spatial orientation of neighbouring atoms. By advanced biophysical calculations the solution structure of biomolecules can be deduced from NMR spectra. The most common nuclear resonances come from 1H and ^{13}C . This makes RNA more difficult to study by NMR than proteins because the proton density of RNA is looser than this of proteins (Barnwal *et al*, 2017). Moreover, because RNA has only 4 different building blocks (A, U, G and C nucleosides) the NMR spectra from RNA tend to overlap more than those of proteins (Barnwal *et al*, 2017). There are other technical difficulties involved with RNA structure determination by NMR resulting in limiting the method's capacity to the study of RNAs of up to ~ 100-200 nucleotides in length (Barnwal *et al*, 2017). This

methodological limitation excludes NMR from the list of potential options for the study of *MEG3* since a biologically relevant, minimal structure of *MEG3* would include its active core, comprising domains D2 and D3.

Another attempt towards a high-resolution structure of the *MEG3* active core could be to use my construct comprising domains D2-D3 with deleted helix H8 and the H11-TR pseudoknot forced to form via TR1, which showed structural conformer homogeneity (Figure 3:20) for studies with cryo-EM. I would suggest an additional step in the sample preparation procedure to include cross-linking for instance by the GraFix method ([Stark, 2010](#)). Cross-linking could potentially decrease sample flexibility enough to allow for 2D class averaging and subsequent 3D modelling of the sample.

In conclusion, solving the tertiary structure of *MEG3* in high resolution is near impossible given the current state of the technology. My work has provided the first steps towards producing a stable and biologically relevant construct of *MEG3* by targeted mutations aiming to limit its flexibility while maintaining function.

3.4 Investigation of the evolutionary conservation of the *MEG3* functional mechanism

3.4.1 Résumé en français

La découverte récente d'ARN non-codants fonctionnels a suscité de nombreuses questions. Par exemple : les fonctions et le mécanisme des ARN non-codants sont-ils conservés entre les espèces? Jusqu'où pouvons-nous remonter dans l'évolution pour retrouver les ARN non-codants individuels? Pour la classe spécifique des lncRNAs, la réponse à cette question n'est pas simple. Les lncRNAs sont une classe d'ARN non-codant tellement diversifiée que les membres individuels de la classe des lncRNAs sont susceptibles d'avoir des chemins évolutifs très différents les uns des autres. Dans ce chapitre, j'explore la conservation évolutive du rôle de *MEG3* dans la régulation génique dépendante de p53. Je décris d'abord l'utilisation de méthodes computationnelles pour l'identification d'éléments lncRNA potentiellement conservés. Ensuite, je décris mon travail sur la caractérisation structurelle et fonctionnelle de novo de *Meg3* murine.

3.4.2 Chapter introduction

The recent discovery of functional non-coding RNAs (ncRNA) has sparked many questions. For example: are the functions and mechanism of ncRNAs conserved across species? How far back in evolution can we trace individual ncRNAs? For the specific class of lncRNAs the answer to this question is not simple. LncRNAs are such a diverse class of ncRNA that individual members of the lncRNA class are likely to have a different evolutionary paths from one another. In this chapter I am exploring the evolutionary conservation of the role of *MEG3* in p53-dependent gene regulation. I first describe the use of computational methods for the identification of potentially conserved lncRNA elements. Then I describe my work on the *de novo* structural and functional characterisation of murine *Meg3*.

3.4.3 Sequence and structure alignment provide evidence supporting that the *MEG3* functional mechanism is conserved through evolution

The NCBI database contains transcript sequences annotated as *MEG3* belonging to 9 mammalian species. The highest level of conservation in the sequences of those *MEG3* genes was observed in exon E3 which is the exon containing helix H11 and the TRs in human *MEG3*. Using exon E3 as a seed I employed the programme BLAT (Kent, 2002) to search for genomic sequences that could potentially encode *MEG3* in the genomes of all species which have had their genomes sequenced. The BLAT search extended our collection of *MEG3* sequences to 40 mammals. Next I aligned the obtained sequences using CLUSTAL OMEGA (Madeira *et al*, 2019). The alignment showed a reasonable fit for 19 of the 46 available sequences but failed to fit the remaining 27 (Uroda *et al*, 2019). I then performed the alignment using the programme Infernal (Nawrocki *et al*, 2009). Infernal is a programme specifically designed for RNA alignments and can take secondary structural information as input, calculating compatible mutation rates and covariation events. Compatible mutations and co-variations are terms used to refer to mutations occurring on nucleotides that base-pair with each other, for instance on either side of a helical stem. In the example of a helical stem, in the case of compatible mutations only one nucleotide is mutated on one side of the stem however the capability to form a base-pair with its partner at other side of the helix is maintained for example through a non-Watson-Crick pair. In the case of co-variation both nucleotides would mutate in a manner that maintains the structure of the helix whilst the overall sequence is different. When using classical sequence alignment tools such as CLUSTAL OMEGA any mutation would be detected as a change in the sequence and would therefore skew the alignment suggesting that the region in question is not under strict evolutionary pressure. However, when studying structured RNAs, regions with highly compatible mutation and/or co-variation rates are on the contrary considered as highly conserved since the ability of the RNA to form the correct structural elements is more important to its function rather than its exact sequence. Analysis of *MEG3* with Infernal expanded the alignment to 44 species spanning all kingdoms of mammals (Figure 3:21 and Table 3:2) (Uroda *et al*, 2019). Of the 46 total *MEG3* sequences that could be identified, *MEG3* from armadillos, bison, sheep, tarsiers and Tasmanian devils seemed not to contain TRs. This could be the result of poor cDNA read qualities. Since the sequences appeared to be incomplete, we did not include the *MEG3*

sequence from those animals in our covariation analysis. *MEG3* could not be found in any non-mammalian species. The highest degree of conservation was found in the region of helix H11 where the loop region was ultraconserved, i.e. it did not have any mutations across all species, whilst the stem showed occurrences of compatible mutations and co-variations (Figure 3:21).

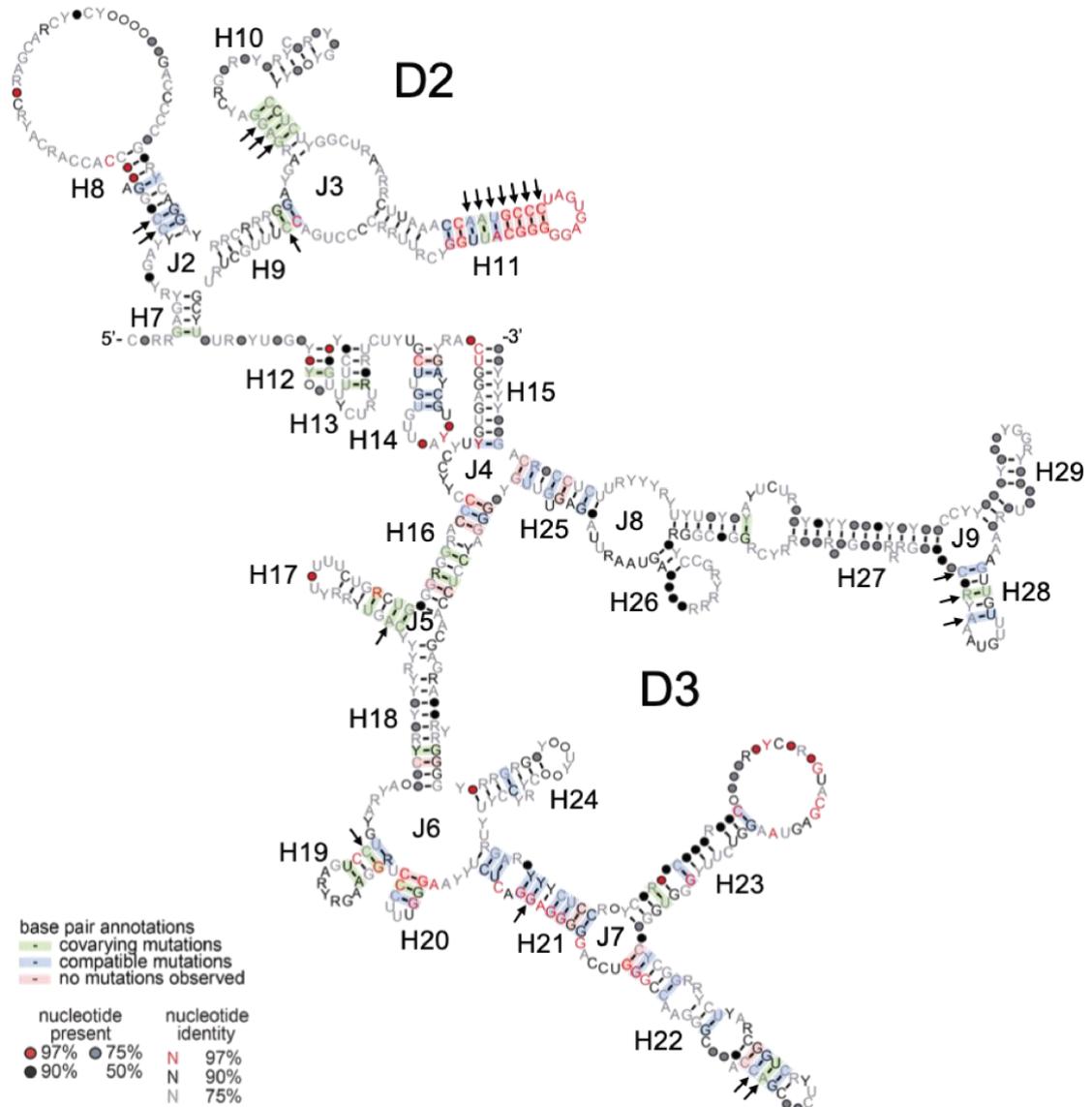


Figure 3:21 Sequence and structural alignment of 41 mammalian *MEG3* analogues mapped on the secondary structure model of human *MEG3v1*

Domains D2-D3 are shown helices (H) and junctions (J) are marked according to their order of appearance in *MEG3v1*. Absolutely conserved residues are in red, covarying and compatible mutation events are highlighted in green and blue respectively. Arrows point to conserved base-pairs. Original figure published in (Uroda *et al*, 2019).

Table 3:2 Tandem repeats identified in different mammals

	Common name	species	TR1	TR2	TR3	TR4	TR5	TR6
1	Human	<i>Homo sapiens</i>	UCUUGCU	CUUGU	UCUUACU	CUUGU	UCUUAAU	UUUAAU
2	Chimpanzee	<i>Pan troglodytes</i>	UCUUGCU	CUUGU	UCUUACU	CUUGU	UCUUAAU	UUUAAU
3	Bonobo	<i>Pan paniscus</i>	UCUUGCU	CUUGU	UCUUACU	CUUGU	UCUUAAU	UUUAAU
4	Gorilla	<i>Gorilla gorilla</i>	UCUUGCU	CUUGU	UCUUACU	CUUGU	UCUUAAU	UUUAAU
5	Orangutan	<i>Pongo abelii</i>	UCUUGCU	CUUGU	UCUUACU		UCUUAAU	UUUAAU
6	Gibbon	<i>Nomascus leucogenys</i>	UCUUGCU	CUUGU	UCUUACU		UCUUAAU	UUUAAU
7	Rhesus macaque	<i>Macaca mulatta</i>	UCUUGCU	CUUGU	UCUUACU	CUCGU	UCUUAAU	UUUAAU
8	Baboon	<i>Papio anubis</i>	UCUUGCU	CUUGU	UCUUACU	CUCGU	UCUUAAU	UUUAAU
9	Golden monkey	<i>Rhinopithecus roxellana</i>	UCUUGCU	CUUGU	UCUUACU		UCUUAAU	UUUAAU
10	Green monkey	<i>Chlorocebus sabeus</i>	UCUUGUU	UUUGU	UCUUACU		UCUUAAU	UUUAAU
11	Proboscis monkey	<i>Nasalis larvatus</i>	UCUUGCU	CUUGU	UCUUACU		UCUUAAU	UUUAAU
12	Marmoset	<i>Callithrix jacchus</i>	UUUUGCUU	CUUGU	UCUUACU	CUCAU		UUUAAU
13	Squirrel monkey	<i>Saimiri boliviensis</i>	UUUGCU		UCUUGUU			
14	Bushbaby	<i>Otolemur garnettii</i>			UCUCACU	CUUGU	UCUUAAU	UUUAAU
15	Malayan lemur	<i>Galeopterus variegatus</i>	UCUUGCU	CUUGU	UCUUACU	CUCAU	UCUUAAU	UUUAAU
16	Mouse lemur	<i>Microcebus murinus</i>			UCUCAC		UCUUAAU	UUUAAU
17	Tarsier	<i>Tarsius syrichta</i>						
18	Mouse	<i>Mus musculus</i>			UUCAU	CUCAU	UCUCA	UCUCUC
19	Rat	<i>Rattus norvegicus</i>	CUCAC		UCUCAU		UCUCGU	UCUAUU
20	Squirrel	<i>Spermophilus tridecemlineatus</i>	CUCAUU	UUUGC	CCUCACU	CUCGU	UCUUAAU	UUUAAU
21	Kangaroo rat	<i>Dipodomys ordii</i>			CUUUAAU		UCUUAAU	UUUAAU
22	Naked mole-rat	<i>Heterocephalus glaber</i>			UCUUAAU		UCUUAAU	UUUAAU
23	Chinese hamster	<i>Cricetulus griseus</i>			UCUCAU			
24	Hedgehog	<i>Erinaceus europaeus</i>	CCUCAC	CCUGU	UCUCACU	CUCAU		
25	Shrew	<i>Sorex araneus</i>						
26	Megabat	<i>Pteropus vampyrus</i>	UCUCAU		UCUCAC		UCUCAUU	UUUGUU
27	Microbat	<i>Myotis lucifugus</i>					UCUUGUU	UUUAAU
28	Cow	<i>Bos taurus</i>	CCUCACU		UCUCAC		UCUCGUU	UUUAAU
29	Bison	<i>Bison bison</i>						
30	Pig	<i>Sus scrofa</i>			UCUCAC		UCUCAUU	UUUAAU
31	Alpaca	<i>Vicugna pacos</i>		CUUGU	UCUUAC		UCUUGU	UCUAUU
32	Sheep	<i>Ovis aries</i>						
33	Minke whale	<i>Balaenoptera acutorostrata</i>	UCUCAC		UCUUAC		UCUUAAU	UUUAAU
34	Dolphin	<i>Tursiops truncatus</i>	UCUCA		UCUUAC		UCUUAAU	UUUAAU
35	Horse	<i>Equus caballus</i>	UCUCACU	CUUGU				CUUGUU
36	White rhinoceros	<i>Ceratotherium simum</i>	UCUCACU	CUUGU				CUUAAU
37	Chinese pangolin	<i>Manis pentadactyla</i>		CUUGU	UCUCAC	CUUGU	UCUUAAU	UUUAAU
38	Dog	<i>Canis lupus familiaris</i>		CCUCGU	UCUCAU		UCUCGUU	UUUAAU
39	Cat	<i>Felix catus</i>			UCUCACU		UCUCGUU	UUUAAU
40	Panda	<i>Ailuropoda melanoleuca</i>					UCUCGUU	UUUAAU
41	Ferret	<i>Mustela putorius furo</i>		UCUCGC	CCUCAU		UCUUGUU	UUUAAU
42	Armadillo	<i>Dasyopus novemcinctus</i>						
43	Elephant	<i>Loxodonta africana</i>			CUCACU	CUUGU	UCUUAAU	UUUAAU
44	Manatee	<i>Trichechus manatus</i>		CUCACUU		CUUGU	UCUUAAU	UUUAAU
45	Tenrec	<i>Echinops telfairi</i>	UCUUGCU		CUCACU	UUUUUU	UCUCGUU	UUCAUU
46	Tasmanian devil	<i>Sarcophilus harrisii</i>						

For the species studied, I could identify at least one consensus sequence compatible with base-pairing with the loop of helix H11 and thus having the potential to act as a pseudoknot receptor like the TRs in human *MEG3*. Notably, the exact sequence and total number of the potential TRs varied across species (Table 3:2). The bovine *MEG3* homologue present an interesting case study. The bovine *MEG3* transcript sequence deposited in the NCBI database can be mapped on the bovine genome in

locations 65,727,804 to 65,746,033 according to the ARS.UCD1.2 assembly bovine genome, (April 2018). I could not align this transcript with human *MEG3* by sequence or structural alignments (Figure 3:22). I therefore used Infernal to search the entire bovine genome for potential genomic sequences in the bovine *MEG3* locus that could fit the criteria for the TRs. Through this approach, I found that the transcript sequence deposited in the NCBI database contained only the 3' part of exon E3 (from which domains D2 and D3 are formed) (Figure 3:22). However, the sequence matching the entire exon E3 was present in the bovine genome mapping just upstream of the NCBI-deposited transcript from 65,724,253 to 65,746,033 (Figure 3:22). By analysing this sequence, perfect fits for helix H11 and 4 TRs were identified (Table 3:2). This suggests that perhaps *MEG3* homologues are present in the genomes of other species and are simply waiting to be correctly annotated. In order to verify whether the genomic sequence of bovine *MEG3* is actively translated in real animals, I established a collaboration with Maša Roller, a postdoctoral fellow in Paul Flicek's lab at the European Bioinformatics Institute. By mining Chromatin Immunoprecipitation (ChIP) data targeting active promoters, Maša found a promoter just upstream of the new 5' exon boundary which I proposed based on the structural alignments (Figure 3:22). Furthermore, some alternative cDNA reads could be mapped onto this genomic region providing evidence for inclusion of the candidate exon E3 sequence in bovine *MEG3*. The information gathered so far from analysing the sequences of mammalian *MEG3* analogues is compatible with pseudoknot formation between the loop residues of helix H11 and the TRs. Taken together, sequence and structural alignments as well as bioinformatic analysis of active transcriptional centres supported that there is a potential for conservation of the regulatory mechanism of *MEG3* depending on formation of the helix H11-TR pseudoknot across mammals.

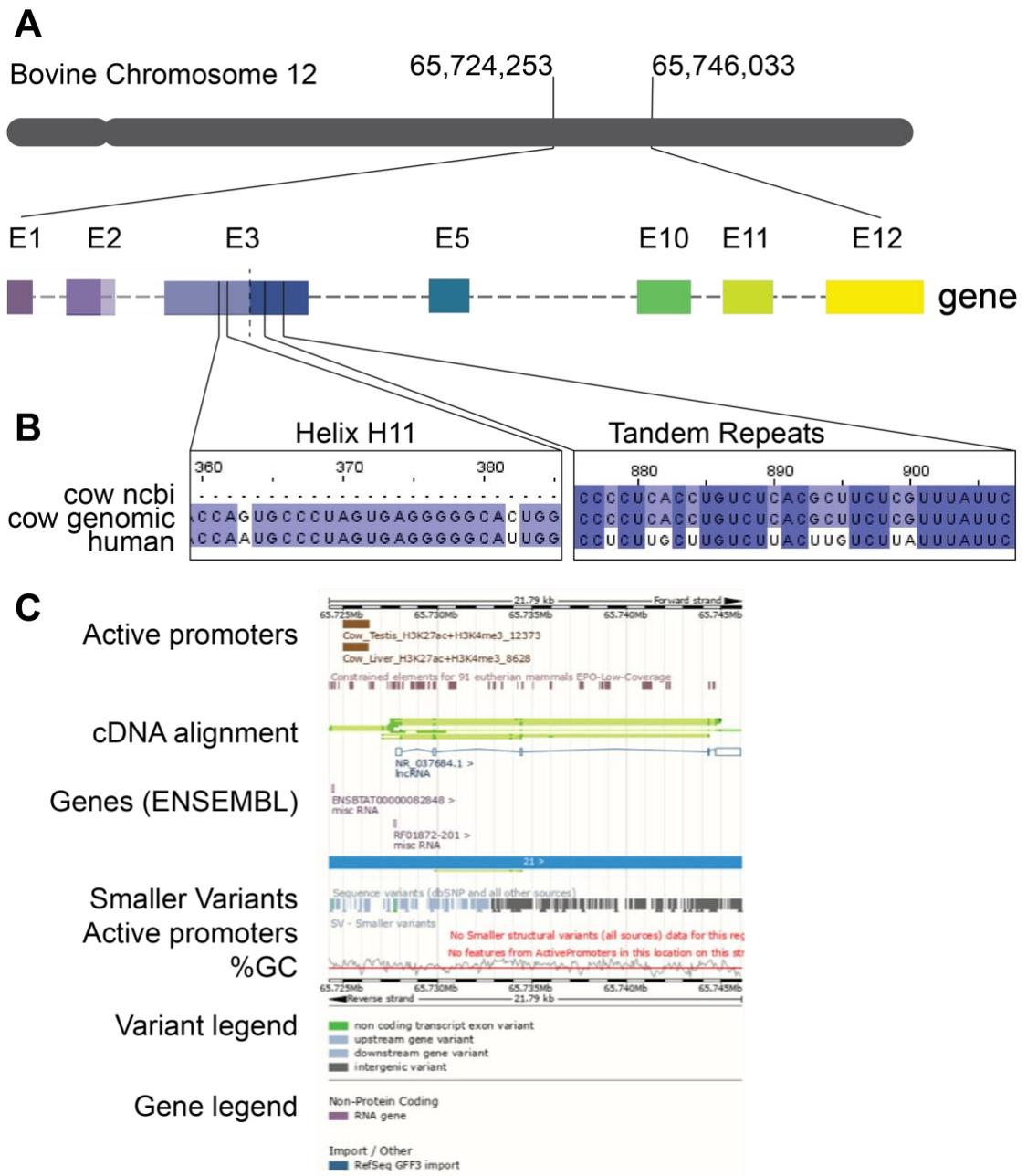


Figure 3:22 Genomic location of bovine *MEG3* and analysis of its sequence

(A) Schematic representation of the gene organisation of bovine *MEG3*. Coloured boxes represent different exons as they map to human *MEG3*. The parts of bovine *MEG3* that are not present in the NCBI-deposited transcript appear faded, a dotted line shows separated portion of exon E3 that is present in the NCBI-deposited transcript from the one that is not. (B) Alignment at the regions of helix H11 and the tandem repeats of the NCBI-deposited bovine *MEG3* sequence and of the putative genomic sequence discovered by Infernal onto human *MEG3*. (C) Screen shot from genome browser page summarising data mining experiment for the identification of active promoters and cDNA reads corresponding to the sequence of exon E3 as identified by infernal (this image was kindly provided by Maša Roller full results can be viewed at http://www.ensembl.org/Bos_taurus/Location/View?r=21:65724257-65746044).

3.4.4 Structural characterisation of mouse *Meg3* provides unbiased proof of structural conservation of *MEG3* lncRNA in mammals

Having revealed evidence for the potential of conservation of the functional mechanism of *MEG3*-driven p53-dependent gene regulation across species, I next wanted to experimentally test this hypothesis. Thus, I chose to structurally characterise murine *Meg3* (hereafter referred to as *mMeg3*), which is the murine counterpart of human *MEG3*. Through structural mapping of *mMeg3* and its functional characterisation via our cell-based luciferase reporter system I aim to provide an unbiased assessment of the extent of evolutionary conservation in the biological mechanism of *MEG3* between humans and mice.

The ENSEMBLE database lists 11 splicing isoforms of *mMeg3* (Figure 3:23). Unlike the known human *MEG3* isoforms (Figure 3:1), some *mMeg3* splicing variants appear to contain intronic sequences (Figure 3:23). Only one *mMeg3* variant, namely *mMeg3-207* seems to follow the consensus splicing pattern that guides human *MEG3* splicing where exons E1-E2-E3 and E10-E11-E12 flank the variable middle exons (see 3.1.3). Variant *mMeg3-207* is equivalent to murine *Meg3* variant 3 (*mMeg3v3*) found on the NCBI transcript database. *mMeg3-207* is 1,896 nucleotides long, a length that is comparable to human *MEG3* splicing isoforms that I have been studying thus far. Finally, a murine transcript compatible with the splicing pattern of *mMeg3-207* has been found to be important for the correct neuronal tissue development of mouse embryos, suggesting that this isoform is functional (Schuster-Gossler *et al*, 1998). Therefore, based on (1) the compatibility of its splicing pattern and absence of intronic sequences, (2) the comparability of lncRNA lengths and (3) the evidence for functional relevance linked to *mMeg3-207*, I chose this *mMeg3* isoform for structural characterisation.

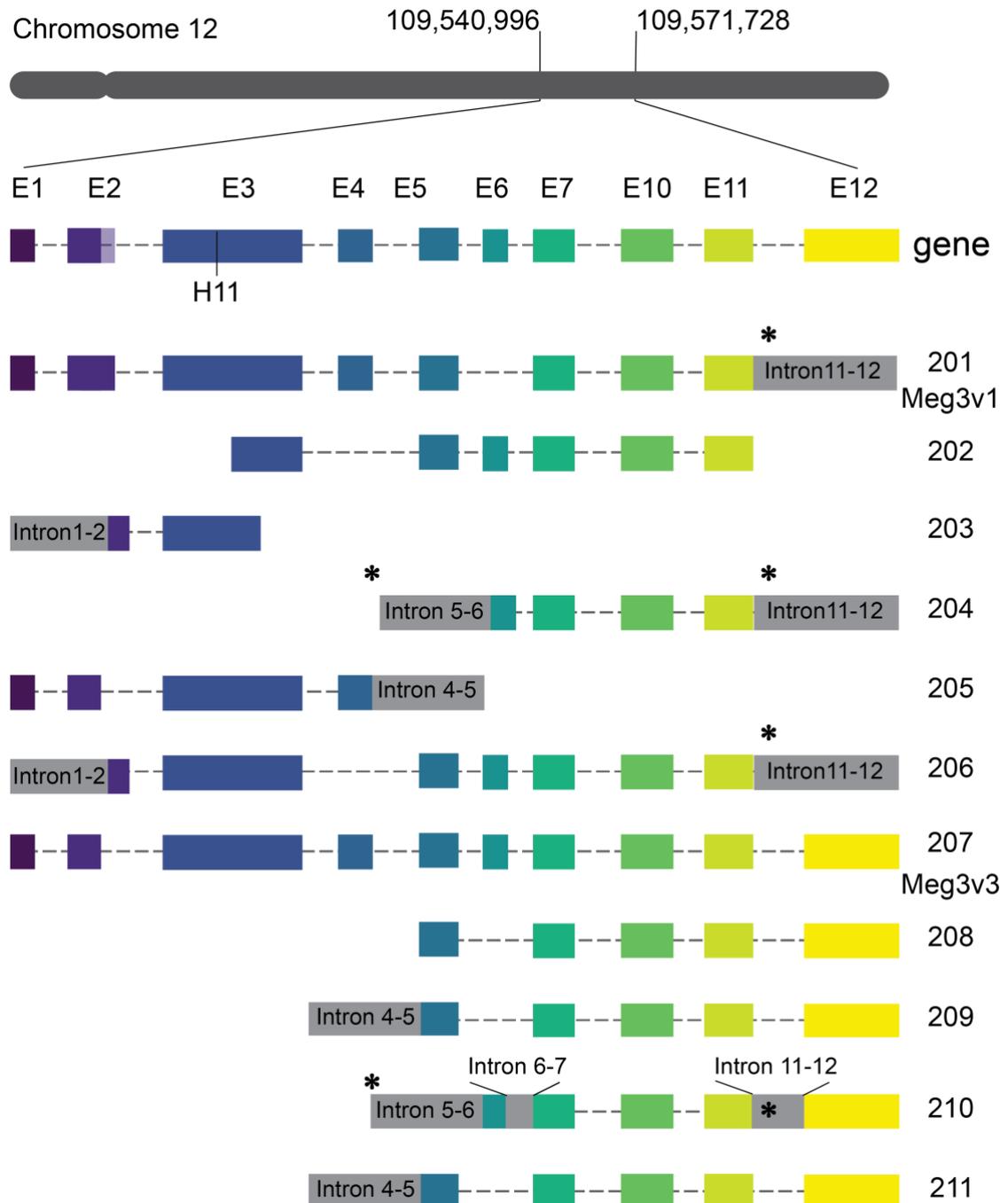


Figure 3:23 Exon organisation and splicing isoforms of mouse *Meg3*

Graphic representation of the mouse *Meg3* gene exon structure (top) and the 11 splicing isoforms of *mMeg3* found in murine cells. Isoforms are numbered according to their ENSEMBL codes where possible the corresponding NCBI names are noted under the ENSEMBL codes. Exons are represented by coloured boxes and are labelled E1-E9, the opaque box next to exon E2 represents an alternative 3' splice site. The colours of the exons are chosen to match the homologous exon sequence of the human *MEG3*. Retained introns are represented by grey boxes. Asterisks mark the mapped NOVA protein binding sites identified by iCLIP in (Ule *et al*, 2003).

First, I established an *in vitro* transcription and purification protocol for *mMeg3v3* using the approach described in (Chillón et al, 2015) (Figure 3:24). To test the quality of the purified RNA, I performed agarose gel electrophoresis with increasing $[\text{MgCl}_2]$ concentrations. It was shown that *mMeg3v3* had the capacity to gain compaction in a $[\text{Mg}^{2+}]$ concentration-dependent manner.

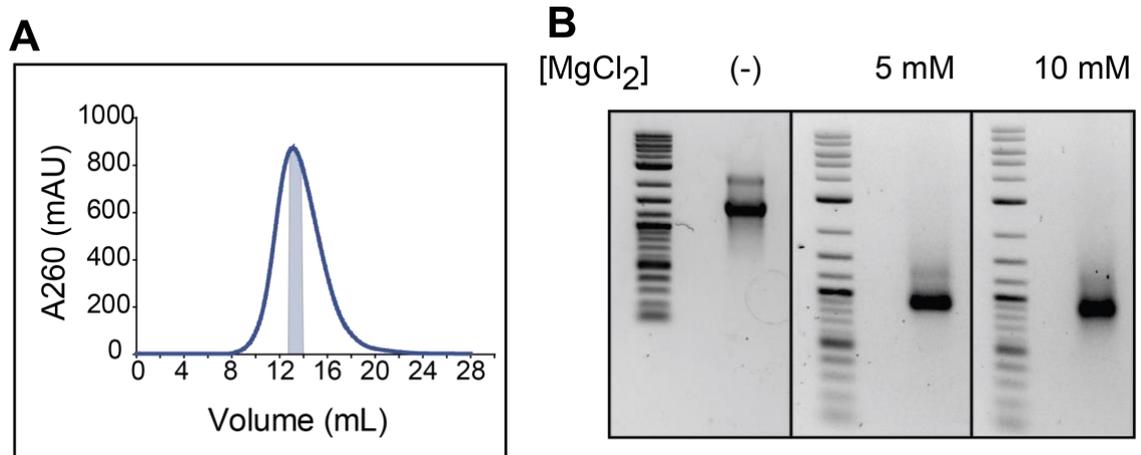


Figure 3:24 Purification and initial biophysical characterisation of *mMeg3v3*

(A) Size exclusion chromatography of *mMeg3v3* shaded area represents the peak fraction.
 (B) Agarose gel electrophoresis with increasing MgCl_2 concentrations.

Next, I performed sedimentation-velocity analytical ultra-centrifugation (AUC) of *mMeg3v3* with titration of $[\text{MgCl}_2]$ in collaboration with Isabel Chillón (Figure 3:25). The hydrodynamic radius (R_h) of *mMeg3v3* at saturating $[\text{Mg}^{2+}]$ was $R_h = 10.4 \text{ nm} \pm 0.0$, a value close to the one of human *MEG3v1* ($R_h = 8.8 \text{ nm} \pm 0.1$) (Uroda et al, 2019). Altogether, these data suggested that mouse *Meg3v3* can be produced as a homogenous RNA, and folded at physiological Mg^{2+} concentration similarly to human *MEG3* and to other lncRNAs such as *HOTAIR* (Somarowthu et al, 2015) and *RepA* (Liu et al, 2017).

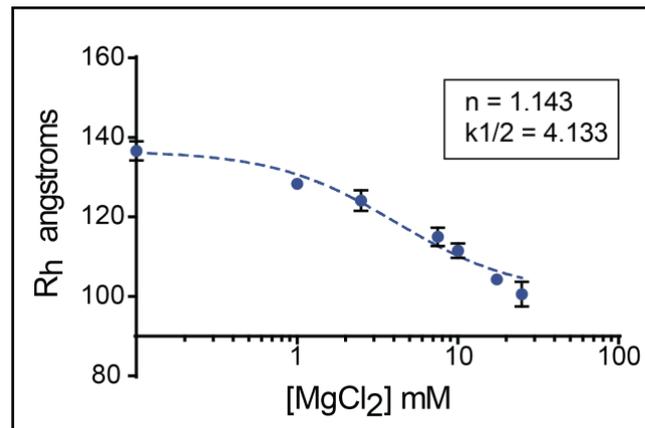


Figure 3:25 Sedimentation-velocity analytical ultracentrifugation of *mMeg3*

The plot shows the distribution of hydrodynamic radius (Rh) of *mMeg3* with MgCl₂ titration fitted into a Hill plot. The coefficient n is the Hill coefficient and it indicated folding cooperativity. The $k_{1/2}$ value is the Mg²⁺ ion concentration in mM at which about 50% of the RNA is folded. This experiment was conducted in collaboration with Isabel Chillón.

Having established that *mMeg3v3* folded in a compact structure, I probed the secondary structure of *mMeg3* by *in vitro* SHAPE with 1M7 (Wilkinson *et al*, 2006). *In vitro* SHAPE probing was performed in triplicates and produced highly consistent results. (Figure 3:26). In addition, the resulting secondary structure model was thermodynamically favourable with overall low Shannon entropy values across the length of the RNA and pairing energetically favourable predicted base-pairs (Figure 3:26). These results are in line with well-defined structural elements across the RNA as opposed to transient structures (Siegfried *et al*, 2014). The secondary structure model revealed that mouse *Meg3* formed 7 modular domains (Figure 3:27). The model included 16 multiway junctions, 36 terminal loops, 55 internal loops and 51 helices (Figure 3:27). Of the 577 base-pairs in the secondary structure map, 8.5% comprised non-Watson Crick pairs. Furthermore, the secondary structure map revealed a GNRA tetraloop in domain D7 (nucleotides G1831-A1834) suggesting potential tertiary interactions with GNRA receptors. All together, these features indicated that mouse *Meg3* adopted higher order structure.

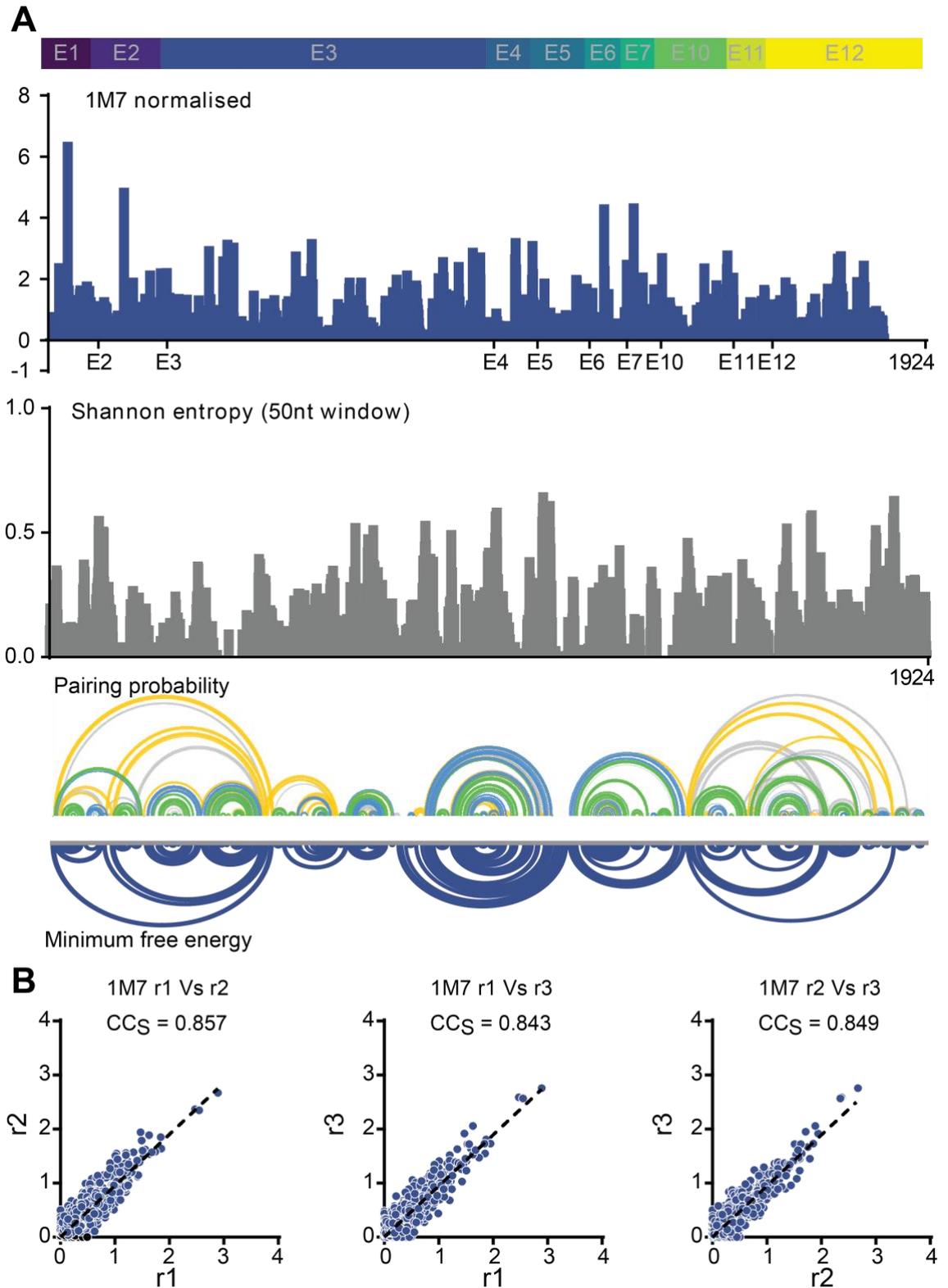


Figure 3:26 Summary of *mMeg3v3* *in vitro* SHAPE probing

(A) Top to bottom: graphical representation of the exon composition of mouse *Meg3v3*; Individual nucleotide 1M7 reactivities across *mMeg3v3* exon boundaries are noted at the x-axis; calculated Shannon entropy values for individual nucleotides across *mMeg3v3* averaged by a 50 nucleotide sliding window; arch plots showing pairing probability and minimum free energies of predicted base-pairs; the pairing probability chart is colour-coded

so that green is the most highly probable pairs and grey is the least probable pairs (B) Correlation of 1M7 individual nucleotide reactivities was assessed for $n = 3$ independent experiments by the Spearman correlation test. The correlation co-efficient for each comparison is noted on top of each graph.

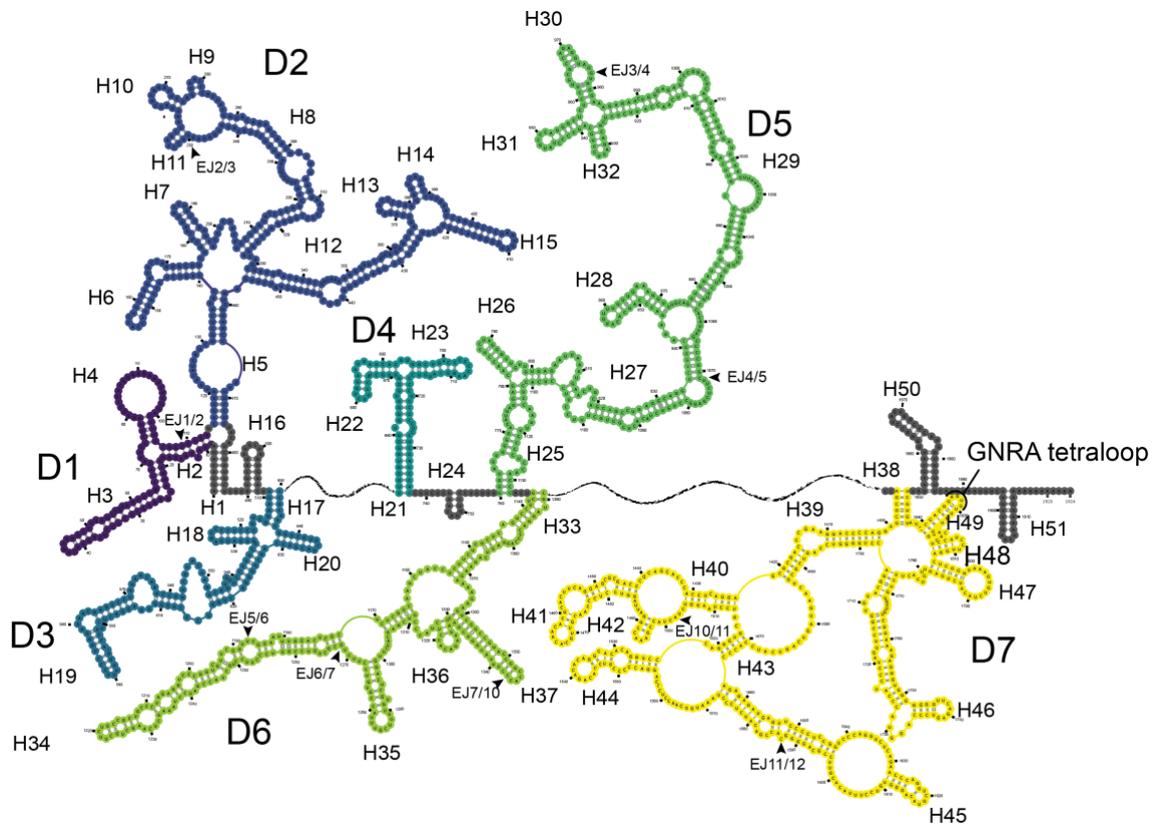


Figure 3:27 Secondary structural map of *mMeg3v3* highlighting its domain organisation

Secondary structure map of mouse *Meg3v3* coloured to highlight domain organisation. Domain colours are in a continuous gradient from dark purple (domain D1) to yellow (domain D7). Nucleotides in grey do not belong to any domain. The potential GNRA tetraloop in domain D7 is marked.

Next, I investigated the individual nucleotide SHAPE reactivities across the entire length of mouse *Meg3* and compared them to those of its human counterpart (*MEG3v1*). This analysis revealed several striking similarities in domains representing regions with the highest evolutionary conservation (Figure 3:28). Specifically, the SHAPE reactivities and in extend the secondary structures of helix H11 in human *MEG3v1* and helix H15 in mouse *Meg3v3* respectively were absolutely conserved (Figure 3:28). Structural conservation persisted despite some sequence variation (the stem residues of helix H11 contain the co-variant pair G399-C418 instead of the human A360-U379 (Figure 3:28). Finally, low SHAPE reactivities were observed on loop nucleotides 408-411 of mouse *Meg3v3*, which correspond to loop nucleotides 368-373 of human *MEG3v1*. This was in line with the functional roles

assigned to helix H11 and the formation of pseudoknot interactions. Finally, the helical structure of H11 (human)/H15 (mouse) was conserved even though the surrounding secondary structural elements differed (Figure 3:28). Taken together, the experimentally determined secondary structure of *mMeg3* was compatible with a functional and mechanistic evolutionary conservation model for *MEG3* lncRNA from mice to humans.

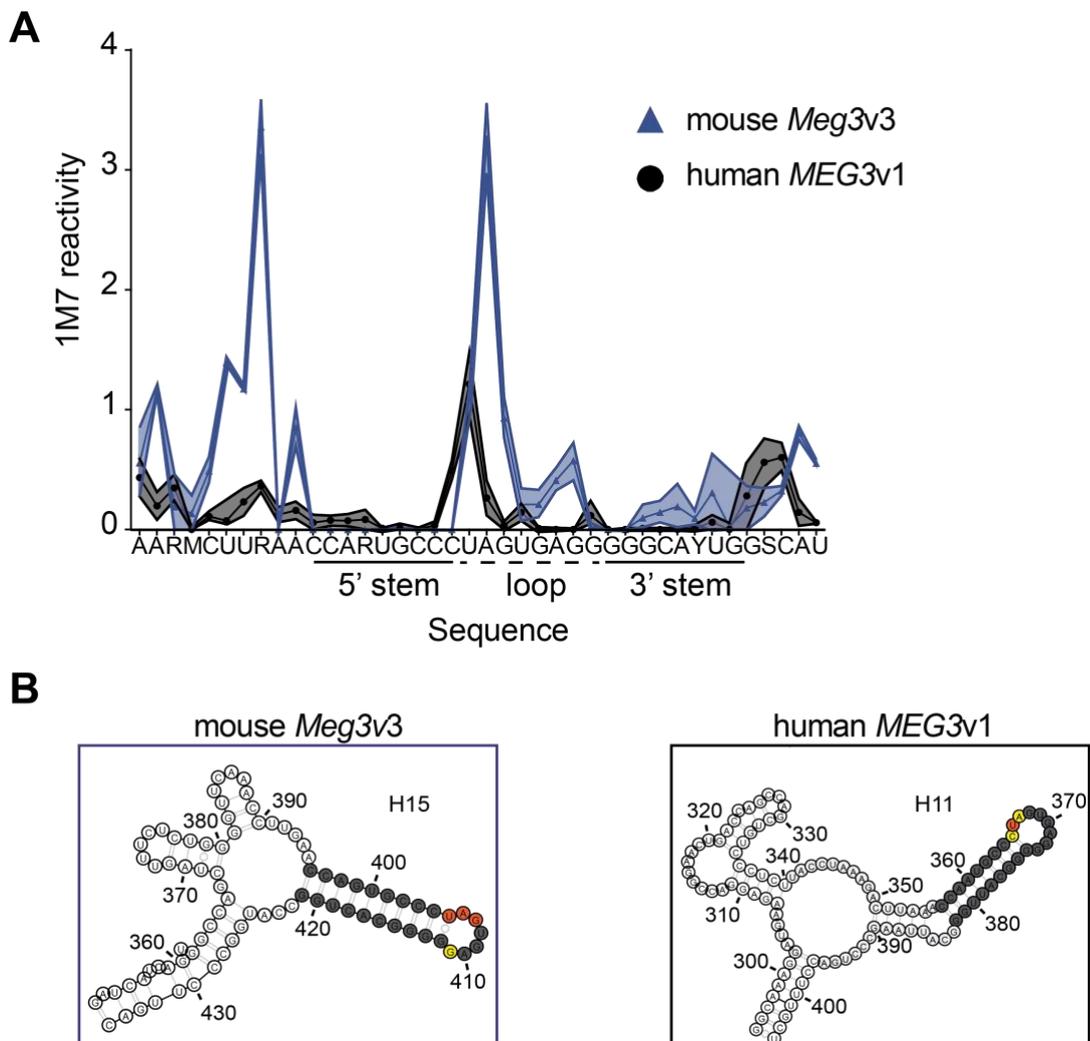


Figure 3:28 1M7 reactivities of individual nucleotides in the helix H11/H15 region
 (A) The 1M7 SHAPE reactivities of individual nucleotides across the aligned sequences of mouse *Meg3* variant v3 (dark blue) and human *MEG3* variant v1 (black) are shown. The solid line in the middle of each trace represents the mean reactivity and the opaque frame is the standard error of the mean from $n = 3$ replicates. Residues belonging to the stem of helix H15/H11 respectively are marked by a solid line under the sequence and the loop is marked by a dashed line. (B) Schematic representation of the secondary structural maps of mouse *Meg3* (left - dark blue frame) and human *MEG3v1* (right - black frame) showing the region of helix H15/H11 respectively colour-coded according to individual nucleotide SHAPE reactivity. Colour code is as in Figure 3:7; Grey is SHAPE low reactivity, yellow is medium and red is high reactivity suggesting high, medium and low probability for base-pairing respectively. The SHAPE data for human *MEG3v1* shown in this figure were produced by Tina Uroda.

3.4.5 Functional characterisation of murine *Meg3* in p53-dependent gene activation

Having established that mouse *Meg3* folded into a well-structured globular-like shape in a $[Mg^{2+}]$ -dependent manner and that its secondary structural features were compatible with the formation of tertiary interactions such as the H11-TR pseudoknot of human *MEG3*, I moved on to characterise the function of *mMeg3* in the regulation of p53-dependent gene expression. Human p53 and murine TP53 are very well conserved (see Appendix 5.2). I therefore tested whether mouse *Meg3* could activate human p53 in our functional reporter assay in human HCT116 cells. This experiment showed that *mMeg3* did not activate the human p53 pathway from a variety of different p53-response elements (Figure 3:29). This result was independent of the species origin on the p53 protein since when I repeated the experiment in HCT116 p53^{-/-} cells, transiently expressing either human p53 or mouse TP53 protein respectively a similar pattern was observed (Figure 3:30). Thus, whereas human *MEG3* was able to drive p53-dependent gene activation via either human p53 or murine p53 in the human HCT116 cell line, mouse *Meg3* did not promote p53-dependent gene activation inside human cells regardless of the p53 protein analogue.

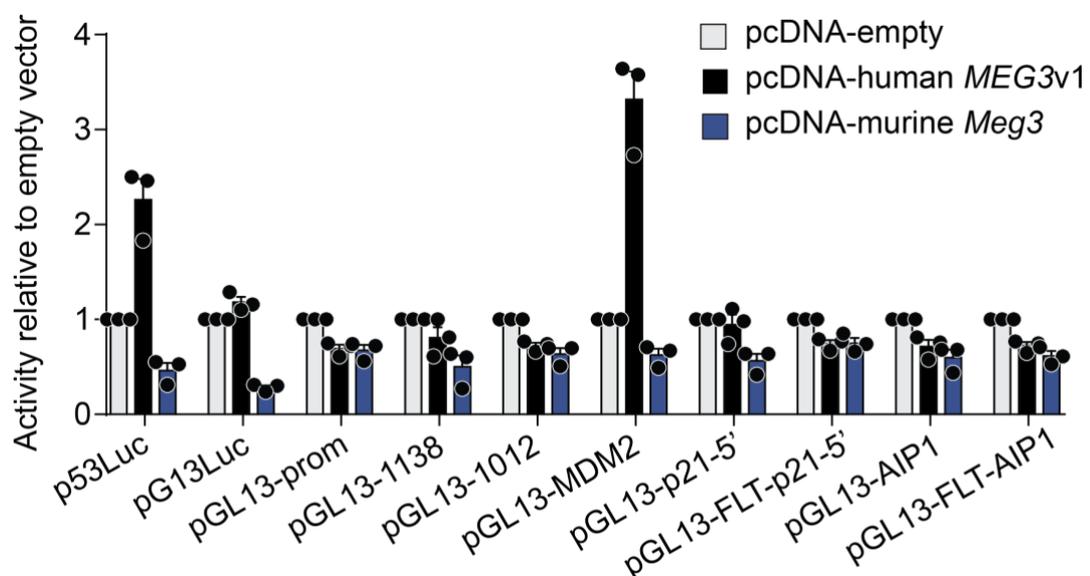


Figure 3:29 Activity assay for human *MEG3v1* and mouse *Meg3* in human HCT116 cells

Error bars show the standard error of the mean from n = 3 biological replicates.

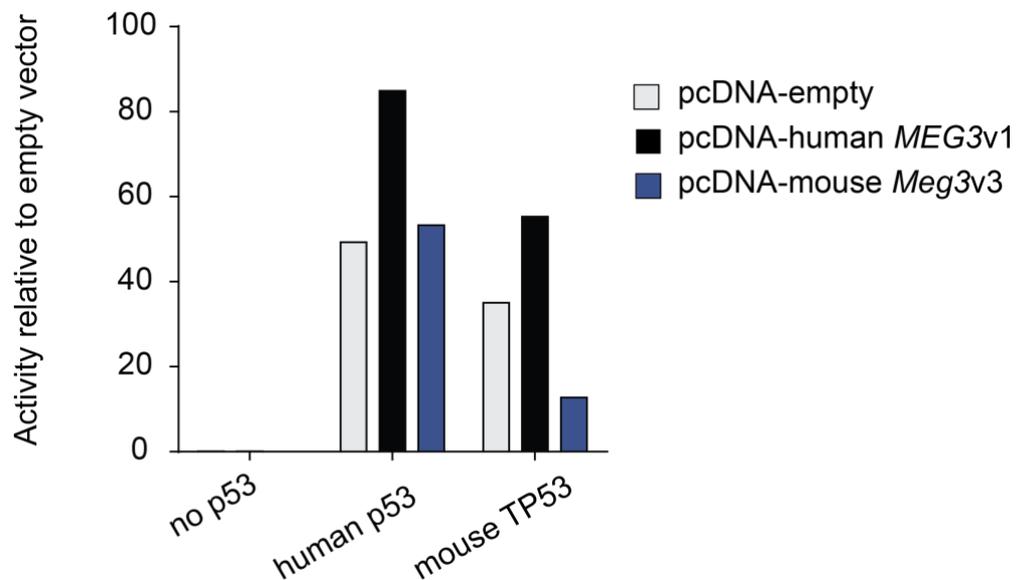


Figure 3:30 Activity assay using HCT116 p53^{-/-} cells

HCT116 p53^{-/-} cells which are genetically identical to normal HCT116 except that their gene coding for p53 has been knocked out were used to assess the ability of human *MEG3* (black) and mouse *Meg3* (blue) to activate p53-dependent gene expression of the reporter plasmid when transfected with a plasmid encoding either human or murine p53. This experiment was only performed once.

Hence, I moved to testing the function of mouse *Meg3* in a murine cell line. Since we did not have a murine cell line with validated lack of *mMeg3* expression, I first characterised Hepa1-6 cells by RT-qPCR to quantify the levels of *mMeg3* RNA present in the cells (Figure 3:31). No significant *mMeg3* levels could be detected in 3 biological replicates. Hence, Hepa1-6 cells were chosen for further functional assays of *mMeg3*. The functional assay showed that *mMeg3* was able to drive p53-dependent gene activation in the mouse cell line Hepa1-6 to a level slightly less efficient than human *MEG3v1* (Figure 3:32). Curiously, deleting helix H15 (H11 in human *MEG3*) seemed to slightly increase the p53-dependent gene activatory effect of *mMeg3* (Figure 3:32). Those results suggested that murine *Meg3* could activate p53-dependent gene expression in murine cells but that the mechanism of activation depended on a motif other than helix H15.

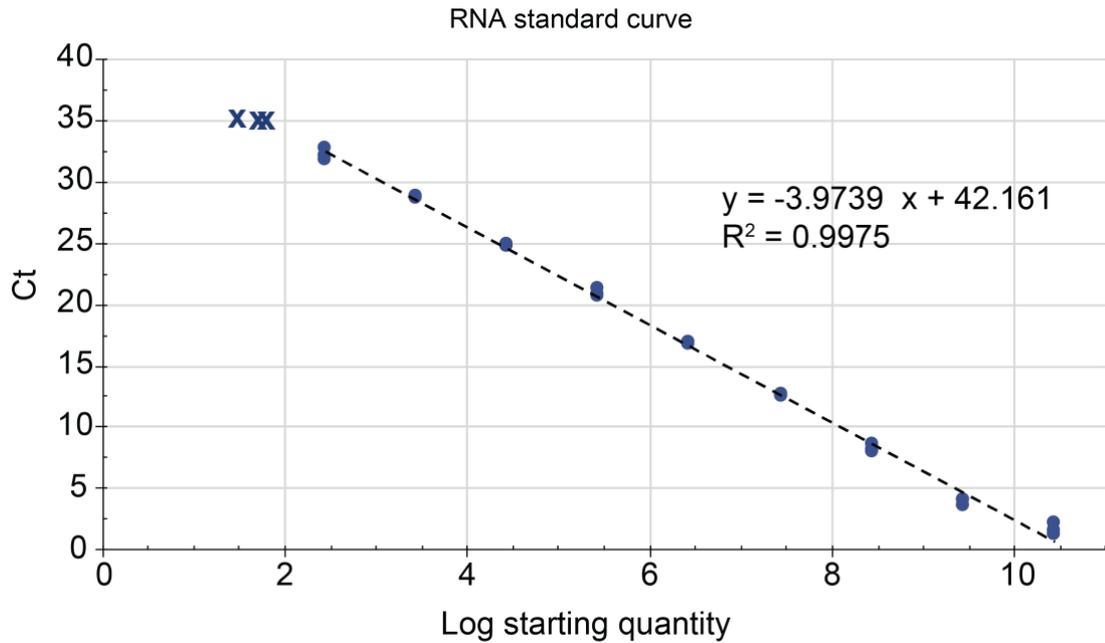


Figure 3:31 RT-qPCR quantification of endogenously expressed *mMeg3* in the murine cell line Hepa1-6

Endogenous RNA levels were quantified by RT-qPCR against an RNA standard curve created by plasmid dilutions. Individual dots Log starting quantity values show the results from $n = 3$ independent experiments. The quantified *mMeg3* RNA from the Hepa1-6 cell line is marked by X. This experiment was conducted in collaboration with Isabel Chillón.

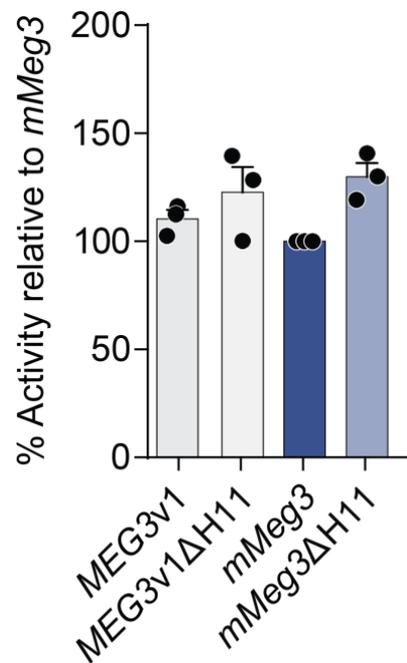


Figure 3:32 Activity assay of human *MEG3v1* and mouse *Meg3v3* on mouse Hepa1-6 cells

Luciferase reporter assay testing the activation efficiency of human *MEG3v1* (grey), mouse *Meg3* (blue) and their helix H11/H15 deletion mutants (light coloured bars) on p53-dependent gene expression. Error bars show the mean from $n = 3$ independent experiments.

To further investigate the active motifs of murine *Meg3*, I designed individual exon deletion constructs and tested their effect on p53-dependent gene regulation using our reporter system in Hepa1-6 cells (Figure 3:33). I found that deleting different individual exons had a distinct effect on the ability of *mMeg3* to regulate p53-dependent gene expression. Interestingly, deleting exon E3 did not significantly affect p53-dependent reporter gene activation. From this experiment it appears that the essential motifs for *mMeg3* activity within the p53 pathway reside in exons E1, E6, E11 and E12 since deleting those exons seemed to affect reporter gene activation the most. Those results support the idea of mouse *Meg3* being a modular lncRNA composed of domains acting in concert to regulate the function of the entire RNA. Taken together, the findings of the functional assays on *mMeg3* were compatible with a mechanism where the arrangement of surrounding domains around the active core of *mMeg3* regulate the exact functional effect of the lncRNA similar to what we observed for the human *MEG3* splicing isoforms.

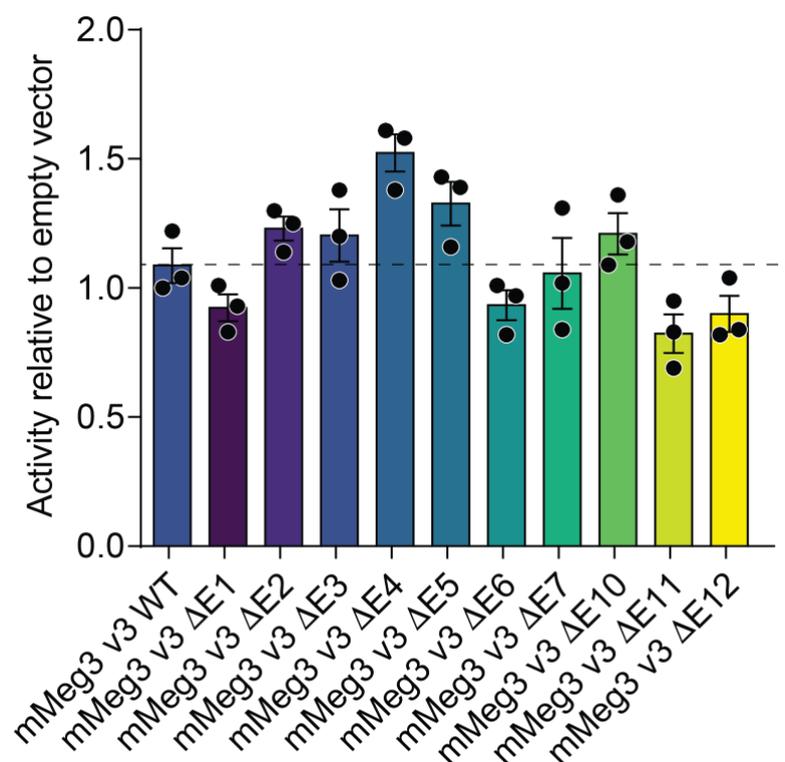


Figure 3:33 Activity assay of individual exon deletion constructs of murine *Meg3v3*
Luciferase reporter assay testing the activation efficiency on p53-dependent gene expression of mouse *Meg3* (blue) and constructs with deletions of individual exons (Δ Ex – coloured according to Figure 3:23). Error bars show the mean from $n = 3$ independent experiments.

3.4.6 Chapter discussion

In this chapter, I characterised the lncRNA *MEG3* from an evolutionary perspective. *MEG3* was shown to be conserved across mammals in terms of its sequence. Through covariation and compatible mutation analysis it was shown that the ability to form the structure of helix H11 was maintained across mammals, suggesting that the mechanism of *MEG3* had the potential of also being conserved.

LncRNA genes are not as well-studied as protein-coding genes. Through my analysis I discovered a potential new splicing variant of bovine *MEG3* or a possible incomplete transcript deposited in the NCBI database. This highlighted the strength of structure-based alignments for assessing lncRNA conservation. Perhaps the publicly available list of *MEG3* homologues could be extended by this targeted genome scanning approach. The data mining approach employed by our collaborators, provided initial evidence for the transcription of those genomic sequences as well as function at the RNA level. However, we still lack experimental data proving the presence of bovine *MEG3* transcripts in animal tissues. Given the tissue-specific expression pattern and low abundance of *MEG3*, it will be difficult to obtain such experimental data. The best approach to this problem would be to extract total RNA from different bovine tissues including neuronal and endocrine tissues since these show the highest levels of *MEG3* transcription in humans. Then 3'-end and 5'-end rapid amplification of cDNA ends (RACE) with primers for bovine *MEG3* should be used to verify transcription ([Yeku & Frohman, 2011](#)). One major drawback of this method is that it is low throughput and relies on the availability of tissue samples from healthy animals. Additionally, since *MEG3* is most abundant in neuronal and endocrine tissues, tissue sample acquisition would inevitably require culling animals which raises ethical considerations. As technology advances and RNA sequencing data for more cell types and organisms start to become available, the presence of *MEG3* transcripts and splicing isoforms might be revealed in more species. Until experimental evidence is provided validating transcription and splicing patterns of *MEG3* in different species, bioinformatic analyses provide us with a steppingstone for exploring the potential of evolutionary conservation of the *MEG3* functional mechanism.

To investigate the extent of evolutionary conservation of *MEG3* through unbiased, experimental evidence, I characterised murine *Meg3* structurally and functionally.

Biophysical characterisation of *mMeg3v3* supported that it consistently folds into a, compact structure in response to increasing $[Mg^{2+}]$. Secondary structural characterisation showed that, like human *MEG3*, *mMeg3v3* organised in distinct domains (Figure 3:27). The structure of helix H11 (H15 in the mouse), was also conserved and in addition the SHAPE reactivity pattern of the loop residues on helix H15 was compatible with the formation of pseudoknot interactions (Figure 3:28). I did not yet provide strong evidence supporting or disproving the formation of a pseudoknot between the loop nucleotides of helix H15 and the TRs. Hence, I propose the next step in the study of *mMeg3*, should focus on this region. For instance, we could follow the same strategy as with human *MEG3v1* and perform SHAPE, AUC and AFM on the non-pseudoknot forming mutant (Uroda *et al*, 2019). In addition, we could use SHAPE-JuMP to explore the long-range interactions formed in *mMeg3* (Christy *et al*, 2021) (see chapter 3.1.5).

Through functional analysis of *mMeg3* by our luciferase reporter system, I showed that the activity of *mMeg3* as a regulator of p53-dependent gene expression is species specific since *mMeg3* activated reporter gene expression in mouse cells but not in human cells. Interestingly, the human *MEG3v1* was able to drive p53-dependent gene expression in the murine cell line Hepa1-6. Moreover, I showed that this regulatory mechanism does not depend on the species origin of p53 itself. Overexpression of exogenous human or mouse p53 upregulated p53-dependent gene expression in the human cell line HCT116 p53^{-/-} in the presence of human *MEG3v1* but failed to do so in the presence of mouse *Meg3*. Using the murine cell line Hepa1-6, I could show some initial evidence supporting that *mMeg3* organised in domains and that domain organisation was important for the regulation of the functional effect of *mMeg3* on the p53 pathway. Taken together these results support that the lncRNA *Meg3* can act as a regulator of p53-dependent gene expression in mouse cells. The mechanism of *Meg3*-mediated p53-dependent gene expression seems to depend on the relative domain organisation of the lncRNA similarly to what we observe in the human system. I also provided initial evidence suggesting that the mechanism of *Meg3*-driven p53-dependent gene expression in murine cells relies on different active motifs compared to the human system. Taken together, my functional characterisation of murine *Meg3v3* in light of its domain organisation suggests that *mMeg3* can regulate p53-dependent gene expression via a mechanism largely

different than this followed by its human counterpart, *MEG3*. Conservation of function despite mechanistic divergence is not uncommon in lncRNAs from different species. For instance, the lncRNA *JPX* is poorly conserved from humans to mice both in terms of its nucleotide sequence and its secondary structure (Kärner *et al*, 2020). Nevertheless, in both species, *JPX* maintains its capacity to interact with CTCF and regulate random X-chromosome inactivation (Kärner *et al*, 2020).

Admittedly, there is still a lot of unknown factors to discover before we can fully understand the level of mechanistic conservation for *Meg3*-mediated, p53-dependent gene expression. For example, why did human *MEG3* activate p53-dependent gene expression in the murine system but *mMeg3* failed to do so in human cells? The set of genes regulated by p53 in humans is very different than this in mice (Fischer, 2019). Protein pathways are generally conserved across species. This makes sense since protein pathways are absolutely essential for the survival of cells and the function of organisms. Our conserved protein pathways between human and mice could be viewed as (evolutionarily speaking) what define us as mammals. However, with so many well conserved proteins and protein pathways, it may be difficult to explain genetically, what drives the variation between different mammals. Divergence in the p53 gene regulatory networks between the two mammals has mainly occurred via mutations on the p53 response elements of target genes (Fisher, 2019). Since the same reporter plasmid was used for the activity assays in both human and murine cell lines, we might be looking at a secondary level of divergence which does not depend on DNA regulatory elements such as REs but on lncRNAs. Non-protein coding elements of the genome are under lower evolutionary pressure and thus lncRNA homologs from different species commonly exhibit more differences than homologous proteins do. As mentioned earlier (see chapter 1.2), lncRNAs are very tightly regulated; their expression can be restricted to specific tissues, cell-types and subcellular compartments, their transcription level and parental origin are strictly regulated. Finally, there are so many ncRNAs in each cell that it wouldn't be surprising if some ncRNAs were present for the sole purpose of fine-tuning protein pathways. Taken together, the higher evolutionary turnover and the tight spatio-temporal regulation of lncRNAs suggest that perhaps lncRNAs are there to provide an extra layer of regulation on conserved molecular pathways that is unique to a species or even cell-type.

Chapter 4. Discussion

4.1 Résumé en français

Dans ce chapitre, je résume les résultats de mon projet de thèse et je suggère des perspectives d'avenir possibles pour l'achèvement de notre compréhension de la biologie de *MEG3*. Enfin, je termine ce chapitre en proposant un mécanisme pour l'activité de *MEG3* dans la régulation de l'expression des gènes. En bref, la régulation minutieuse de la transcription et de l'épissage de *MEG3* détermine la composition de ses exons. La composition des exons détermine l'organisation des domaines des isoformes respectives de *MEG3* et, par la suite, l'interaction des domaines individuels formant des structures tertiaires. Enfin, les différents conformères de structure tertiaire de *MEG3* interagiront avec des protéines dans le noyau cellulaire et guideront l'activité des protéines pour réguler d'autres gènes *en trans*.

4.2 Summary, discussion and future outlook

The work presented in this thesis provided a comprehensive study of the long non-coding RNA *MEG3*. I described the characterisation of *MEG3* at the structural, functional and evolutionary levels. In summary, my work helped to reveal that, *MEG3* is organised in distinct domains and that its relative domain organisation depends on exon composition. In addition, together with my colleagues, we identified the active core of *MEG3* which is formed by a pseudoknot between the loop residues of helix H11 on domain D2 (nomenclature for canonical isoform *MEG3v1*) and either one of the six receptor sequences on domain D3. We named the pseudoknot receptors tandem repeats (TRs) because they were found one after the other. The formation of this pseudoknot is essential for the function of *MEG3* in the p53 pathway in all splicing isoforms of *MEG3* that we studied. This finding was in line with the previous observation showing that minimal activity of *MEG3* required the combination of domains D2 and D3 (Uroda *et al*, 2019). By hydroxyl-radical probing experiments, I showed that the formation of the H11-TR pseudoknot affects the entire spatial organisation of *MEG3*. This result was complementary to my colleague's work who studied *MEG3* by atomic force microscopy. Through detailed functional characterisation of the TRs, helped by site-directed mutagenesis forcing pseudoknot formation by a pre-defined TR whilst keeping the amount of point mutations to the minimum, I showed that the choice of the pseudoknot-forming TR has the potential to shift the arrangement of folded *MEG3* towards an orientation that may favour the interaction with certain p53-responsive elements more than with others. However, this assay suggested that the choice of the pseudoknot-forming TR is not the sole determinant of *MEG3*-driven specificity of p53-dependent gene regulation for the majority of p53-responsive promoters. Taken together, this part of my PhD work contributed towards connecting the structure and function of *MEG3* and in extent to understanding its functional mechanism.

There is still a lot to learn about the functional mechanism of *MEG3*. One major question that remains unanswered, concerns its interaction network. In the literature there is conflicting evidence about the direct interaction of *MEG3* with p53 protein. The Klibanski lab suggest that p53 interacts directly with *MEG3* based on results from RNA immunoprecipitations (RIP), *in vitro* interaction assays and proximity localisation microscopy assays (Zhu *et al*, 2015; Bauer *et al*, 2021). In my opinion,

the evidence is nevertheless weak. The RIP assay relies on immunoprecipitation of a protein from whole cell lysates and identification of specific RNAs in the purified sample by RT-qPCR. This method can carry many artefacts since RNA can be notoriously sticky on protein surfaces or pulldown resin, especially if the proteins are positively charged. The *in vitro* interaction assays included *in vitro* transcription and biotinylation of *MEG3* followed by artificial refolding of the RNA. Then, whole cell lysate or recombinant purified p53 protein were added in excess to the purified RNA. The RNA complexes formed were further purified by affinity purification and then the presence of interacting proteins determined by immunoblotting (Zhu *et al*, 2015). Due to the artificial folding of this approach, there are concerns about the conformation of *MEG3* not being biologically relevant. Moreover, adding protein in excess in this *in vitro* set up may lead to non-specific binding of p53 onto the RNA without necessarily being a biologically relevant interaction. Finally, whilst the fluorescence microscopy experiment provided undoubtable evidence for the co-localisation of *MEG3* with p53, it does not prove *in cellulo* zero-distance which would be necessary for a direct interaction. On the other hand, other labs including ours presented evidence rejecting the hypothesis of direct p53-binding on *MEG3* based on p53 individual-nucleotide cross-linking and immunoprecipitation (iCLIP) and p53-targeted RNA immunoprecipitation (Riley & Maher, 2007a; Uroda *et al*, 2019). All in all, even though there is some evidence supporting a direct and an indirect p53-*MEG3* interaction, it is still unclear how *MEG3* influences the promoter selectivity during p53-dependent gene expression.

The ultimate answer to whether *MEG3* is bound directly by p53 or not will come by analysis of the entire protein interactome of the lncRNA. For instance, we could adapt the interactome capture protocol which identifies proteins associated with all polyadenylated RNAs inside cells, to target *MEG3* specifically (Castello *et al*, 2012). Interactome capture relies on UV-light mediated cross-linking of RNA-protein complexes *in vivo* before extracting polyadenylated RNA by oligo-deoxythymine covered magnetic beads and identifying the co-purified proteins by mass-spectrometry (Castello *et al*, 2012). UV-light cross-linking is very powerful in the study of RNA-protein complexes because it only cross-links nucleic acids to proteins found a distance less than 0.1 Å and without any significant base preference, leading to capture of true *in vivo* interactions (Castello *et al*, 2012). An approach to identifying

proteins interaction with a specific RNA, used MS2 loops to tag the RNA of interest (Slobodin & Gerst, 2010). MS2 loops are an RNA secondary structural element first identified in the MS2 bacteriophage genome, they are specifically recognised by the MS2 coat proteins (MS2-CP) of the bacteriophages and can be harnessed by researchers for the isolation of specific RNAs (Johansson *et al*, 1997). By fusing the MS2-CP with green fluorescent protein (GFP) and streptavidin binding protein (SBP) the MS2-tagged RNA of interest can be visualised inside the cell and specifically purified respectively (Slobodin & Gerst, 2010). By coupling MS2 loop fusion with formaldehyde-induced crosslinking of proteins to nucleic acids and mass-spectrometry to identify the co-eluted proteins, it was possible to identify the protein interactomes of *ASH1* and *OXA1* mRNAs in yeast and that of *β -Actin* mRNA within human cells (Slobodin & Gerst, 2010). Since formaldehyde can cross-link proteins to other proteins as well as proteins to nucleic acids, it is possible that *MEG3* was detected in the purified sample through another bridging entity (RNA or protein). For the study of *MEG3* and its interacting proteins, I would suggest using elements from those two protocols and purify MS2 loop-tagged *MEG3* after UV-light cross-linking, followed by mass spectrometry of the co-eluted protein partners. Identifying the proteins which interact with *MEG3* inside cells will cast light on the exact mechanism of function of *MEG3* and explain how *MEG3* regulates p53-dependent gene expression independent of a direct or indirect *MEG3*-p53 contact. Another possible method aiming to capture proteins that interact with *MEG3* would be to utilise MS2 loops in order to tether *MEG3* to the engineered peroxidase APEX2 resulting in the biotinylation of *MEG3*-interacting proteins (Han *et al*, 2020). Subsequently, biotinylated proteins can be isolated from the cell by affinity purification using streptavidin beads and analysed by mass spectrometry. Notably, there has already been an effort to characterise the proteins binding *MEG3* using MS2 loops (Liu *et al*, 2015). However, this study lacked validation of the identified hits and therefore the results need to be taken with caution until sufficient validation has been provided.

Some members of the lncRNA research community argue that fusing MS2 loops on a lncRNA can affect its native folding and thus alter its function inside the cell. Those concerns are reasonable however, at least in the case of *MEG3* there has not been any evidence supporting that MS2 loop has or does not have any effect on *MEG3* function. Hence, before planning any large-scale interactions experiment the effect

of MS2 fusion on *MEG3* function should be tested. In the case MS2 loop fusion is proven to interfere with the biological function of *MEG3*, then other techniques for the identification of interacting proteins can be used (reviewed in Barra & Leucci, 2017). For instance, complementary oligonucleotides could be used to isolate *MEG3* complexes from inside cells following the Capture Hybridization Analysis of RNA Targets (CHART) (Simon *et al*, 2011).

By identifying the interaction network of *MEG3* we can begin to understand its functional mechanism and involvement in development and in disease. It is not unlikely that the lncRNA follows different paths to perform its developmental and growth regulatory functions. For example, whilst our laboratory found that exon E3 was indispensable for the action of human *MEG3* within the p53 pathway, the Klibanski lab showed that mouse embryos with exons E2, E3 and E4 deleted from their genome were viable (Zhu *et al*, 2019). Since the aim of the latter study was to determine the imprinting mechanism of the *Dlk/Meg3* locus, it does not report on whether those pups developed cancers or had any other growth regulation related diseases in their adulthood. Hence, my hypothesis is that *MEG3* has distinct mechanisms of function during development and growth control regulation. Accordingly, its relevant active motifs may differ between the two functions and the same may be true about its protein interactome. Those potential differences highlight the value of our *in vitro* approach for the structural characterisation of *MEG3* and its splicing isoforms. Even though undeniably the *in vivo* structures of any RNA are more biologically relevant, our *in vitro* structural map provided us with an unbiased scaffold for the identification of active motifs without the influence of proteins and other cellular factors that may be only present in the cell type and developmental stage of the cells used for the experiment.

Another still elusive mechanistic detail of *MEG3* is its DNA-targeting and specificity. One study suggested that *MEG3* targets chromatin by forming RNA-DNA triple helices (Mondal *et al*, 2015). This particular study focused on the gene regulatory effect of *MEG3* through interactions with the *Polycomb* group family proteins and therefore we do not know whether this mechanism also applies to *MEG3*-driven, p53-dependent gene expression. Again, information on the interactome of *MEG3* can illuminate the possibilities for DNA targeting. For instance, if a specific transcription

factor was found amongst the *MEG3* interactome it could be a likely bridge between *MEG3* and DNA.

In this study, I focused my efforts to functionally characterise *MEG3* on its effect on the p53 pathway. The active core formed by the helix H11-TR pseudoknot was functionally validated only in the context of p53-dependent gene expression. However, *MEG3* has also been characterised to interact with proteins of the *Polycomb* group family (see chapter 1.3.2). We do not yet know whether the helix H11-TR pseudoknot is involved in regulating the function of *MEG3* in this pathway. Moreover, it has not been explored whether *MEG3* splicing isoforms have any difference in their respective abilities to drive *Polycomb*-dependent chromatin regulation. It would be very interesting to see whether the active core formed by the helix H11-TR pseudoknot is a universal active motif of *MEG3* or whether its effect is not involved in the *Polycomb*-dependent functions of *MEG3*. To do this, we should couple a systematic mutagenesis assay with functional characterisation targeting *Polycomb* proteins as we did for the p53 pathway in this study. This may be a laborious method but can generate clear connections between structural elements and functional effects.

Finally, to complete our understanding of *MEG3*-dependent gene regulation we should explore whether different splicing isoforms are linked to different developmental stages or specialised tissues. In fact, different human tissues seem to contain different relative abundances of the various *MEG3* splicing isoforms (Zhang *et al*, 2010). Perhaps variation comes due to the presence of different interacting proteins during different developmental stages or within different tissues. In mice, the splicing of *mMeg3* is tightly regulated by the presence of NOVA proteins which themselves are strictly neuronal proteins thus resulting in the active form of *mMeg3* present in the mouse nervous system but not other cell types (Ule *et al*, 2003). Whether such a mechanism also exists in humans is not known. Once again understanding the *MEG3* interactome can help elucidate these questions especially if we could identify proteins that interact with some *MEG3* splicing isoforms but not others.

Furthermore, I described my efforts towards acquiring a high-resolution tertiary structure of *MEG3* which unfortunately did not yield a structural model due to the high flexibility of the lncRNA. Inevitably, characterising the RNA-protein interaction

network of *MEG3* can provide an array of possible *MEG3*-protein complexes which could be better targets for structural studies. Other RNAs have been studied in the context of their RNPs in the past even though they are too flexible in isolation. For instance, the structure of ribosomal RNA has been solved in the context of the entire ribosome (Ramakrishnan, 2002).

Finally, I described my work on exploring the evolution of the *MEG3*-driven pathway for p53-dependent gene regulation. By sequence and structure-based alignments, I showed that *MEG3* has the potential to maintain its H11-TR driven active core across mammals. I then focused on providing experimental evidence for the extent of evolutionary conservation of *MEG3* biology across humans and mice. By functional characterisation, I found that *mMeg3* is a species-specific regulator of p53-dependent gene expression. In comparison, the human *MEG3v1* did not exhibit species bias as it was able to drive p53-dependent gene expression in human and in murine cells. Functional characterisation of *mMeg3* showed that *mMeg3* organised in domains and that domain organisation was important for functional regulation of *mMeg3* in the p53 pathway. Although the functional characterisation of *mMeg3* is still in progress, thus far my results support that the mechanism of *mMeg3* for the regulation of p53-dependent gene expression depends on its relative domain organisation. This is in line with the hypothesis of functional and mechanistic conservation of *MEG3*-mediated p53-dependent gene expression across humans and mice. However, since this part of my work is not yet completed no final conclusions can be drawn. Next steps in the functional characterisation of *mMeg3* will include a careful dissection of its active motifs as they appear in its secondary structural map and characterisation of its active core as exemplified by our studies of human *MEG3*. Through such experiments we can systematically connect the structural features of *mMeg3* to its functional effect and start comparing its mechanistic similarities and differences to the human *MEG3*.

Altogether, my work on human *MEG3* and its splicing variants contributed to establishing a mechanism for the function of *MEG3* in fine-tuning the p53 pathway for the selective upregulation of cell growth related genes. In brief, I suggest a mechanism where careful regulation of *MEG3* transcription and splicing of *MEG3* determine its exon composition. Exon composition determines the domain organisation of respective *MEG3* isoforms and subsequently this will determine the

interaction of individual domains forming tertiary structures. Finally, the different tertiary structural conformers of *MEG3* will interact with proteins in the cell nucleus and will guide the activity of the proteins to regulate other genes *in trans* (Figure 4:1).

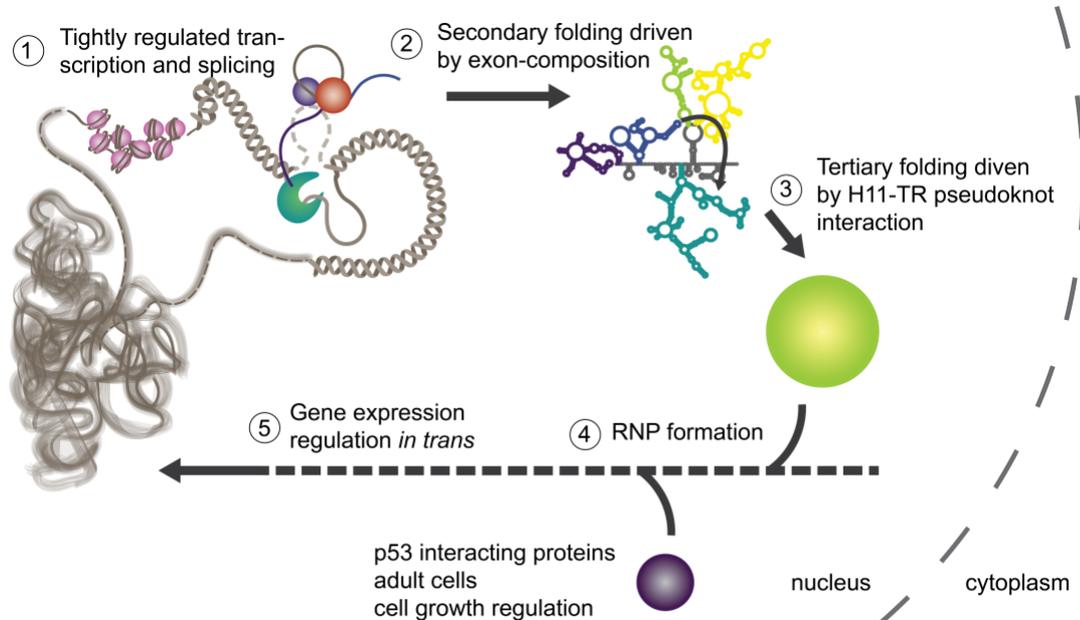


Figure 4:1 Cartoon representation of the proposed mechanism for *MEG3*-driven gene expression regulation in trans

(1) The expression pattern and splicing of *MEG3* are tightly regulated by imprinting of the genome. (2) Secondary folding of *MEG3* occurs likely co-transcriptionally and depends on its exon-composition. (3) Tertiary folding of *MEG3* is driven by the interaction of helix H11 and the tandem repeats (TRs). A pseudoknot motif is formed holding the tertiary fold of *MEG3*. The final conformation of folded *MEG3* will depend on its domain organisation as dictated by exon-composition and on the choice of pseudoknot-forming TR. (4) Folded *MEG3* interacts with proteins in the nucleus. (5) Ribonucleoprotein complexes act on the genome *in trans* to regulate expression of genes critical for embryonal development and tissue homeostasis (cancer) in the adult.

Chapter 5. Appendix

5.1 Appendix 1: Human *MEG3* splicing isoforms

Table 5:1 Annotations of human *MEG3* splicing isoforms

NCBI accession number	<i>MEG3</i> variant number	Klibanski lab letter code	Length (nt)
NR_002766.2	v1	<i>MEG3</i>	1,595
NR_003530.2	v2	-	1,855
NR_003531.3	v3	a	1,731
NR_033358.1	v4	-	1,735
NR_033359.1	v5	-	1,714
NR_033360.1	v6	b	1,621
NR_046464.1	v7	c	1,687
NR_046465.2	v8	d	1,653
NR_046466.1	v9	e	1,506
NR_046467.1	v10	f	1,793
GQ183499.1	v11	g	1,745
NR_046469.1	v12	h	1,656
NR_046470.2	v13	i	1,785
NR_046471.1	v14	j	1,547
NR_046472.1	v15	k	1,687

5.2 Appendix 2: Reporters plasmids with p53 response elements

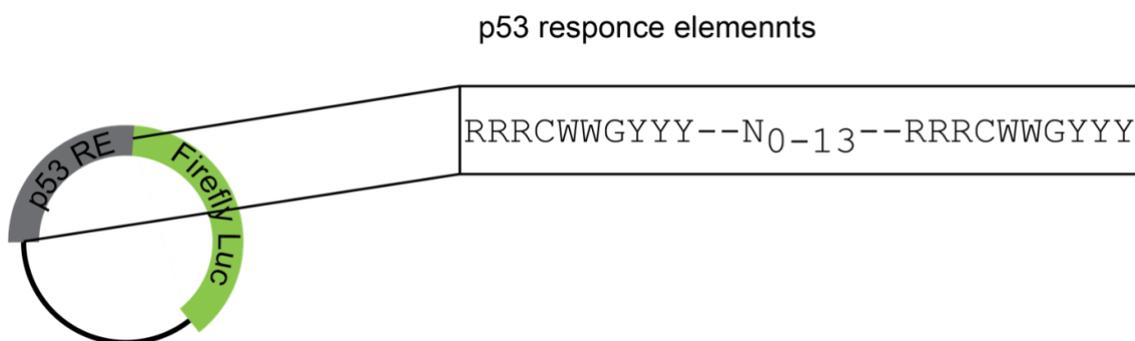


Figure 5:1 Representation of p53-responsive reporter plasmids

The Firefly luciferase gene (green) is controlled by a p53-responsive cassette containing a p53 response element (RE-grey box) followed by a promoter sequence.

Table 5:2 List of genomic p53 response elements used in this work

Name	p53 RE	Comments
01-pGL3-promoter	SV40 promoter	entire promoter cassette
02-pGL3-1138	GAACATGTC-N1-CAACATGTT	p53 RE of p21 gene
03-pGL3-1012	TCACAAGTT-N1-AGACAAGCC	
04-pGL3-MDM2	GGTCAAGTT-N1-GGACACGGT	has an additional non-canonical p53-RE
05-pGL3-p21-5'	GAACATGTC-N1-CAACATGTT	only 5' part of promoter cassette
06-pGL3-AIP1	GAGCATGTT-N1-ACACAAGCC	
07-pGL3-FLT-p21-5'	GAACATGTC-N1-CAACATGTT	only 5' part of p21 RE in FLT promoter
08-pGL3-FLT-AIP1	GAGCATGTT-N1-ACACAAGCC	AIP1 gene RE in FLT promoter
09-pGL3-FLT-GADD45	AGACAAGAC	non-canonical second repeat

5.3 Appendix 3: Sequence alignment of human and murine p53 proteins

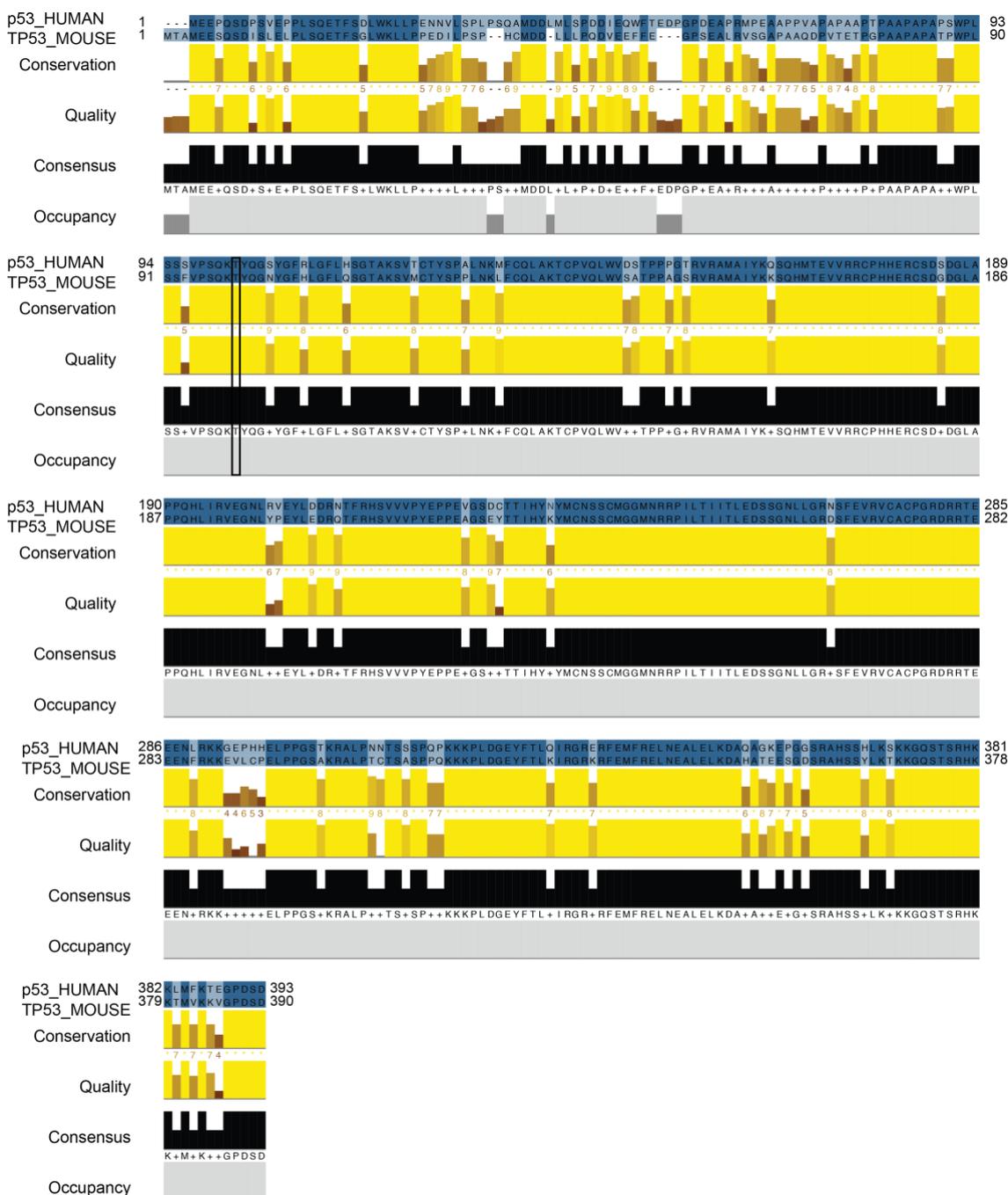


Figure 5:2 Sequence alignment of human and murine p53 proteins

The amino acid sequences of human and murine p53 were aligned using Clustal Omega and visualised using the Java applet Jalview. The amino acids are highlighted in shades of blue based on sequence similarity from dark to opaque corresponding to most and least conserved accordingly. The yellow and black bars represent the percent of conservation, the quality of the alignment and the consensus sequence respectively. The representative consensus sequence is marked in black. The occupancy of the consensus sequence is shown as grey bars.

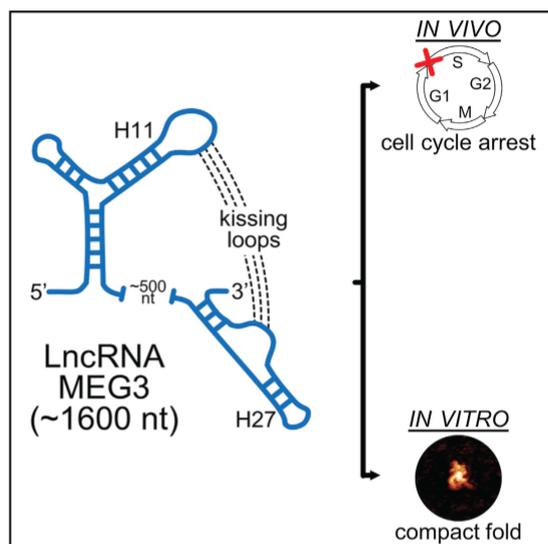
5.4 Appendix 4: Publications

Article

Molecular Cell

Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway

Graphical Abstract



Authors

Tina Uroda, Eleni Anastasakou, Annalisa Rossi, ..., Alberto Inga, Isabel Chillón, Marco Marcia

Correspondence

ichillon@embl.fr (I.C.),
mmarcia@embl.fr (M.M.)

In Brief

Uroda et al. identify the structural and functional core of the human lncRNA MEG3, revealing that point mutations designed to disrupt the MEG3 fold impair function. The work proves the importance of lncRNA tertiary structure motifs for selective and specific control of key cellular processes such as the p53 stress response.

Highlights

- lncRNA MEG3 is a tumor suppressor that stimulates the p53 pathway
- The p53-stimulating core of MEG3 comprises of two conserved, structured domains
- Two distal motifs in the MEG3 core form pseudoknot interactions ("kissing loops")
- Mutations in these pseudoknots disrupt MEG3 architecture and impair its function

Uroda et al., 2019, Molecular Cell 75, 1–14
September 5, 2019 © 2019 The Authors. Published by Elsevier Inc.
<https://doi.org/10.1016/j.molcel.2019.07.025>

CellPress

Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway

Tina Uroda,¹ Eleni Anastasakou,¹ Annalisa Rossi,² Jean-Marie Teulon,³ Jean-Luc Pellequer,³ Paolo Annibale,⁴ Ombeline Pessey,¹ Alberto Inga,² Isabel Chillón,^{1,*} and Marco Marcia^{1,5,*}

¹European Molecular Biology Laboratory (EMBL) Grenoble, 71 Avenue des Martyrs, Grenoble 38042, France

²Department CIBIO, University of Trento, via Sommarive 9, 38123 Povo (Trento), Italy

³Université Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), 38000 Grenoble, France

⁴Max Delbrück Center, Robert-Rössle-Straße 10, 13092 Berlin, Germany

⁵Lead Contact

*Correspondence: ichillon@embl.fr (I.C.), mmarcia@embl.fr (M.M.)

<https://doi.org/10.1016/j.molcel.2019.07.025>

SUMMARY

Long non-coding RNAs (lncRNAs) are key regulatory molecules, but unlike with other RNAs, the direct link between their tertiary structure motifs and their function has proven elusive. Here we report structural and functional studies of human maternally expressed gene 3 (MEG3), a tumor suppressor lncRNA that modulates the p53 response. We found that, in an evolutionary conserved region of MEG3, two distal motifs interact by base complementarity to form alternative, mutually exclusive pseudoknot structures ("kissing loops"). Mutations that disrupt these interactions impair MEG3-dependent p53 stimulation *in vivo* and disrupt MEG3 folding *in vitro*. These findings provide mechanistic insights into regulation of the p53 pathway by MEG3 and reveal how conserved motifs of tertiary structure can regulate lncRNA biological function.

INTRODUCTION

Long non-coding RNA (lncRNA) structures are increasingly being recognized as important modulators of cellular processes, including chromatin remodeling, DNA repair, and translation (Mercer et al., 2009). Among the more than 32,000 human lncRNAs (Volders et al., 2013), a subgroup emerged as particularly suited for mechanistic studies based on their evolutionary conservation (Necsulea et al., 2014), specific cellular distribution (Cabilli et al., 2015), tissue localization (Kaushik et al., 2013), and clinical relevance (Sauvageau et al., 2013; Wapinski and Chang, 2011). The secondary structures of a handful of such lncRNAs have been experimentally mapped, including SRA (Novikova et al., 2012b), HOTAIR (Somarowthu et al., 2015), XIST (Smola et al., 2016), RepA (Liu et al., 2017), roX (Ilik et al., 2013), BRAVEHEART (Xue et al., 2016), COOLAIR (Hawkes et al., 2016), NEAT1 (Lin et al., 2018), and parts of lincRNA-p21 (Chillón and Pyle, 2016). However, unlike with other RNAs, it has not yet

been possible to systematically connect information regarding tertiary structure motifs of lncRNAs with their biological function, partly because the size and complexity of these molecules present unprecedented challenges for biophysical studies.

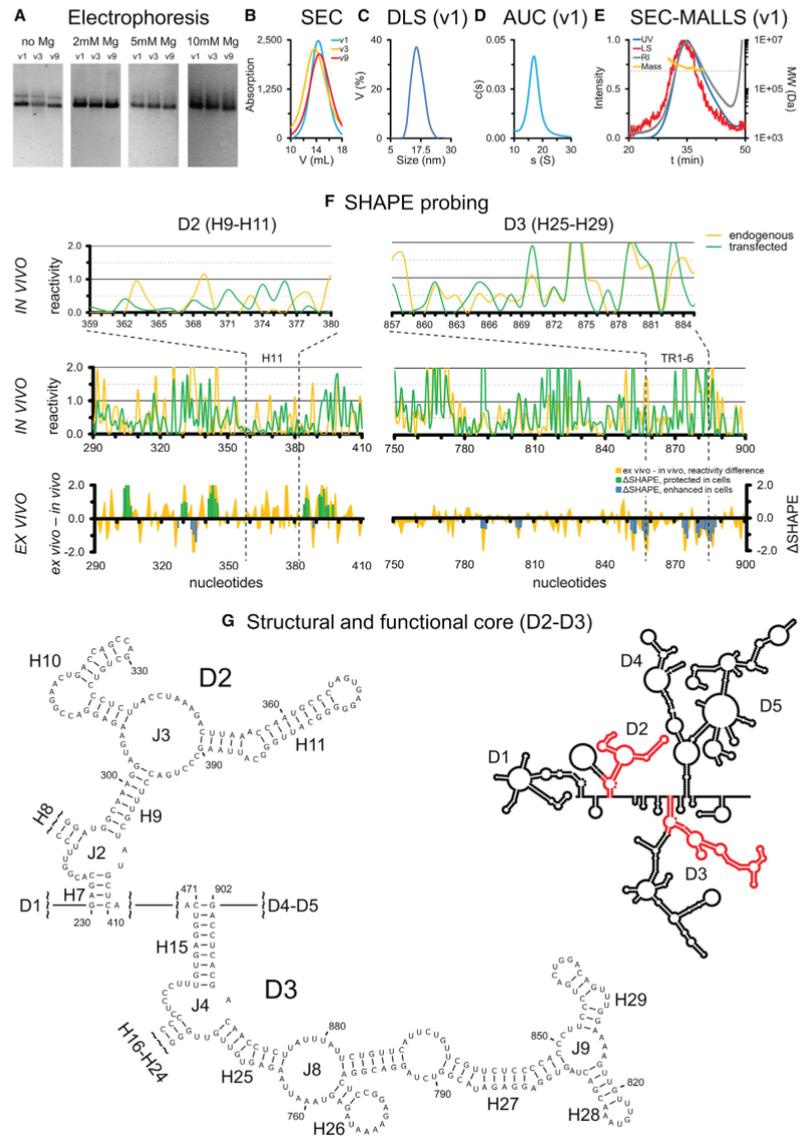
Human maternally expressed gene 3 (MEG3) is an alternatively spliced nuclear lncRNA abundant in the brain, placenta, and endocrine glands (Mondal et al., 2015; Zhang et al., 2003). MEG3 is expressed under the control of differentially methylated promoters from the Dlk1-MEG3 imprinted locus on chromosome 14q32, which also encodes other ncRNAs, none of which overlap with MEG3 exons (McMurray and Schmidt, 2012; Miyoshi et al., 2000). In embryonic cells, where it is not imprinted (McMurray and Schmidt, 2012), MEG3 silences genes involved in neurogenesis by regulating chromatin targeting of *Polycomb* proteins, and MEG3 expression is needed during neuronal development (Kaneko et al., 2014; Mercer et al., 2008; Mondal et al., 2015). Instead, in adult cells, where it becomes imprinted, MEG3 stimulates the p53 pathway, inducing cell cycle arrest and apoptosis (Zhou et al., 2007). In most human cancer cell lines and certain primary tumors, such as pituitary adenoma, MEG3 is downregulated via hypermethylation of the maternal allele, but its ectopic expression reduces tumor progression; thus, MEG3 acts as a tumor suppressor (Cheunsuchon et al., 2011; Zhou et al., 2012). Therefore, understanding the molecular mechanism of MEG3 is crucial to improve our knowledge of specific p53-related carcinogenic pathways.

In vitro and *in vivo* studies suggest that MEG3 interacts with p53 protein, leading to selective upregulation of p53 target genes (Zhou et al., 2007; Zhu et al., 2015). The 27 known splice variants of MEG3, which contain variable middle exons flanked by common exons at the 5' (E1-E3) and 3' (E10-E12) ends, vary in their ability to stimulate the p53 pathway (Zhang et al., 2010b). Changes in the MEG3 splicing pattern under stress lead to fluctuations in the p53 stress response (Zhang et al., 2010b). Interestingly, deletion mutagenesis of MEG3 impairs stimulation of the p53 pathway, suggesting that specific regions of this lncRNA are important for the p53 response (Zhang et al., 2010b; Zhou et al., 2007). However, the link between the structure of MEG3 and its functional effects on p53 remains to be defined.

To address this issue, we set out to characterize the secondary and tertiary structures of three MEG3 splice variants *in vitro*

Please cite this article as: Uroda et al., Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway, Molecular Cell (2019), <https://doi.org/10.1016/j.molcel.2019.07.025>

CellPress



(legend on next page)

and *in vivo*. Guided by cell-based functional assays, we identified the functional core of MEG3. We found that two distal motifs complementary to each other in sequence are strictly required for stimulation of the p53 pathway. Point mutations designed to break the long-range base-pairing interactions between these motifs and to perturb the global 3D fold of MEG3 severely impair its ability to stimulate the p53 pathway. These findings underscore the importance of specific MEG3 tertiary structural elements for stimulation of the p53 pathway and provide mechanistic insights into this important tumor suppressor lncRNA.

RESULTS

Three Splicing Variants of Human MEG3 Share a Common Structural Core that Is Evolutionarily Conserved in Mammals

We expressed and purified the most abundant MEG3 splice variant (v1 or MEG3) as well as the two variants that induce the lowest (v3 or MEG3a) and the highest (v9 or MEG3e) degree of stimulation of the p53 pathway. After confirming homogeneity (Figures 1A–1E; Table S1), we mapped the secondary structures of these three variants by *in vitro* selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) using 3 reagents (1-methyl-7-nitroisatoic anhydride [1M7], 1-methyl-6-nitroisatoic anhydride [1M6], and N-methylisatoic anhydride [NMIA]; Figures 1G and S1–S3). We then used a fourth reagent (dimethyl sulfate [DMS]) to validate the map of v1, which we take as the reference isoform for structural description.

Variant v1, which spans 1,595 nt defined by common exons E1–E3 and E10–E12 and by varying E5, adopts a modular organization, including 5 structural domains (D1–D5) whose boundaries localize close to exon junctions, so that D2 (nt 230–410) and D3 (nt 471–902) together comprise E3 (Figures S1A, S2A, and S3). D2 and D3 are also preserved in v3, which starts 24 nt downstream of v1 and spans 1,712 nt defined by variable E6 along with E1–E3/E5/E10–E12 (Figures S1B, S2B, and S3), and in v9, which spans 1,481 nt defined by common E1–E3/E10–E12 and lacks any variable middle exon (Figures S1C, S2C, and S3). The most structurally stable portion of MEG3 (D2–D3; Figure 1G) folds reproducibly independent of the computational algorithm used to compute its secondary structure (Superfold, Siegfried et al., 2014; or RNA structure, Reuter and Mathews, 2010; Figure S4).

This motif also corresponds to the most evolutionarily conserved region of MEG3. For instance, Rfam family RF01872 covers a region in E3 (H21–H23) common to 53 putative MEG3 sequences from 40 mammals. Using the sequence of E3 and BLAT (Kent, 2002), we identified the complete E3 in 46 mammals, covering all mammalian orders except Monotremata (orders

defined according to Tarver et al., 2016). MEG3 transcription has been confirmed and annotated in the NCBI for only 6 of these 46 species (human, orangutan, mouse, rat, cow, and pig). Besides E3, we also identified E1 and E2 in 33 species and E10–E12 in 19 species. We additionally identified E12 but not E10 or E11 in 21 species. We could not detect, or only partially detected E1 and E2 in Prosimians, Eulipotyphla, Xenarthra, Afrotheria, and Marsupialia, and we could not identify E10–E12 in Xenarthra and Marsupialia. Finally, we could not identify MEG3 beyond mammals (Table S2; Data S1). Because E3 is the most conserved region of MEG3, we aligned this region in infernal, including the corresponding secondary structure information from v1 (Figure 2). The final alignment includes 41 mammalian MEG3 sequences and reveals that the most conserved region is the H11 stem-loop structure, in which 6 bp and the entire terminal loop are invariant and 3 additional bp co-vary.

In summary, our data suggest that, in human MEG3 exon E3, which comprises domains D2 and D3, represents the most conserved portion in sequence and secondary structure.

Systematic Cell-Based Assays Dissect the Functional Contribution of Each Structural Domain of MEG3

To assess the relationship between the MEG3 secondary structure and the function of this lncRNA as a tumor suppressor, we performed cell cycle (Lu et al., 2013) and luciferase reporter assays (Zhang et al., 2010a) in HCT116, a human cell line that expresses wild-type p53 and negligible levels of endogenous MEG3 (Figure S5N). Cell cycle assays showed that MEG3 induces arrest specifically at the G1/S but not the G2/M checkpoint and does not induce apoptosis in HCT116 (Figures 3A–3C, 3F–3H, S5A–S5C, and S5F–S5H). Cell cycle arrest by MEG3 is exclusively p53-dependent because we did not observe any MEG3 effect in HCT116-p53^{-/-} cells (Figures 3I and S5I). Furthermore, luciferase reporter assays showed that MEG3 stimulates expression of p53 target genes. The MEG3 effect is dose-dependent and exclusively p53-dependent because luciferase production in HCT116-p53^{-/-} cells is minimal and identical in cells transfected with MEG3 or a control vector (Figures 3J and S5J–S5M). Importantly, the intensity of the MEG3 effect depends on the p53 response element (p53RE) in the promoter of the luciferase reporter gene (Figure 3K). We identified two reporters on which all three MEG3 variants are active: one possessing an optimized p53RE used in previous MEG3 studies (p53Luc; Zhang et al., 2010a) and another possessing the p53RE of the endogenous MDM2 gene (pGL-MDM2; Menendez et al., 2010).

Using the luciferase assay and the p53Luc reporter, we systematically probed the structure of v1 (Figures 4A–4D). We found that none of the v1 domains in isolation can activate p53 and that

Figure 1. The MEG3 D2–D3 Structural Core
(A and B) Native agarose gel electrophoresis (A) and size exclusion chromatography (SEC; B) of v1, v3, and v9.
(C–E) Dynamic light scattering (DLS; C), analytical ultracentrifugation (AUC; D), and SEC coupled to multi-angle laser light scattering (MALLS; E) profiles of v1.
(F) SHAPE reactivity values of individual nucleotides in D2 and D3 (H11 and TRs motifs are delimited by the dotted vertical lines). Bottom: difference between *ex vivo* minus *in vivo* reactivity values and deltaSHAPE values (endogenous MEG3 datasets). Center: *in vivo* 1M7 reactivity values in endogenous MEG3 and transfected v1. Top: magnification of *in vivo* 1M7 reactivity values for H11 and the TRs. Structure maps and complete data from *in vitro* probing are reported in Figures S1–S4.
(G) Structure of selected motifs in the D2–D3 core (D indicates domains, H helices, and J junctions). Inset: schematic of the complete v1 structure (from Figure S1), with the core shown in red.

Please cite this article as: Uroda et al., Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway, *Molecular Cell* (2019), <https://doi.org/10.1016/j.molcel.2019.07.025>

CellPress

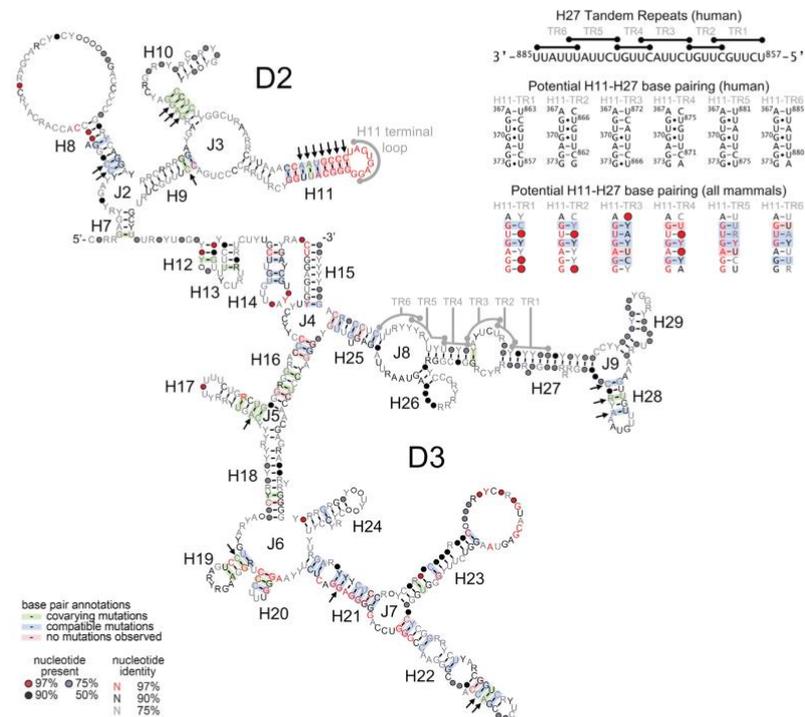


Figure 2. Evolutionary Conservation in the MEG3 Core

R2R plot of 41 D2-D3 (E3) sequences aligned in Infernal (color legend at the bottom left; Weinberg and Breaker, 2011). Arrows indicate covariant base pairs of potential statistical significance (see STAR Methods for details). Human sequences of the H27 TRs, corresponding base-pairing to the H11 terminal loop, and potential covariation of the base-pairing interaction are shown at the top right.

D1/D4/D5 are dispensable for stimulation of the p53 pathway. In contrast, deletion of either D2 or D3 (constructs Δ D2 and Δ D3) abolishes stimulation of the p53 pathway. Curiously, deletion of D4 or D5 enhances v1 activity. The same stimulation/inhibition pattern can be observed for individual exons. Deleting E5 (as in v9), E10, E11, or E12 stimulates the p53 pathway, but neither E1 or E2 nor E10–E12 in isolation can stimulate the p53 pathway. The only exon that can stimulate the p53 pathway in isolation is E3, which covers D2–D3. Interestingly, D2, which is inactive in isolation, can partially stimulate the p53 pathway when co-transfected with Δ D2, which is also inactive *per se*, suggesting that D2 and D3 functionally cooperate when supplied either in *cis* or in *trans* (Figure 4A).

Considering that D2–D3 is also the most conserved portion of MEG3, we systematically focused on this region for detailed functional probing. We found that deleting H8 (D2) or H17, H28, and H21–H23 (D3) does not have a significant effect on

activity, but deletion of H11 (D2) or H25–H29 (D3) dramatically reduces stimulation of the p53 pathway (Figure 4B). Remarkably, within H11, disrupting the helical stem abolishes p53 stimulation, which can be recovered to near wild-type levels by compensatory mutations, suggesting that the structure of this element is essential for function (Figures 4C and 4D). Furthermore, deleting the H11 terminal loop (nt 366–373 in v1) or mutating all of its residues to adenosines (mutant H11LpA) impairs stimulation of the p53 pathway. Most strikingly, nearly all point mutations at positions 368–372 in the H11 terminal loop (GUGAG motif) also abolish stimulation of the p53 pathway (Figure 4C).

We thus focused on the H11LpA and G³⁷⁰C mutants to analyze the effects of H11 mutations on p53 regulation. First, we found that H11LpA and G³⁷⁰C reduce luciferase expression not only from the p53Luc but also from the pGL-MDM2 promoter (~35% residual activity with respect to v1), and, most importantly, both mutants failed to arrest the cell cycle (Figures 3D and 3E). Second, we

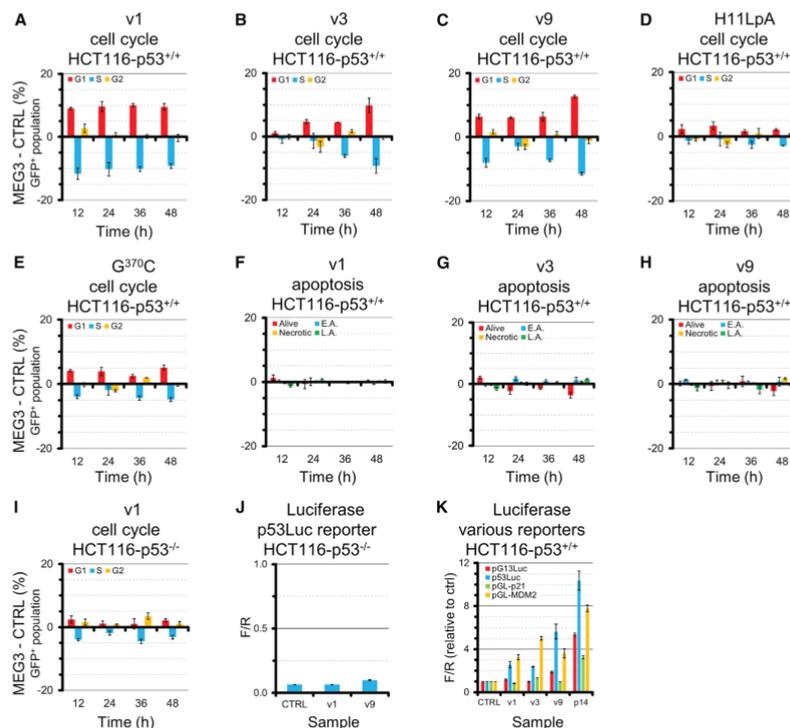


Figure 3. Selective Stimulation of the p53 Pathway by MEG3 Variants

(A–I) Cell cycle and apoptosis analysis of v1 (A), v3 (B), v9 (C), H11LpA (D), and G³⁷⁰C (E) in HCT116-p53^{+/+}, apoptosis analysis of v1 (F), v3 (G), v9 (H) in HCT116-p53^{+/+}, and cell cycle analysis of v1 in isogenic p53^{-/-} cell lines (I). E.A., early apoptotic; L.A., late apoptotic.

(J) Luciferase assay performed in HCT116-p53^{-/-} cells (absolute ratio of firefly luciferase versus *Renilla* luciferase chemiluminescence).

(K) Stimulation of the p53 pathway by v1, v3, v9, and p14^{ARF} on 4 reporter vectors possessing different p53REs (pG13Luc, p53Luc, pGL-p21, and pGL-MDM2). Data were normalized to the signal of corresponding empty vectors. For this experiment, 500 ng of MEG3 vectors and 50 ng of p14^{ARF} vectors were used for transfection in 12-well plates.

Error bars indicate SEM of 3 experiments.

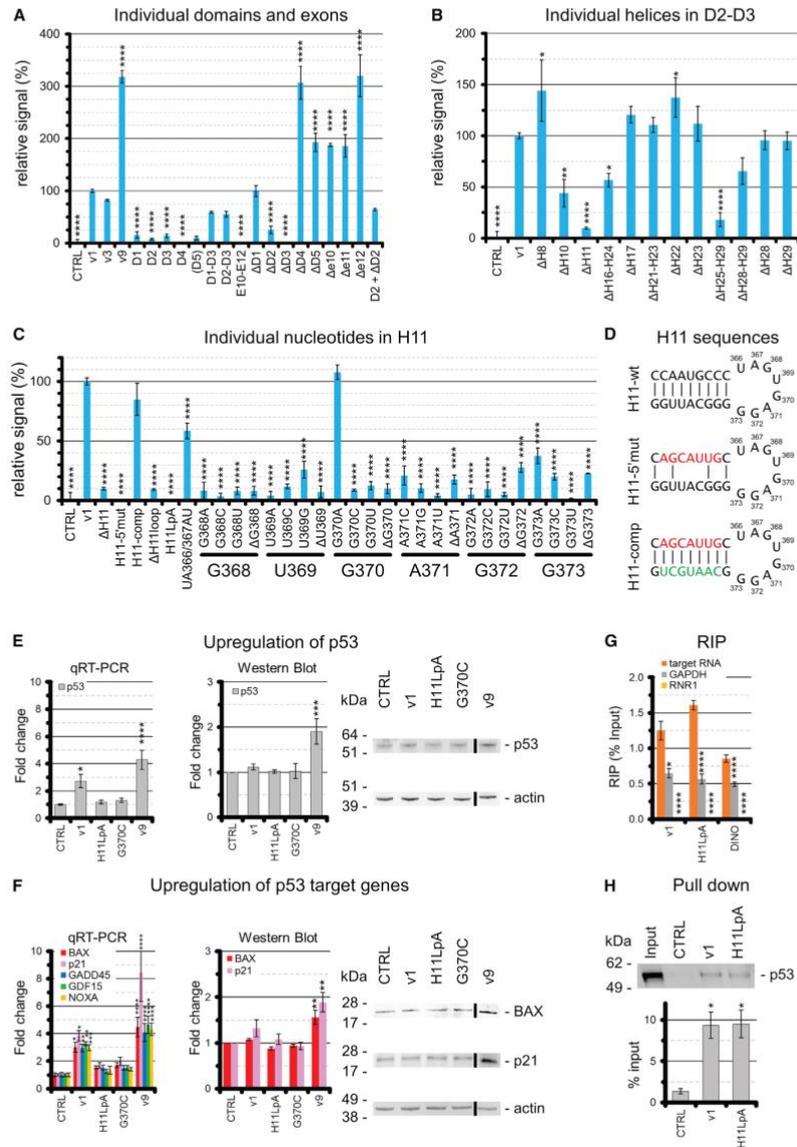
established, by qRT-PCR and western blot, that v1 induces p53 expression, as reported previously (Zhou et al., 2007), but the H11LpA and G³⁷⁰C mutants do not (Figure 4E). Finally, by qRT-PCR and western blot, we revealed that v1 induces expression of endogenous p53 target genes, in line with previous reports (Zhou et al., 2007), but the H11LpA and G³⁷⁰C mutants do not, as expected from our luciferase reporter data (Figure 4F).

In summary, the p53-stimulating core of MEG3 is formed by D2 and D3 and specifically involves the two structural motifs H11 (D2) and H25–H29 (D3). Both the structure of the highly conserved H11 stem and individual nucleotides in the invariant H11 terminal loop are essential for stimulation of the p53 pathway and for cell cycle regulation.

The Functional and Evolutionarily Invariant H11 Stem-Loop Structure Is Not a Protein-Binding Site

Because the effects of MEG3 on p53 and its target genes have been proposed to depend on a direct MEG3-p53 interaction (Zhou et al., 2007; Zhu et al., 2015), we performed RNA immunoprecipitation (RIP) and pull-down experiments to compare v1 with the H11LpA mutant and examine whether the functional importance of H11 is related to protein binding, as reported previously for other lncRNA structural motifs (Ilik et al., 2013; Lu et al., 2016; Somarowthu et al., 2015; Xue et al., 2016). By RIP, we found that p53 interacts to a similar extent with both v1 and H11LpA and comparably to lncRNA DINO (Schmitt et al., 2016; Figure 4G). Analogously, by pull-down using biotinylated RNA, we could

Please cite this article as: Uroda et al., Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway, Molecular Cell (2019), <https://doi.org/10.1016/j.molcel.2019.07.025>



(legend on next page)

detect similar levels of p53 bound to v1 and to H11LpA (Figure 4H). Thus, we conclude that H11 is not involved in p53 binding.

We subsequently analyzed whether H11 is involved in other intermolecular interactions in the cell. To this end, we performed *in vivo/ex vivo* chemical probing analysis. These techniques both compare the chemical reactivity of MEG3 folded in the cellular environment, either in the presence of protein partners (*in vivo*) or after gentle extraction and protein removal under experimental conditions that are not expected to denature the RNA secondary structure (*ex vivo*) (Smola et al., 2015a; Figure 1F). To overcome the challenge posed for cellular probing by the complex alternative splicing pattern of MEG3, in our study we used W138 fibroblasts to probe endogenous MEG3 (expression levels are reported in Figure S5N; expressed variants are v1 [68%] and v9 [26%]; Zhang et al., 2010b) and HCT116 cells to probe transfected v1 (expression levels of endogenous MEG3 in this cell line are negligible; Figure S5N). *In vivo* SHAPE reactivity values of individual nucleotides in the MEG3 core (D2-D3) in endogenous and transfected samples are similar (Figure 1F), suggesting that this region folds reproducibly in different cell lines. This observation reinforces the notion that D2-D3 is an important structural element of MEG3 and corroborates the use of the transfected system as a proxy to mimic near-physiological conditions for our functional assays. Surprisingly, within the H11 motif, we could not detect any significant reactivity difference between *ex vivo* and *in vivo* conditions (deltaSHAPE), suggesting that nucleotides in this region are unlikely to be involved in protein binding (Figure 1F).

In summary, our data reveal that non-functional H11 mutants preserve affinity for p53. More broadly, our *in vivo/ex vivo* SHAPE probing establishes that the H11 stem-loop structure is not a protein binding site, despite its high degree of sequence and structural conservation.

H11 Forms Conserved and Functional Pseudoknot Structures ("Kissing Loops") with H27

Seeking a molecular explanation for the functional importance of H11, we noticed that its terminal loop (residues 368–372, GUGAG motif) is chemically unreactive *in vitro* in all of our

SHAPE maps despite being single stranded (Figures S1 and S3). However, by performing SHAPE on the H11LpA mutant, we noticed that this motif becomes highly reactive, whereas the reactivity of all other structural regions except H25–H29 remains nearly unchanged (Figure 5A). Considering these differences in *in vitro* SHAPE reactivity and the fact that, in the cell, D2 can stimulate the p53 pathway *in trans* with Δ D2 (Figure 4A), we hypothesized that MEG3 may adopt a higher-order structure in which H11 is constrained by intramolecular long-range tertiary interactions.

To address this hypothesis, we first analyzed the folding behavior of v1 and H11LpA in solution by analytical ultracentrifugation (AUC). Strikingly, these two constructs, which differ by only 6 of 1,595 nt, display the same hydration radius (R_h) at Mg^{2+} concentrations below the $C_{Mg^{1/2}}$, but above the $C_{Mg^{1/2}}$, the mutant compacts up to 8% less than the wild type (Figure 5B).

Having established by AUC that MEG3 folding *in vitro* is specific, we performed hydroxyl radical footprinting (HRF). We found that the D2-D3 core of v1 displays a solvent accessibility pattern typical of highly structured RNAs, with solvent-protected regions flanked by highly solvent-exposed residues (Figures 5C, S6A, and S6B). The number of solvent-protected nucleotides in MEG3 is proportional to that of other highly structured large RNAs (Figure S6C). These observations suggest that MEG3 folds in a specific conformation *in vitro*. Moreover, HRF analysis of H11LpA revealed that both the GUGAG motif and the H25–H29 motif, which display increased SHAPE reactivity (see above and Figure 5A), also become significantly more solvent-exposed than in v1 (Figure 5D).

Curiously, nt 857–885 in H25–H29 (3' side of H27) comprise a conspicuous series of 6 tandem repeats (TR1–TR6) with sequences complementary to the GUGAG motif in H11, suggesting that H11 and H27 could base pair with each other to form long-range pseudoknot structures (kissing loops; Figure 2). Importantly, the potential to form H11–H27 pseudoknots is conserved in evolution because all mammalian sequences where we could identify MEG3 possess at least three such tandem repeats (Figure 2; Table S2). To test whether the putative H11–H27 pseudoknots actually form *in vivo* and are functionally relevant, we used the functionally impaired $G^{370}C$ mutant and introduced a

Figure 4. Functional Importance of MEG3 Structural Motifs

(A–C) p53-dependent luciferase assays using the p53Luc reporter on MEG3 variants, individual exons, and domains (A); on D2-D3 mutants (B); and on selected H11 mutants (C). Construct D5 is indicated in parenthesis because it is more than 10,000-fold less expressed than v1. Expression levels of all other constructs are reported in Figure S5O. Error bars indicate SEM of at least 3 experiments. Asterisks indicate a significant difference in relative luciferase signal with respect to v1 based on one-way ANOVA statistical tests in GraphPad (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, and **** $p \leq 0.0001$).

(D) Sequences used to disrupt the H11 stem (H11-5' mut, red nucleotides) and corresponding compensatory mutations (H11-comp, green nucleotides).

(E) qRT-PCR and western blot analysis of p53 upregulation by v1, v9, H11LpA, and $G^{370}C$.

(F) qRT-PCR analysis measuring upregulation of p53 target genes (BAX, p21, GADD45A, GDF15, and NOXA) by v1, v9, H11LpA, and $G^{370}C$. Representative western blots for BAX and p21 are reported on the right (the endogenous signal for GADD45A, GDF15, and NOXA is too low in our system for accurate quantification).

(G) RNA immunoprecipitation using the DO1 anti-p53 antibody for cells transfected with the indicated constructs. Control samples using unspecific immunoglobulin G (IgG) produced negligible amplification and are not plotted. Values are reported as percent input. Error bars indicate SEM of 4 biological replicates, each performed in technical duplicates. Asterisks indicate significant variations in the amounts of immunoprecipitated control RNAs (GAPDH and RNR1) with respect to the target RNA (v1, H11LpA, or DINO) based on unpaired parametric t tests in GraphPad (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, and **** $p \leq 0.0001$).

(H) Pull-down of p53 (detected by western blot) using *in vitro*-transcribed and biotinylated v1 and H11LpA and non-biotinylated v1 (CTRL). Values are reported as percent of input. Error bars indicate SEM of 3 biological replicates. Asterisks indicate significant variation of p53 pulled down by v1 and H11LpA with respect to CTRL, based on one-way ANOVA statistical tests in GraphPad (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, and **** $p \leq 0.0001$).

Please cite this article as: Uroda et al., Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway, Molecular Cell (2019), <https://doi.org/10.1016/j.molcel.2019.07.025>

CellPress

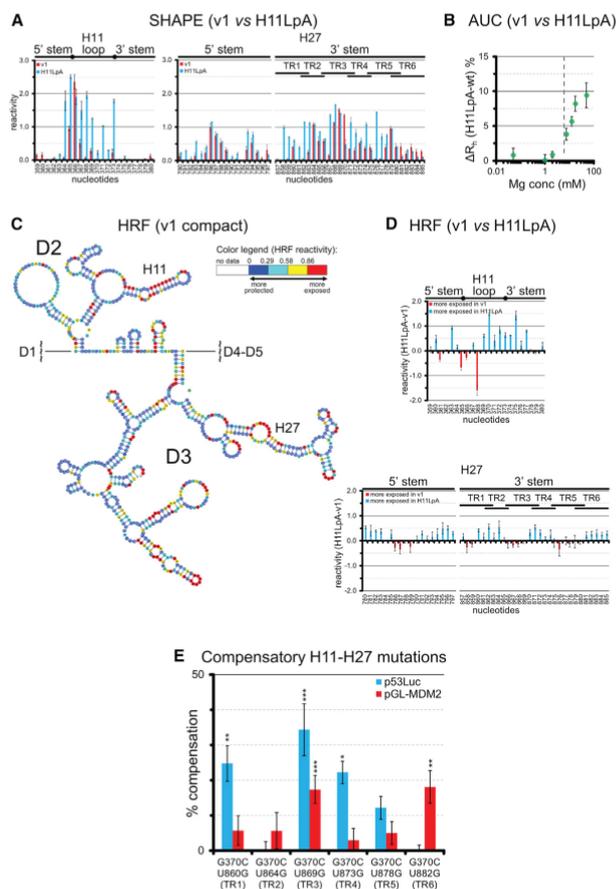


Figure 5. Structural and Functional Interconnections between H11 and H27

(A) SHAPE (1M7) reactivity values of individual nucleotides in H11 and H27 from compact v1 (from Figures S1–S3) and H11LpA.

(B) Difference in hydration radius (ΔR_h) between v1 and H11LpA at increasing Mg^{2+} concentrations, as measured by AUC (the vertical dotted line indicates the $C_{Mg^{1/2}}$ of v1 from Table S1).

(C) Secondary structure map of compact v1 color-coded according to the HRF reactivity values of individual nucleotides. HRF analysis and normalization procedures are described in the STAR Methods.

(D) HRF reactivity values of individual nucleotides in H11 and H27 from compact v1 (from C) and H11LpA in K^+ and Mg^{2+} . Reactivity values of H11LpA were normalized to the reactivity values of v1 following the scaling procedure of QuSHAPE, as described previously (Karabiber et al., 2013). Raw reactivity plots and correlations between replicates are reported in Figure S6.

(E) p53-dependent luciferase assay (p53Luc and pGL-MDM2 reporters) on $G^{370}C$ and compensatory double mutants.

Error bars in (A) and (C)–(E) indicate SEM of 3 experiments. Asterisks indicate significant variation in relative luciferase signal with respect to the $G^{370}C$ single mutant (Figure 4C) based on one-way ANOVA statistical tests in GraphPad (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$).

secondary and tertiary structures of H27 (Figures 5A and 5D). Moreover, *in vivo*, H11 and H27 are both required for stimulation of the p53 pathway (Figure 4B) and form functionally important base-pairing interactions because inactive H11 mutants can be rescued by compensatory H27 mutations (Figure 5E).

The Functional H11–H27 Pseudoknots Are Required for Compaction of MEG3 *In Vitro*

Having established that MEG3 compacts in a specific manner *in vitro* and that it forms a functionally important pseudoknot *in vitro* and *in vivo*, we attempted to visualize this lncRNA by single-particle 3D imaging. For our study, we used atomic force microscopy (AFM), a technique used previously to image other multi-domain structured RNAs and to monitor RNA conformational changes (García-Sacristán et al., 2015; Giro et al., 2004; Hansma et al., 1996; Lyubchenko et al., 2011; Schön, 2016; Yu et al., 2015).

We visualized v1 in three different folding states (Figures 6 and S7A; Table S1). In the presence of formamide, which denatures RNA, v1 forms elongated unstructured filaments, as expected. In the presence of K^+ ions, which induce RNA to form secondary but not tertiary structure motifs, v1 particles become shorter and taller, forming globular domains connected by flexible linkers.

compensatory point mutation within each H27 TR. We tested all six resulting double mutants by luciferase assay using both the p53Luc and the pGL-MDM2 reporter. Strikingly, all double mutants except $G^{370}C/U^{864}G$ (H11/TR2) partially rescued activity on at least one reporter. Interestingly, the extent of rescue depends on the TR concerned: $G^{370}C/U^{869}G$ (H11/TR3) was the most efficient compensatory mutant on both p53Luc and the pGL-MDM2 reporters, $G^{370}C/U^{873}G$ (H11/TR4) rescued activity only on p53Luc but not on pGL-MDM2, and $G^{370}C/U^{882}G$ (H11/TR6) rescued activity only on pGL-MDM2 but not on p53Luc (Figure 5E).

Taken together, our data show that H11 and H27 are structurally connected *in vitro* because mutations on H11 affect the sec-

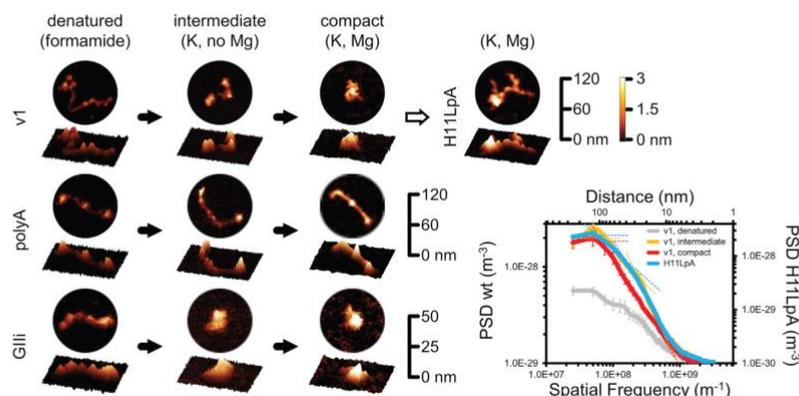


Figure 6. In Vitro Single-Particle Analysis of v1

Shown are representative AFM particles of v1, poly(A) RNA, and group II intron (GII) in formamide (denatured state), K^+ (intermediate state), and K^+ and Mg^{2+} (compact state). A representative particle of H11LpA in K^+ and Mg^{2+} is reported at the top right. Each particle is displayed in 2D and 3D representations. The xy scale bars are on the right for each molecule, and the z color scale bar is common to all samples. The bottom right panel shows PSD plots from images acquired at $1,024 \times 1,024 \text{ pixel}^2$ with a pixel size of 0.98 nm/pixel for v1 in the denatured, intermediate, and compact states and of H11LpA in K^+ and Mg^{2+} . We obtained similar PSD plots for images acquired at $512 \times 512 \text{ pixel}^2$ with a pixel size of 1.96 nm/pixel . Intercepts between linear fits to autofluorescence regions in the spectra (dashed lines) indicate characteristic spatial frequencies. Error bars indicate SEM. We imaged ~ 100 – 110 particles in total per condition (see STAR Methods for details). The AFM processing pipeline and raw scan are reported in Figure S7.

Power spectral density (PSD) analysis, which provides an overview of the spatial features present in images (Calò et al., 2009; Higuchi, 1988), reveals two characteristic peaks. One peak likely corresponds to the globular domains of v1 (average size of 30 nm) and the second peak to the entire v1 ($\sim 85 \text{ nm}$), in line with dimensions measured in solution by small-angle X-ray scattering (SAXS; maximum particle size $[D_{\text{max}}]$, $\sim 70 \text{ nm}$; Figure S7B; Table S1). Finally, in the presence of K^+ and Mg^{2+} , which induce RNA tertiary folding (Su et al., 2005; Wadley et al., 2007; Woodson, 2005, 2010), v1 particles become even shorter and taller than with K^+ and lose their multi-domain organization to adopt an overall compact shape. Correspondingly, the PSD spectrum becomes steeper and loses the 30-nm shoulder. The intercept with the low-frequency plateau is now at $\sim 65 \text{ nm}$, suggesting a degree of compaction ($\sim 25\%$) similar to that observed in solution by AUC ($\sim 20\%$; Figure S7C; Table S1). The folding behavior observed for v1 differs from that of low-complexity RNAs, such as poly(A) RNA homopolymers, and is similar to that of highly structured RNAs, like the *Oceanobacillus iheyensis* group II intron (Figure 6). Furthermore, we also imaged the H11LpA mutant in K^+ and Mg^{2+} . This mutant displays a dramatic defect in compaction. PSD analysis shows emergence of an inflection point at $\sim 40 \text{ nm}$ in the mutant, more closely resembling the intermediate than the compact state of v1 (Figures 6 and S7A).

In summary, our AFM analysis, supported by our highly correlated biochemical, biophysical, and functional characterization *in vitro* and *in vivo*, reveals that MEG3 folding is dictated by the formation of evolutionarily conserved and functionally important pseudoknots between H11 and H27.

DISCUSSION

Alternative Splicing Dictates Structural Organization in IncRNA MEG3

In this work, we characterized the secondary and tertiary structures of MEG3, a human lncRNA that functions as a tumor suppressor by stimulating the p53 pathway.

Our secondary structure maps of three MEG3 splice variants show that MEG3 forms highly structured domains, displaying Shannon entropy and SHAPE reactivity values comparable with highly structured RNAs (Mathews, 2004; Figures 1G, S1–S4, and S6C). The boundaries between structural domains define exon E3 as an independent module comprising domains D2 and D3 (Figures S1–S4). Correspondence between exon junctions and domain boundaries have been observed previously for another lncRNA, BRAVEHEART (Xue et al., 2016), but not for HOTAIR (Somarowthu et al., 2015) or XIST (Smola et al., 2016). For MEG3, which possesses at least 27 splice variants exhibiting different levels of p53 stimulation, such correspondence is remarkable and provides support for the previously proposed correlation between MEG3 exonic organization, structural architecture, and ability to stimulate the p53 pathway (Zhang et al., 2010b). D2 and D3 likely adopt a similar secondary structure across human MEG3 splicing variants (Figure S1) because they have similar chemical reactivity in different cell lines (endogenous and transfected; Figure 1F) and a distinct solvent protection pattern *in vitro* (Figure 5C). Importantly, we could identify D2–D3 (E3) in early-diverging mammals such as Marsupialia (i.e., Tasmanian devil; Data S1; Table S2),

Reference List

- Agarwal SK & Jothi R (2012) Genome-Wide Characterization of Menin-Dependent H3K4me3 Reveals a Specific Role for Menin in the Regulation of Genes Implicated in MEN1-Like Tumors. *PLoS One* **7**: e37952
- Andergassen D, Muckenhuber M, Bammer PC, Kulinski TM, Theussl H-C, Shimizu T, Penninger JM, Pauler FM & Hudson QJ (2019) The Airn lncRNA does not require any DNA elements within its locus to silence distant imprinted genes. *PLoS Genet.* **15**: e1008268
- Antonarakis SE, Blouin JL, Maher J, Avramopoulos D, Thomas G, Talbot CC & Jr (1993) Maternal uniparental disomy for human chromosome 14, due to loss of a chromosome 14 from somatic cells with t(13;14) trisomy 14. *Am. J. Hum. Genet.* **52**: 1145–52
- Arun G, Diermeier S, Akerman M, Chang KC, Wilkinson JE, Hearn S, Kim Y, MacLeod AR, Krainer AR, Norton L, Brogi E, Egeblad M & Spector DL (2016) Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev.* **30**: 34–51
- Arun G, Diermeier SD & Spector DL (2018) Therapeutic Targeting of Long Non-Coding RNAs in Cancer. *Trends Mol. Med.* **24**: 257–277
- Augui S, Nora EP & Heard E (2011) Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat. Rev. Genet.* **2011** **12**: 429–442
- Avner P & Heard E (2001) X-chromosome inactivation: Counting, choice and initiation. *Nat. Rev. Genet.* **2**: 59–67
- Ayed A, Mulder FAA, Yi G-S, Lu Y, Kay LE & Arrowsmith CH (2001) Latent and active p53 are identical in conformation. *Nat. Struct. Biol.* **8**: 756–760
- Bae E, Calhoun VC, Levine M, Lewis EB & Drewell RA (2002) Characterization of the intergenic RNA profile at abdominal-A and abdominal-B in the *Drosophila* bithorax complex. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 16847–16852
- Barnwal RP, Yang F & Varani G (2017) Applications of NMR to structure determination of RNAs large and small. *Arch. Biochem. Biophys.* **628**: 42–56
- Barra J & Leucci E (2017) Probing long non-coding RNA-protein interactions. *Front. Mol. Biosci.* **4**: 45
- Bartel DP (2009) MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**: 215–233
- Bartel DP (2018) Metazoan MicroRNAs. *Cell* **173**: 20–51
- Bauer NC, Yang A, Wang X, Zhou Y, Klibanski A & Soberman RJ (2021) A cross-nearest neighbor/Monte Carlo algorithm for single-molecule localization microscopy defines interactions between p53, Mdm2, and MEG3. *J. Biol. Chem.* **0**: 100540
- Beltran M, Puig I, Peña C, García JM, Álvarez AB, Peña R, Bonilla F & De Herreros AG (2008) A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev.* **22**: 756–769
- Berends MJW, Hordijk R, Scheffer H, Oosterwijk JC, Halley DJJ & Sorgedraeger N (1999) Two cases of maternal uniparental disomy 14 with a phenotype overlapping with the Prader-Willi phenotype. *Am. J. Med. Genet.* **84**: 76–79
- Berghoff EG, Clark MF, Chen S, Cajigas I, Leib DE & Kohtz JD (2013) Evf2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. *Development* **140**: 4407–16
- Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ, Gingeras TR, Schreiber SL & Lander ES (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**: 169–181
- Bitetti A, Mallory A, Golini E, Carrieri C, Carreño Gutiérrez H, Perlas E, Pérez-Rico Y, Tocchini-Valentini G, Enright A, WHJ N, Mandillo S, O'Carroll D & Shkumatava A (2018) MicroRNA degradation by a conserved target RNA regulates animal behavior. *Nat. Struct. Mol. Biol.* **25**: 244–251
- Bond AM, Vangompel MJW, Sametsky EA, Clark MF, Savage JC, Disterhoft JF & Kohtz JD (2009) Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat. Neurosci.* **12**: 1020–1027
- Borah P, Das A, Milner MJ, Ali A, Bentley AR & Pandey R (2018) Long non-coding RNAs as endogenous target mimics and exploration of their role in low nutrient stress tolerance in plants. *Genes (Basel)*. **9**: 1. – 17
- Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D, Capra V, Grompe M, Pizzuti A, Muzny D, Lawrence C, Willard HF, Avner P & Ballabio A (1991) Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**: 325–329
- Botti G, Marra L, Malzone M, Anniciello A, Botti C, Franco R & Cantile M (2016) LncRNA HOTAIR as Prognostic Circulating Marker and Potential Therapeutic Target in Patients with Tumor Diseases. *Curr. Drug Targets* **18**: 27–34
- Bourdon JC, Fernandes K, Murray-Zmijewski F, Liu G, Diot A, Xirodimas DP, Saville MK & Lane DP (2005) p53 isoforms can regulate p53 transcriptional activity. *Genes Dev.* **19**: 2122–2137
- Braconi C, Kogure T, Valeri N, Huang N, Nuovo G, Costinean S, Negrini M, Miotto E, Croce CM & Patel T (2011) MicroRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. *Oncogene* **30**: 4750–4756

- Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R & Willard HF (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**: 38–44
- Cabili M, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A & Rinn JL (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**: 1915–1927
- Cai X & Cullen BR (2007) The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA* **13**: 313
- Cajigas I, Chakraborty A, Swyter KR, Luo H, Bastidas M, Nigro M, Morris ER, Chen S, VanGompel MJW, Leib D, Kohtz SJ, Martina M, Koh S, Ay F & Kohtz JD (2018) The Efv2 Ultraconserved Enhancer lncRNA Functionally and Spatially Organizes Megabase Distant Genes in the Developing Forebrain. *Mol. Cell* **71**: 956–972.e9
- Cajigas I, Leib DE, Cochrane J, Luo H, Swyter KR, Chen S, Clark BS, Thompson J, Yates JR, Kingston RE & Kohtz JD (2015) Efv2 lncRNA/BRG1/DLX1 interactions reveal RNA-dependent inhibition of chromatin remodeling. *Dev.* **142**: 2641–2652
- Carlevaro-Fita J, Lanzós A, Feuerbach L, Hong C, Mas-Ponte D, Pedersen JS, Abascal F, Amin SB, Bader GD, Barenboim J, Beroukhim R, Bertl J, Boroevich KA, Brunak S, Campbell PJ, Carlevaro-Fita J, Chakravarty D, Chan CWY, Chen K, Choi JK, et al (2020) Cancer lncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun. Biol.* **3**: 1–16
- Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J & Hentze MW (2012) Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**: 1393–1406
- Cazalla D, Yario T & Steitz J (2010) Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA. *Science* **328**: 1563–1566
- Chance MR, Sclavi B, Woodson SA & Brenowitz M (1997) Examining the conformational dynamics of macromolecule with time-resolved synchrotron X-ray 'footprinting'. *Structure* **5**: 865–869
- Chillemi G, Kehrlöesser S, Bernassola F, Desideri A, Dötsch V, Levine AJ & Melino G (2017) Structural Evolution and Dynamics of the p53 Proteins. *Cold Spring Harb. Perspect. Med.* **7**: a028308
- Chillón I, Marcia M, Legiewicz M, Liu F, Somarowthu S & Pyle AM (2015) Native Purification and Analysis of Long RNAs. *Methods Enzymol.* **558**: 3–37
- Christy TW, Giannetti CA, Houlihan G, Smola MJ, Rice GM, Wang J, Dokholyan N V., Laederach A, Holliger P & Weeks KM (2021) Direct Mapping of Higher-Order RNA Interactions by SHAPE-JuMP. *Biochemistry: acs.biochem.1c00270*
- Chunharojrith P, Nakayama Y, Jiang X, Kery RE, Ma J, De La Hoz Ulloa CS, Zhang X, Zhou Y & Klibanski A (2015) Tumor suppression by MEG3 lncRNA in a human pituitary tumor derived cell line. *Mol. Cell. Endocrinol.* **416**: 27–35
- Clore GM, Ernst J, Clubb R, Omichinski JG, Kennedy WMP, Sakaguchi K, Appella E & Gronenborn AM (1995) Refined solution structure of the oligomerization domain of the tumour suppressor p53. *Nat. Struct. Mol. Biol.* **2**: 321–333
- Comijn J, Bex G, Vermassen P, Verschuere K, Van Grunsven L, Bruyneel E, Mareel M, Huylebroeck D & Van Roy F (2001) The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Mol. Cell* **7**: 1267–1278
- Cotter PD, Kaffe S, McCurdy LD, Jhaveri M, Willner JP & Hirschhorn K (1997) Paternal uniparental disomy for chromosome 14: A case report and review. *Am. J. Med. Genet.* **70**: 74–79
- Coviello DA, Panucci E, Mantero MM, Perfumo C, Guelfi M, Borrone C & Bricarelli FD (1996) Maternal Uniparental Disomy for Chromosome 14. *Acta Genet. medicae Gemellol. twin Res.* **45**: 169–172
- Crick F (1970) Central Dogma of Molecular Biology. *Nature* **227**: 561–563
- Darty K, Denise A & Ponty Y (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**: 1974–1975
- Davidovich C, Zheng L, Goodrich KJ & Cech TR (2013) Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* **20**: 1250–1257
- Dey B, Pfeifer K & Dutta A (2014) The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration. *Genes Dev.* **28**: 491–501
- Dimitrova N, Zamudio JR, Jong RM, Soukup D, Resnick R, Sarma K, Ward AJ, Raj A, Lee JT, Sharp PA & Jacks T (2014) lincRNA-p21 Activates p21 In cis to Promote Polycomb Target Gene Expression and to Enforce the G1/S Checkpoint. *Mol. Cell* **54**: 777–790
- Eddy SR (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**: 919–929
- Ei-Deiry WS, Kern SE, Pietenpol JA, Kinzler KW & Vogelstein B (1992) Definition of a consensus binding site for p53. *Nat. Genet.* **1**: 45–49
- ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74
- ENCODE Project Consortium TEP (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**: 636–40

- Feng C, Chan D, Joseph J, Muuronen M, Coldren WH, Dai N, Corrêa IR, Furche F, Hadad CM & Spitale RC (2018) Light-activated chemical probing of nucleobase solvent accessibility inside cells. *Nat. Chem. Biol.* **14**: 276–283
- Feng J, Bi C, Clark BS, Mady R, Shah P & Kohtz JD (2006) The Evi-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* **20**: 1470–1484
- Fischer M (2019) Conservation and divergence of the p53 gene regulatory network between mice and humans. *Oncogene* **38**: 4095–4109
- Fleming-Waddell JN, Olbricht GR, Taxis TM, White JD, Vuocolo T, Craig BA, Tellam RL, Neary MK, Cockett NE & Bidwell CA (2009) Effect of DLK1 and RTL1 but not MEG3 or MEG8 on muscle gene expression in callipyge lambs. *PLoS One* **4**: e7399
- Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, García JA & Paz-Ares J (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* **39**: 1033–1037
- Ghini F, Rubolino C, Climent M, Simeone I, Marzi MJ & Nicassio F (2018) Endogenous transcripts control miRNA levels and activity in mammalian cells by target-directed miRNA degradation. *Nat. Commun.* **2018** **9**: 1–15
- Griffiths-Jones S, Bateman A, Marshall M, Khanna A & Eddy SR (2003) Rfam: an RNA family database. *Nucleic Acids Res.* **31**: 439–41
- Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, Van De Vijver MJ, Sukumar S & Chang HY (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**: 1071–1076
- Gutschner T, Hämmerle M, Eißmann M, Hsu J, Kim Y, Hung G, Revenko A, Arun G, Stentrup M, Groß M, Zörnig M, MacLeod AR, Spector DL & Diederichs S (2013) The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* **73**: 1180–1189
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES, et al (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**: 223–227
- Ha M & Kim VN (2014) Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**: 509–524
- Haerty W & Ponting CP (2013) Mutations within lncRNAs are effectively selected against in fruitfly but not in human. *Genome Biol.* **14**: 1–16
- Hampel KJ & Burke JM (2001) Time-resolved hydroxyl-radical footprinting of RNA using Fe(II)-EDTA. *Methods* **23**: 233–239
- Han S, Zhao BS, Myers SA, Carr SA, He C & Ting AY (2020) RNA-protein interaction mapping via MS2- or Cas13-based APEX targeting. *Proc. Natl. Acad. Sci. U. S. A.* **117**: 22068–22079
- Hannon GJ (2002) RNA interference. *Nature* **418**: 244–251
- Harris DA, Tinsley RA & Walter NG (2004) Terbium-mediated Footprinting Probes a Catalytic Conformational Switch in the Antigenomic Hepatitis Delta Virus Ribozyme. *J. Mol. Biol.* **341**: 389–403
- Harris DA, Todd GC & Walter NG (2014) Terbium(III) Footprinting as a Probe of RNA Structure and Metal Binding Sites. In *Handbook of RNA Biochemistry: Second, Completely Revised and Enlarged Edition* pp 255–268. Wiley Blackwell
- Healey S, Powell F, Battersby M, Chenevix-Trench G & McGill J (1994) Distinct phenotype in maternal uniparental disomy of chromosome 14. *Am. J. Med. Genet.* **51**: 147–149
- Hellen CUT & Sarnow P (2001) Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.* **15**: 1593–1612
- Hezroni H, Koppstein D, Schwartz MGG, Avrutin A, Bartel DPP & Ulitsky I (2015) Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Rep.* **11**: 1110–1122
- Higgs PG & Lehman N (2015) The RNA World: Molecular cooperation at the origins of life. *Nat. Rev. Genet.* **16**: 7–17
- Hoagland MB, Stephenson ML, Scott JF, Hecht LI & Zamecnik PC (1958) A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.* **231**: 241–57
- Hordijk R, Wierenga H, Scheffer H, Leegte B, Hofstra RM & Stolte-Dijkstra I (1999) Maternal uniparental disomy for chromosome 14 in a boy with a normal karyotype. *J. Med. Genet.* **36**: 782–5
- Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T & Rinn JL (2010) A large intergenic non-coding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**: 409–419
- Hulscher RM, Bohon J, Rappé MC, Gupta S, D'Mello R, Sullivan M, Ralston CY, Chance MR & Woodson SA (2016) Probing the structure of ribosome assembly intermediates in vivo using DMS and hydroxyl radical footprinting. *Methods* **103**: 49–56
- Iyer M, Niknafs Y, Malik R, Singhal U, Sahu A, Hosono Y, Barrette T, Prensner J, Evans J, Zhao S,

- Poliakov A, Cao X, Dhanasekaran S, Wu Y, Robinson D, Beer D, Feng F, Iyer H & Chinnaiyan A (2015) The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**: 199–208
- Iyer S, Modali SD & Agarwal SK (2017) Long Noncoding RNA MEG3 Is an Epigenetic Determinant of Oncogenic Signaling in Functional Pancreatic Neuroendocrine Tumor Cells. *Mol. Cell. Biol.* **37**: e00278-17
- De Jesus BB, Marinho SP, Arros S, Sousa-Franco A, Alves-Vale C, Carvalho T & Carmo-Fonseca M (2018) Silencing of the lncRNA Zeb2-NAT facilitates reprogramming of aged fibroblasts and safeguards stem cell pluripotency. *Nat. Commun.* **9**: 1–11
- Ji P, Diederichs S, Wang W, Böing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, Thomas M, Berdel WE, Serve H & Müller-Tidow C (2003) MALAT-1, a novel noncoding RNA, and thymosin β 4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **22**: 8031–8041
- Joerger AC & Fersht AR (2008) Structural Biology of the Tumor Suppressor p53. *Annu. Rev. Biochem.* **77**: 557–582
- Johansson HE, Liljas L & Uhlenbeck OC (1997) RNA recognition by the MS2 phage coat protein. *Semin. Virol.* **8**: 176–185
- Jonas S & Izaurralde E (2015) Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16**: 421–433
- Kagami M, Mizuno S, Matsubara K, Nakabayashi K, Sano S, Fuke T, Fukami M & Ogata T (2015) Epimutations of the IG-DMR and the MEG3-DMR at the 14q32.2 imprinted region in two patients with Silver-Russell Syndrome-compatible phenotype. *Eur. J. Hum. Genet.* **23**: 1062–7
- Kagami M, Sekita Y, Nishimura G, Irie M, Kato F, Okada M, Yamamori S, Kishimoto H, Nakayama M, Tanaka Y, Matsuoka K, Takahashi T, Noguchi M, Tanaka Y, Masumoto K, Utsunomiya T, Kouzan H, Komatsu Y, Ohashi H, Kurosawa K, et al (2008) Deletions and epimutations affecting the human 14q32.2 imprinted region in individuals with paternal and maternal upd(14)-like phenotypes. *Nat. Genet.* **40**: 237–242
- Kaneko S, Bonasio R, Saldaña-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A & Reinberg D (2014) Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin. *Mol. Cell* **53**: 290–300
- Kapranov P, Cheng J, Dike S, Nix DA, Dutttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, et al (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science (80-.)*. **316**: 1484–1488
- Karabiber F, McGinnis JL, Favorov O V. & Weeks KM (2013) QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA* **19**: 63–73
- Karner H, Webb CH, Carmona S, Liu Y, Lin B, Erhard M, Chan D, Baldi P, Spitale RC & Sun S (2020) Functional Conservation of LncRNA JPX Despite Sequence and Structural Divergence. *J. Mol. Biol.* **432**: 283–300
- Kedde M, van Kouwenhove M, Zwart W, Oude Vrielink JAF, Elkon R & Agami R (2010) A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. *Nat. Cell Biol.* **12**: 1014–1020
- Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, Smits G & Reik W (2012) The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat. Cell Biol.* **2012** **14**: 659–665
- Kent WJ (2002) BLAT---The BLAST-Like Alignment Tool. *Genome Res.* **12**: 656–664
- Kern SE, Kinzler KW, Bruskin a, Jarosz D, Friedman P, Prives C & Vogelstein B (1991) Identification of p53 as a sequence-specific DNA-binding protein. *Science* **252**: 1708–1711
- Kirk JM, Kim SO, Inoue K, Smola MJ, Lee DM, Schertzer MD, Wooten JS, Baker AR, Sprague D, Collins DW, Horning CR, Wang S, Chen Q, Weeks KM, Mucha PJ & Mauro Calabrese J (2018) Functional classification of long non-coding RNAs by k-mer content. *Nat. Genet.*
- Kiss T (2002) Small Nucleolar RNAs: An Abundant Group of Noncoding RNAs with Diverse Cellular Functions. *Cell* **109**: 145–148
- Kitayner M, Rozenberg H, Kessler N, Rabinovich D, Shaulov L, Haran TE & Shakked Z (2006) Structural Basis of DNA Recognition by p53 Tetramers. *Mol. Cell* **22**: 741–753
- Kleaveland B, Shi C, Stefano J & Bartel D (2018) A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. *Cell* **174**: 350-362.e17
- Kmita M & Duboule D (2003) Organizing axes in time and space; 25 years of colinear tinkering. *Science (80-.)*. **301**: 331–333
- Kota S, Lières D, Bouschet T, Hirasawa R, Marchand A, Begon-Pescia C, Sanli I, Arnaud P, Journot L, Girardot M & Feil R (2014) ICR noncoding RNA expression controls imprinting and DNA replication at the Dlk1-Dio3 domain. *Dev. Cell* **31**: 19–33
- Latos PA, Pauler FM, Koerner M V., Şenergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH,

- Klement RM, Warczok KE, Aumayr K, Pasierbek P & Barlow DP (2012) Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science (80-.)*. **338**: 1469–1472
- Lee S, Kopp F, Chang TC, Sataluri A, Chen B, Sivakumar S, Yu H, Xie Y & Mendell JT (2016) Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* **164**: 69–80
- Li MZ & Elledge SJ (2012) SLIC: A method for sequence- and ligation-independent cloning. *Methods Mol. Biol.* **852**: 51–59
- Li XL, Subramanian M, Jones MF, Chaudhary R, Singh DK, Zong X, Gryder B, Sindri S, Mo M, Schetter A, Wen X, Parvathaneni S, Kazandjian D, Jenkins LM, Tang W, Elloumi F, Martindale JL, Huarte M, Zhu Y, Robles AI, et al (2017) Long Noncoding RNA PURPL Suppresses Basal p53 Levels and Promotes Tumorigenicity in Colorectal Cancer. *Cell Rep.* **20**: 2408–2423
- Li Z, Li C, Liu C, Yu S & Zhang Y (2015) Expression of the long non-coding RNAs MEG3, HOTAIR, and MALAT-1 in non-functioning pituitary adenomas and their relationship to tumor behavior. *Pituitary* **18**: 42–47
- Libri V, Helwak A, Miesen P, Santhakumar D, Borger JG, Kudla G, Grey F, Tollervey D & Buck AH (2012) Murine cytomegalovirus encodes a miR-27 inhibitor disguised as a target. *Proc. Natl. Acad. Sci.* **109**: 279–284
- Lion M, Raimondi I, Donati S, Jousson O, Ciribilli Y & Inga A (2015) Evolution of p53 transactivation specificity through the lens of a yeast-based functional assay. *PLoS One* **10**: 116177
- Liu F, Somarowthu S & Pyle AM (2017) Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat. Chem. Biol.* **13**: 282–289
- Liu S, Zhu J, Jiang T, Zhong Y, Tie Y, Wu Y, Zheng X, Jin Y & Fu H (2015) Identification of lncRNA MEG3 Binding Protein Using MS2-Tagged RNA Affinity Purification and Mass Spectrometry. *Appl. Biochem. Biotechnol.* **176**: 1834–1845
- Luo Z, Zhang QC, Lee B, Flynn RA, Smith MA, Robinson JT, Davidovich C, Gooding AR, Goodrich KJ, Mattick JS, Mesirov JP, Cech TR & Chang HY (2016) RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**: 1267–1279
- Luo Z, Lin C, Woodfin AR, Bartom ET, Gao X, Smith ER & Shilatifard A (2016) Regulation of the imprinted DLK1-Dio3 locus by allele-specific enhancer activity. *Genes Dev.* **30**: 92–101
- Ma B & Levine AJ (2007) Probing potential binding modes of the p53 tetramer to DNA based on the symmetries encoded in p53 response elements. *Nucleic Acids Res.* **35**: 7733–47
- MacMillan AM (2001) From equilibrium to kinetic footprinting of RNA structure. *J. Chem. Soc. Perkin Trans. 2*: 1263–1267
- Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD & Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**: W636–W641
- Marcinowski L, Tanguy M, Krmpotic A, Rädle B, Lisnić VJ, Tuddenham L, Chane-Woon-Ming B, Ruzsics Z, Erhard F, Benkartek C, Babic M, Zimmer R, Trgovcich J, Koszinowski UH, Jonjic S, Pfeffer S & Dölken L (2012) Degradation of Cellular miR-27 by a Novel, Highly Abundant Viral Transcript Is Important for Efficient Virus Replication In Vivo. *PLoS Pathog.* **8**: e1002510
- Marques AC & Ponting CP (2009) Catalogues of mammalian long noncoding RNAs: Modest conservation and incompleteness. *Genome Biol.* **10**: R124
- Martin RA, Sabol DW & Rogan PK (1999) Maternal uniparental disomy of chromosome 14 confined to an interstitial segment (14q23-14q24.2). *J. Med. Genet.* **36**: 633–6
- Mercer TR, Dinger ME & Mattick JS (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10**: 155–159
- Miyoshi N, Wagatsuma H, Wakana S, Shiroishi T, Nomura M, Aisaka K, Kohda T, Surani MA, Kaneko-Ishino T & Ishino F (2000) Identification of an imprinted gene, Meg3/Gtl2 and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q. *Genes to Cells* **5**: 211–220
- Mo C-F, Wu F-C, Tai K-Y, Chang W-C, Chang K-W, Kuo H-C, Ho H-N, Chen H-F & Lin S-P (2015) Loss of non-coding RNA expression from the DLK1-DIO3 imprinted locus correlates with reduced neural differentiation potential in human embryonic stem cell lines. *Stem Cell Res. Ther.* **6**: 1
- Modali SD, Parekh VI, Kebebew E & Agarwal SK (2015) Epigenetic Regulation of the lncRNA MEG3 and Its Target c-MET in Pancreatic Neuroendocrine Tumors. *Mol. Endocrinol.* **29**: 224–237
- Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, Mitra S, Mohammed A, James AR, Hoberg E, Moustakas A, Gyllenstein U, Jones SJM, Gustafsson CM, Sims AH, Westerlund F, Gorab E & Kanduri C (2015) MEG3 long noncoding RNA regulates the TGF- β pathway genes through formation of RNA–DNA triplex structures. *Nat. Commun.* **6**: 7743
- Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R & Fraser P (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science (80-.)*. **322**: 1717–1720
- Nan X, Campoy FJ & Bird A (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88**: 471–481
- Nawrocki EP, Kolbe DL & Eddy SR (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**:

- 1335–1337
- Nelles L, Van De Putte T, Van Grunsven L, Huylebroeck D & Verschuere K (2003) Organization of the mouse *Zfhx1b* gene encoding the two-handed zinc finger repressor Smad-interacting protein-1. *Genomics* **82**: 460–469
- Nieselt K & Herbig A (2013) Non-coding RNA, Classification. In *Encyclopedia of Systems Biology*, Dubitzky W, Wolkenhauer O, Cho K-H & Yokota H (eds) pp 1532–1534. New York, NY: Springer New York
- Okamura K & Lai EC (2008) Endogenous small interfering RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **2008** **9**: 673–678
- Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, Yamanaka I, Kiyosawa H, Yagi K, Tomaru Y, Hasegawa Y, Nogami A, Schönbach C, Gojobori T, Baldarelli R, Hill DP, et al (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563–573
- Ozata DM, Gainetdinov I, Zoch A, O'Carroll D & Zamore PD (2018) PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* **2018** **20**: 89–108
- Penny GD, Kay GF, Sheardown SA, Rastan S & Brockdorff N (1996) Requirement for Xist in X chromosome inactivation. *Nature* **379**: 131–137
- Petty TJ, Emamzadah S, Costantino L, Petkova I, Stavridi ES, Saven JG, Vauthey E & Halazonetis TD (2011) An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity. *EMBO J.* **30**: 2167–2176
- Pyle A (2002) Metal ions in the structure and function of RNA. *JBIC J. Biol. Inorg. Chem.* **7**: 679–690
- Pyle AM, Sigel RKO & Vaidya A (2000) Metal ion binding sites in a group II intron core. *Nat. Struct. Biol.* **7**: 1111–1116
- Quinn JJ & Chang HY (2016) Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* **17**: 47–62
- Ramakrishnan V (2002) Ribosome structure and the mechanism of translation. *Cell* **108**: 557–572
- Rice GM, Leonard CW & Weeks KM (2014) RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA* **20**: 846–854
- Riley KJ-L, Cassidy LA, Kumar A & Maher LJ (2006) Recognition of RNA by the p53 tumor suppressor protein in the yeast three-hybrid system. *RNA* **12**: 620–30
- Riley KJ & Maher JL (2007a) Analysis of p53-RNA interactions in cultured human cells. *Biochem. Biophys. Res. Commun.* **363**: 381–7
- Riley KJ & Maher LJ (2007b) p53 RNA interactions: new clues in an old mystery. *RNA* **13**: 1825–33
- Riley KJ, Ramirez-Alvarado M & James Maher L (2007) RNA-p53 Interactions in Vitro. *Biochemistry*: 2480–2487
- Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E & Chang HY (2007) Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* **129**: 1311–1323
- Da Rocha ST, Edwards CA, Ito M, Ogata T & Ferguson-Smith AC (2008) Genomic imprinting at the mammalian Dlk1-Dio3 domain. *Trends Genet.* **24**: 306–316
- Saito H & Suga H (2002) Outersphere and innersphere coordinated metal ions in an aminoacyl-tRNA synthetase ribozyme. *Nucleic Acids Res.* **30**: 5151–5159
- Sanli I, Lalevée S, Cammisa M, Perrin A, Rage F, Llères D, Riccio A, Bertrand E & Feil R (2018) Meg3 Non-coding RNA Expression Controls Imprinting by Preventing Transcriptional Upregulation in cis. *Cell Rep.* **23**: 337–348
- Sato S, Yoshida W, Soejima H, Nakabayashi K & Hata K (2011) Methylation dynamics of IG-DMR and Gtl2-DMR during murine embryonic and placental development. *Genomics* **98**: 120–127
- Schmitt AM & Chang HY (2016) Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* **29**: 452–463
- Schmitt AM, Garcia JT, Hung T, Flynn RA, Shen Y, Qu K, Payumo AY, Peres-Da-Silva A, Broz DK, Baum R, Guo S, Chen JK, Attardi LD & Chang HY (2016) An inducible long noncoding RNA amplifies DNA damage signaling HHS Public Access lncRNA can create a feedback loop with its cognate transcription factor to amplify cellular signaling networks. *Nat Genet* **48**: 1370–1376
- Schorderet P & Duboule D (2011) Structural and Functional Differences in the Long Non-Coding RNA *Hotair* in Mouse and Human. *PLoS Genet.* **7**: e1002071
- Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**: 1606–1619
- Schüler A, Ghanbarian AT & Hurst LD (2014) Purifying selection on splice-related motifs, not expression level nor RNA folding, explains nearly all constraint on human lincRNAs. *Mol. Biol. Evol.* **31**: 3164–83
- Schuster-Gossler K, Bilinski P, Sado T, Ferguson-Smith A & Gossler A (1998) The mouse *Gtl2* gene is differentially expressed during embryonic development, encodes multiple alternatively spliced transcripts, and may act as an RNA. *Dev. Dyn.* **212**: 214–228
- Schuster-Gossler K, Simon-Chazottes D, Guénet JL, Zachgo J & Gossler A (1996) *Gtl2lacZ*, an insertional mutation on mouse Chromosome 12 with parental origin-dependent phenotype. *Mamm. Genome* **7**:

- 20–24
- Sherpa C, Rausch JW & Le Grice SF (2018) Structural characterization of maternally expressed gene 3 RNA reveals conserved motifs and potential sites of interaction with polycomb repressive complex 2. *Nucleic Acids Res.* **46**: 10432–10447
- Siegfried NA, Busan S, Rice GM, Nelson JAE & Weeks KM (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**: 959–965
- Simon MD, Wang CI, Kharchenko P V., West JA, Chapman BA, Alekseyenko AA, Borowsky ML, Kuroda MI & Kingston RE (2011) The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. U. S. A.* **108**: 20497–20502
- Slobodin B & Gerst JE (2010) A novel mRNA affinity purification technique for the identification of interacting proteins and transcripts in ribonucleoprotein complexes. *RNA* **16**: 2277–2290
- Smola MJ, Calabrese JM & Weeks KM (2015) Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry* **54**: 6867–6875
- Somarowthu S, Legiewicz M, Chillón I, Marcia M, Liu F & Pyle AM (2015) HOTAIR Forms an Intricate and Modular Secondary Structure. *Mol. Cell* **58**: 353–361
- Sommerkamp P, Renders S, Ladel L, Hotz-Wagenblatt A, Schönberger K, Zeisberger P, Przybylla A, Sohn M, Zhou Y, Klibanski A, Cabezas-Wallscheid N & Trumpp A (2019) The long non-coding RNA Meg3 is dispensable for hematopoietic stem cells. *Sci. Rep.* **9**: 1–11
- Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, Kuchelmeister HY, Batista PJ, Torre EA, Kool ET & Chang HY (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**: 486–490
- Stark H (2010) GraFix: Stabilization of fragile macromolecular complexes for single particle Cryo-EM. In *Methods in Enzymology* pp 109–126. Academic Press Inc.
- Sullivan KD, Galbraith MD, Andrysiak Z & Espinosa JM (2018) Mechanisms of transcriptional regulation by p53. *Cell Death Differ.* **25**: 133–143
- Temple IK, Cockwell A, Hassold T, Pettay D & Jacobs P (1991) Maternal uniparental disomy for chromosome 14. *J. Med. Genet.* **28**: 511–4
- Terashima M, Tange S, Ishimura A & Suzuki T (2017) MEG3 Long Noncoding RNA Contributes to the Epigenetic Regulation of Epithelial-Mesenchymal Transition in Lung Cancer Cell Lines. *J. Biol. Chem.* **292**: 82–99
- Tichon A, Gil N, Lubelsky Y, Solomon TH, 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. *Nat. Commun.* **7**:
- Tichon A, Perry RBT, Stojic L & Ulitsky I (2018) SAM68 is required for regulation of pumilio by the NORAD long noncoding RNA. *Genes Dev.* **32**: 70–78
- Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E & Chang HY (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science (80-.).* **329**: 689–693
- Tullius TD & Dombroski BA (1986) Hydroxyl radical ‘footprinting’: high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. *Proc. Natl. Acad. Sci.* **83**: 5469–5473
- Ule J, Jensen KB, Ruggiu M, Mele A, Ule A & Darnell RB (2003) CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**: 1212–5
- Ulitsky I (2016) Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* **17**: 601–614
- Ulitsky I (2018) Interactions between short and long noncoding RNAs. *FEBS Lett.* **592**: 2874–2883
- 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. *Mol. Cell*
- Uroda T, Chillón I, Annibale P, Teulon JM, Pessey O, Karuppasamy M, Pellequer JL & Marcia M (2020) Visualizing the functional 3D shape and topography of long noncoding RNAs by single-particle atomic force microscopy and in-solution hydrodynamic techniques. *Nat. Protoc.* **15**: 2107–2139
- Vandewalle C, Comijn J, De Craene B, Vermassen P, Bruyneel E, Andersen H, Tulchinsky E, Van Roy F & Berx G (2005) SIP1/ZEB2 induces EMT by repressing genes of different epithelial cell-cell junctions. *Nucleic Acids Res.* **33**: 6566–78
- Velema WA & Kool ET (2020) The chemistry and applications of RNA 2'-OH acylation. *Nat. Rev. Chem.* **4**: 22–37
- Vousden KH & Lane DP (2007) p53 in health and disease. *Nat. Rev. Mol. Cell Biol.* **8**: 275–283
- Wan Y, Kertesz M, Spitale RC, Segal E & Chang HY (2011) Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* **12** VN-r: 641–655
- Wang J, Zhang J, Zheng H, Li J, Liu D, Li H, Samudrala R, Yu J & Wong GK-S (2004) Neutral evolution of ‘non-coding’ complementary DNAs. *Nature* **431**: 1–2
- Wang JC, Passage MB, Yen PH, Shapiro LJ & Mohandas TK (1991) Uniparental heterodisomy for chromosome 14 in a phenotypically abnormal familial balanced 13/14 Robertsonian translocation

- carrier. *Am. J. Hum. Genet.* **48**: 1069–74
- Wang X, Song X, Glass CK & Rosenfeld MG (2011) The long arm of long noncoding RNAs: Roles as sensors regulating gene transcriptional programs. *Cold Spring Harb. Perspect. Biol.* **3**: 1–14
- Waterhouse AM, Procter JB, Martin DMA, Clamp M & Barton GJ (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191
- Weinberg Z & Breaker RR (2011) R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics* **12**: 1–9
- Wilkinson KA, Merino EJ & Weeks KM (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1**: 1610–6
- Xu C, Bian C, Yang W, Galka M, Ouyang H, Chen C, Qiu W, Liu H, Jones AE, MacKenzie F, Pana P, Li SSC, Wang H & Min J (2010) Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc. Natl. Acad. Sci. U. S. A.* **107**: 19266–19271
- Yekta S, Tabin CJ & Bartel DP (2008) MicroRNAs in the Hox network: an apparent link to posterior prevalence. *Nat. Rev. Genet.* **9**: 789
- Yeku O & Frohman MA (2011) Rapid amplification of cDNA ends (RACE). *Methods Mol. Biol.* **703**: 107–122
- Yoon J-H, Abdelmohsen K, Kim J, Yang X, Martindale JL, Tominaga-Yamanaka K, White EJ, Orjalo A V., Rinn JL, Kreft SG, Wilson GM & Gorospe M (2013) Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination. *Nat. Commun.* **4**: 2939
- Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, Shen Y, Pervouchine DD, Djebali S, Thurman RE, Kaul R, Rynes E, Kirilusha A, Marinov GK, Williams BA, Trout D, et al (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**: 355–364
- Zerucha T, Stühmer T, Hatch G, Park BK, Long Q, Yu G, Gambarotta A, Schultz JR, Rubenstein JL & Ekker M (2000) A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. *J. Neurosci.* **20**: 709–21
- Zhang X, Rice K, Wang Y, Chen W, Zhong Y, Nakayama Y, Zhou Y & Klibanski A (2010a) Maternally Expressed Gene 3 (MEG3) Noncoding Ribonucleic Acid: Isoform Structure, Expression, and Functions. *Endocrinology* **151**: 939–947
- Zhang X, Zhou Y & Klibanski A (2010b) Isolation and characterization of novel pituitary tumor related genes: a cDNA representational difference approach. *Mol. Cell. Endocrinol.* **326**: 40–7
- Zhang X, Zhou Y, Mehta KR, Danila DC, Scolavino S, Johnson SR & Klibanski A (2003) A Pituitary-Derived MEG3 Isoform Functions as a Growth Suppressor in Tumor Cells. *J. Clin. Endocrinol. Metab.* **88**: 5119–5126
- Zhou Y, Zhong Y, Wang Y, Zhang X, Batista DL, Gejman R, Ansell PJ, Zhao J, Weng C & Klibanski A (2007) Activation of p53 by MEG3 Non-coding RNA. *J. Biol. Chem.* **282**: 24731–24742
- Zhu J, Liu S, Ye F, Shen Y, Tie Y, Zhu J, Wei L, Jin Y, Fu H, Wu Y & Zheng X (2015) Long Noncoding RNA MEG3 Interacts with p53 Protein and Regulates Partial p53 Target Genes in Hepatoma Cells. *PLoS One* **10**: e0139790
- Zhu W, Botticelli EM, Kery RE, Mao Y, Wang X, Yang A, Wang X, Zhou J, Zhang X, Soberman RJ, Klibanski A & Zhou Y (2019) Meg3-DMR, not the Meg3 gene, regulates imprinting of the Dlk1-Dio3 locus. *Dev. Biol.* **455**: 10–18
- Zinshteyn B, Chan D, England W, Feng C, Green R & Spitale RC (2019) Assaying RNA structure with LASER-Seq. *Nucleic Acids Res.* **47**: 43–55