

KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Oncology

VIB Center for Cancer Biology



The long and the short of *NEAT1*: Roles in paraspeckle formation and cancer.

Carmen ADRIAENS

Jury:

Promoter: Prof. Dr. Jean-Christophe Marine

Chair: Prof. Dr. Johan Swinnen

Secretary: Prof. Dr. Anna Sablina

Jury members:

Prof. Dr. Torben Heick Jensen

Prof. Dr. Alexandra Van Keymeulen

Prof. Dr. Jan Cools

Prof. Dr. Anna Sablina

Dissertation presented in
partial fulfilment of the
requirements for the
degree of Doctor in the
Biomedical Sciences.

May 2019

Deze is voor u, Ines

Acknowledgements

First of all, I would like to thank all members of my jury, in particular prof. Jensen, prof. Van Keymeulen, prof. Sablina and prof. Cools for reading and providing feedback on the manuscript, and all the members of the thesis advisory and evaluation committees for their input throughout my doctorate. I would like to thank IWT and the Koning Boudewijn Stichting for the funding they provided.

This thesis is dedicated to my twin sister, without whom I would not have been able to do this. Ines, I admire your drive and effectiveness, you managed to motivate me time and again when I was stuck. You would go through fire for anyone, and you are my example in almost everything I do. I am fortunate I can call you not only my sister, but also my friend.

Chris, I remember your words on Laura's defense: "This was a bumpy road, but we made it". Back then, I didn't understand much of what that meant, but I do now. Thank you for having hosted me in the lab, having taught me the things you know, and especially for having given me the freedom and opportunity to pursue my ideas and projects. I have grown a lot in these five years in your lab. I am grateful for your patience and belief in me. It's been a bumpy road for us, too, but I'm convinced we learned a lot along the way.

Laura, although you have left a very long time ago, I still feel I owe many of my skills and much of my knowledge to you. I enormously admire your steadfastness and intelligence. I have been lucky to have had you as my mentor.

Greet, maybe from the outside, one could say that if I was the brain, you were my hands... But I have to disagree. You cannot imagine how much I feel blessed to have had you as my colleague and friend. The endless times you finished experiments, listened to me trying to make sense of my ideas, read and reread my writings... You were always there for me, you're the rock of the team.

Odessa, our sunny breaks still make me smile every day. You are an awesome mom and I am jealous of your beautiful garden. I hope we keep in touch!! Véro, I am deeply grateful for everything you have done for the project. I feel honored to have gained your trust, you are one of the kindest people I know. George, you were one of my favorite colleagues and I still miss you. I always felt at home when you were around. Karen, I loved our beauty time together and chatting science and beyond! Aljosja, it was a delight to have you around all the late nights and weekends in the lab. Corinna, lots of things happened in our time together, and I hope you find your way! Oskar, I liked our scientific discussions a lot, and although you went on a different road topic-wise, I'm sure you will be successful no matter what challenge you take up.

Takis, I remember I liked you already when you came for interview in the lab. I wish you the best with your projects! Joanna, we have not been colleagues for long, but I can tell you are gentle and talented. I'm glad we met! Dennis & Nina, I really enjoy your kind and easy-going personalities, it was a pleasure to have you around. Ewout, it will be hard, but you can do it! Florian, you will be an awesome PI if you choose to take that path. I look forward to sharing a beer with you (Maybe in a conference? Let's do whiskey instead.) or go for a hike and lunch with Sam and the little one when I'm around. Michael, I owe you so much. You were always patient with my endless technical questions. You are a great teacher and I'm lucky I got to take advantage of your presence in the lab. I really enjoy your bubbly personality and joking around, too! Roberto & Jasmine, I wish you all the best in your future endeavors. Amelie, Iria & Albi, you guys are a joyful bunch and I liked the short time we had in the lab together. I hope you fare well! Veerle, thank you for helping me with all the PhD admin, I enjoyed being your deskmate!

Tom, your cheer enthusiasm for science is contagious. The people in your lab are lucky to have you as their boss (I know nobody as efficient and fair), and I am profoundly grateful to have been able to be part your lab, try new techniques, and especially get in touch with a lot of exciting science and great people.

I would also like to thank those who do not make part of the Marine or the Misteli labs but with whom I have interacted, extensively or not, throughout the project: Paulo, much of this work is through you. Prof. A. Blackford and prof. C. Schmidt, thank you for your input and all the helpful discussions during and after my time in Cambridge. I am grateful to preLights, and in particular Máté Palfy from the Company of Biologists for the opportunity to be part of the preLights community this past year. I would also like to thank prof. A. Fox for her *Neat* enthusiasm.

Doing a Ph.D. isn't just a blip in a life, and I only fully realized this when I was well on the way of obtaining one. Science is difficult, not just because we are constantly trying to figure out the unknown, but also because it involves so much more we are not trained or prepared for. Therefore, I would also like to thank and acknowledge some other special people in my life for having supported me along the way. Mama, papa, Petra, Mona and Hannah, I know the door was always open. Bryce, Jasper, Heleen, Luis, Dan, Paul, Sandra, Martin, Lily, and everyone else who was or wasn't around in these years, I feel tremendously grateful for the many times my busy self could rely on you for anything.

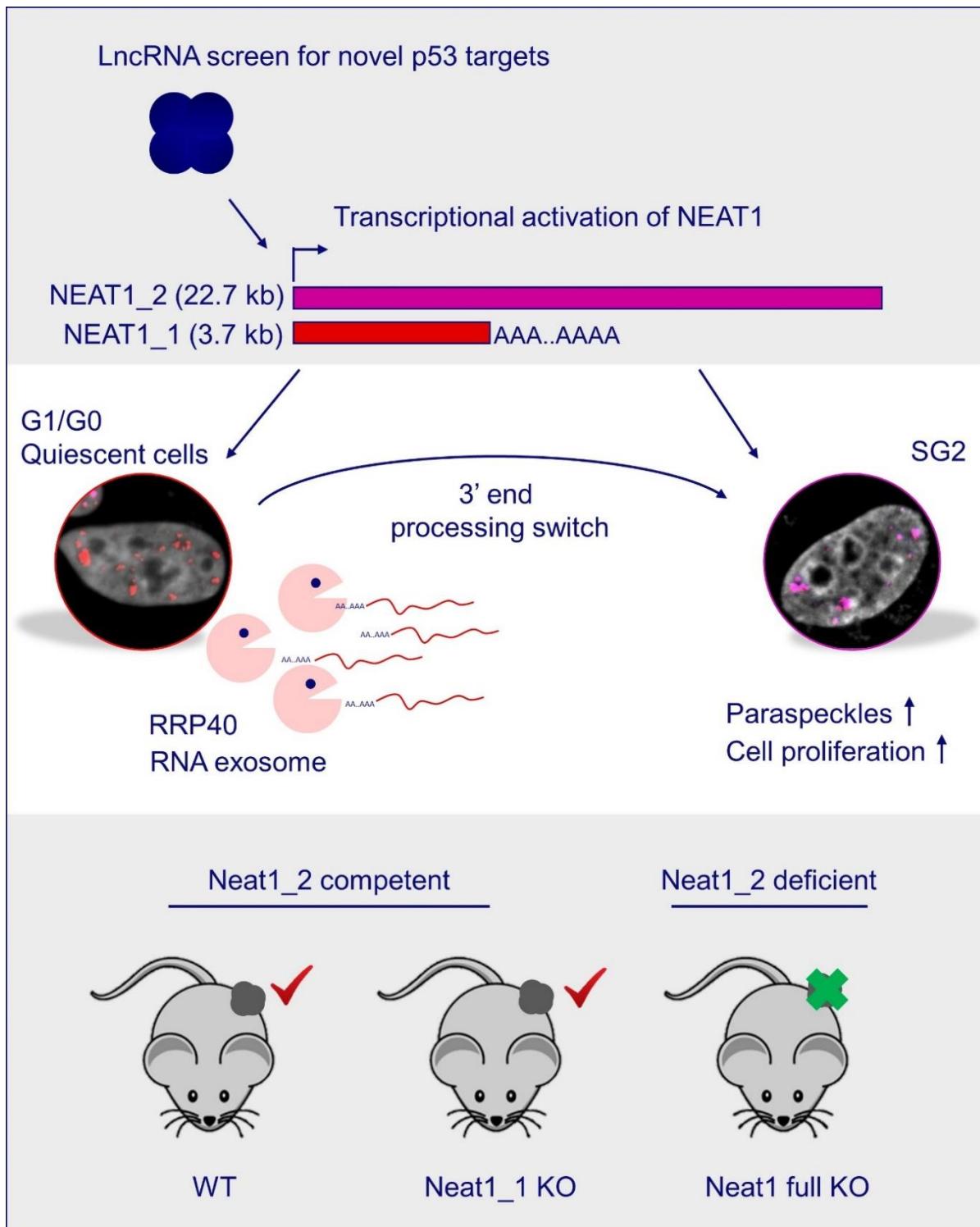
Finally, I dedicate this quote to Delphine, who was there from the start: « On ne voit bien qu'avec le cœur; l'essentiel est invisible pour les yeux. » (The Little Prince)

Table of Contents

Table of Contents	I
Graphical abstract	III
Summary.....	V
Samenvatting	VII
List of Figures and Tables	IX
List of abbreviations	XI
Part I: Introduction	1
<i>Chapter I: Long non-coding RNA</i>	3
Long non-coding RNA: basic features.....	3
RNA turnover.....	5
LncRNA in cancer.....	9
Mammalian <i>in vivo</i> cancer models to study lncRNA	13
<i>Chapter II: NEAT1 and Paraspeckles</i>	16
Paraspeckle structure, components, and definition.....	16
Proposed functions and paraspeckle mechanisms of action.....	22
Conclusion	24
Part II: Objectives	27
Part III: Results.....	31
<i>Chapter I</i>	33
Abstract.....	35
Introduction	37
Results	38
Discussion.....	57
Materials and methods.....	60
Acknowledgements	67
Contributions	67
Personal Contributions (CA)	67
Competing interests	67
References.....	68
Appendix: Supplementary Table I.....	69
<i>Chapter II</i>	73
Abstract	75
Introduction	77
Results.....	79

Discussion.....	99
Materials and methods.....	101
Author contributions	107
Competing interests	107
Acknowledgements	107
References.....	107
Part IV: General Discussion.....	109
<i>Chapter I: On NEAT1 and paraspeckles in cancer.....</i>	111
Altering <i>NEAT1</i> levels in cancer.....	111
<i>NEAT1</i> downstream of p53: not a tumor suppressor?.....	111
Replication stress: cause of <i>NEAT1</i> KD induced DNA damage	113
<i>NEAT1</i> mutations: (not) just bystanders	114
<i>Chapter II: On NEAT1 and paraspeckle function</i>	116
Link with DNA repair.....	116
Potential functions for <i>NEAT1_1</i>	117
<i>Chapter III: On NEAT1 and paraspeckle structure.....</i>	119
Paraspeckle formation requires the right environment	119
Determining specificity for <i>NEAT1</i> nucleation	119
<i>Chapter IV: Global conclusion and outlook</i>	121
Part V: References	123
Part VI: Addendum	141
Curriculum Vitae.....	143
preLights	147

Graphical abstract



Summary

NEAT1 is a well-conserved, highly expressed nuclear RNA that does not code for proteins. It exists in two isoforms of 3.7 and 22.7 kb, of which the longer is the essential architectural component of a nuclear body in the cell, the paraspeckle.

In this thesis, we show that the tumor suppressor p53 upregulates *NEAT1* transcription leading to the formation of paraspeckles in cultured cells, in mouse skin, and in human tumors of epithelial origin. We used a mouse model of chemically-induced skin carcinogenesis combined with a genetic knockout of *Neat1* to demonstrate that this lncRNA contributes to tumor formation. Next, we showed, using *in vitro* studies, that *NEAT1* knockdown sensitizes cancer cells to common chemotherapeutic agents such as doxorubicin and the PARP-inhibitor ABT-888, and that knockdown without external stress already caused an accumulation of DNA damage on its own. In particular, knockdown of the long isoform *NEAT1_2*, the central building block of the paraspeckle nuclear body, is sufficient to do so, indicating that it may be the paraspeckle which is the functional unit in this process. Knockdown of *NEAT1* prevented proper activation of the ATR-Chk1 pathway, decreasing the appropriate DNA repair. This decreased signaling could potentially lead to an increase in collapsed replication forks, causing an increase in S139-phosphorylated Histone 2A.X (γ -H2A.X) and an arrest of the cells in the next G1 phase. Finally, higher expression levels of the long isoform in a cohort of cisplatin-treated ovarian cancer patients could predict poor survival, whereas considering both isoforms together did not have predictive value.

The isoform specificity of these latter observations (i.e. knockdown of the long isoform alone was sufficient to induce a growth phenotype) prompted us to follow up on the individual roles of *NEAT1_1* and *NEAT1_2*. We found that despite a common regulation by p53 at the *NEAT1* promoter, both isoforms are differentially expressed upon different forms of stress, a process mediated by alternative processing of the *NEAT1_1* short isoform at its 3' end. Nutlin-3a, which causes pharmacological activation of p53 and arrests the cells in the G1 phase, led to significant amounts of *NEAT1_1* found outside of the paraspeckle nuclear bodies. In contrast, hydroxyurea stalls the cells mid-S phase in addition to activating p53. When treated with this agent, we failed to detect *NEAT1_1* outside of nuclear bodies in the majority of the cells. In a non-treated, mixed population, the proportion of cells expressing *NEAT1_1* outside of paraspeckles varies greatly. These observations suggested that *NEAT1* isoform configurations in the cell could be cell cycle dependent.

We subsequently found that *NEAT1_1* is the predominant isoform in quiescent (G0) and G1 cells, whereas this isoform disappeared and *NEAT1_2* and paraspeckles became primarily expressed from replication (S phase) onwards. Moreover, CRISPR-mediated knockout of the

polyadenylation signal revealed that the short isoform does not contribute to the observed replication stress and cell cycle arrest upon knockdown of both. In addition, this isoform does not contribute to *NEAT1_2* and paraspeckle function, at least when considering its role of protecting against replication stress.

In fact, the data presented here suggest that a long-standing paradigm in paraspeckle research may need to be challenged: from our work it appears that *NEAT1_1* does not contribute, at all, to paraspeckle function or formation, and that the hypothesis of *NEAT1_1* recruitment into the paraspeckle may be wrong. In effect, in addition to our estimate that *NEAT1_1* expression is almost zero in conditions where only *NEAT1_2* is observed, we also found a mechanism of *NEAT1_1* degradation by the RNA exosome. Here, further studies are needed to establish a cell cycle stage-specific role of the RNA exosome-mediated degradation process.

Finally, repeating our short-term skin carcinogenesis experiment in mice lacking both *Neat1* isoforms as before, but now also using a knockout mouse model of only *Neat1_1* indicated that *Neat1_2* competent cells (WT and knock out for *Neat1_1*) behave in a similar fashion, whereas mice that are *Neat1_2* and paraspeckle incompetent show a decrease in epidermal hyperplasia and an increase in DNA damage as shown in the first part of this work. Similarly, phenotypes described by our lab and others affecting the mammary gland and other aspects of mouse fertility could not be recapitulated in mice lacking only *Neat1_1*, indicating that it is effectively the long isoform alone that confers a survival and growth advantage when cells need to withstand the internal and external stressors.

In summary, the work presented in this thesis revealed that cancer cells can hijack the protective function of *NEAT1_2*/paraspeckles, but, that the more abundant and better conserved *NEAT1_1* short isoform does not seem to contribute to this process, and that, although well studied, many aspects of *NEAT1* and paraspeckle biology still remain to be understood.

Samenvatting

NEAT1 is een abundant nucleair RNA dat sterk geconserveerd is doorheen de evolutie van zoogdieren, en dat niet codeert voor eiwitten. Van het *NEAT1* locus in het DNA worden twee isovormen afgeschreven, *NEAT1_1* (3.7 kb) en een langere *NEAT1_2* (22.7 kb). Deze laatste is het zgn. ‘geraadde’, of de architecturale hoofdcomponent van een membraanloos deeltje in de celkern, de paraspeckle.

In deze thesis tonen we aan dat het tumor-onderdrukkend eiwit p53 de transcriptie van *NEAT1* induceert, wat leidt tot de vorming van paraspeckles in cellen in weefselweek, in de huid van de muis en in menselijke tumoren van epitheliale origine. We gebruikten een muismodel van chemisch-geïnduceerde huidkanker samen met de genetische knock-out van *Neat1* om aan te tonen dat dit lang, niet-coderend RNA bijdraagt tot het ontstaan van tumoren. Vervolgens toonden we aan met *in vitro* studies dat het verlies van *NEAT1* door middel van knockdown kankercellen gevoeliger maakt aan courante chemotherapeutische medicijnen zoals doxorubicine en de PARP-inhibitor ABT-888. Knockdown zonder externe toevoeging van stress veroorzaakt bovendien reeds een accumulatie van schade aan het DNA. De knockdown van de lange isovorm *NEAT1_2*, de centrale bouwsteen van de paraspeckle, is voldoende om deze fenotypes te veroorzaken. Dit suggereert dat de paraspeckles de functionele eenheden zijn in dit proces. Knockdown van *NEAT1* voorkomt een juiste activatie van de ATR-Chk1 pathway wat het vereiste DNA-herstel vertraagt. Deze verlaagde signalisatie kan mogelijk leiden tot een verhoging in gebroken replicatie-vorken in het DNA, wat ervoor zorgt dat de cellen stoppen met delen in de volgende G1 fase. Ten slotte kunnen hogere expressieniveaus van de lange isovorm in een cohort van cisplatinum-behandelde eierstokkanker-patiënten een zwakkere overlevingskans voorspellen, daar wanneer beide isovormen in acht worden genomen is er geen voorspellende waarde voor de patiënt.

Het feit dat deze observaties isovorm-specifiek zijn (bv. knockdown van de lange isovorm alleen was genoeg om een groeifenotype te induceren) gaf ons een goede reden om de individuele bijdrage van elke isovorm te onderzoeken in het tweede deel van dit werk. We vonden dat, ondanks een gezamenlijke regulatie door p53 op promoter-niveau, beide isovormen op een verschillende manier tot expressie komen bij verschillende vormen van stress. Dit proces wordt gemedieerd door alternatieve processsing van het 3' einde van de korte isovorm *NEAT1_1*. Nutlin-3a activeert p53 farmacologisch en stopt de cellen in de G1 fase van de celcyclus. Behandeling met Nutlin-3a leidt tot grote hoeveelheden *NEAT1_1* detecteerbaar buiten de paraspeckles. Hiermee in contrast stopt hydroxyurea de cellen in het midden van de S-fase bovenop de activatie van p53. Wanneer we de cellen met deze stof behandelen konden we *NEAT1_1* niet meer buiten de nucleaire deeltjes vinden in de

meerderheid van de cellen. In een niet-behandelde gemengde populatie varieert de proportie van *NEAT1_1*-tot expressie brengende cellen buiten de paraspeckles sterk. Deze observatie suggereert dat de *NEAT1* isovorm configuraties in de cel celcyclus afhankelijk kunnen zijn.

We vonden vervolgens dat *NEAT1_1* de voornaamste isovorm is in rustende (G0) en G1 cellen, en dat deze isovorm verdwijnt vanaf de S-fase (DNA-rePLICATIE). Bovendien vonden we dat de korte isovorm uit de cel gehaald kon worden zonder de lange vorm aan te tasten door middel van een deletie van het polyadenylatiesignaal met behulp van het CRISPR-systeem. Deze cellen, in dewelke de korte vorm niet tot expressie komt, groeien normaal en vormen normale paraspeckles. Knockdown van de lange vorm in deze cellen zorgt nog steeds voor een groeiarrest wat impliceert dat de lange vorm functioneel blijft wanneer de korte niet aanwezig is.

In feite suggereert de data hier dat een lang aanvaard paradigm in paraspeckle onderzoek niet correct zou kunnen zijn: uit ons werk blijkt immers dat *NEAT1_1* niet bijdraagt tot paraspeckle-functie of vorming en dat de hypothese dat *NEAT1_1* wordt gerekruteerd in de paraspeckles fout zou kunnen zijn. Inderdaad, bovenop onze schatting dat *NEAT1_1* expressie bijna nul is in condities in dewelke enkel *NEAT1_2* wordt geobserveerd, vonden we ook een mechanisme van *NEAT1_1* degradatie door het RNA exosoom. Hier zijn verdere studies noodzakelijk om een celcyclus-specifieke rol van het RNA exosoom-gemedieerde degradatieproces vast te stellen.

Ten slotte herhaalden we ons korte termijn huidcarcinogenese-experiment met muizen die enkel *Neat1_1* ontbreken. Dit experiment toonde aan dat *Neat1_2* competente cellen (WT en knockout voor *Neat1_1*) zich gedragen op een gelijkaardige manier, in tegenstelling tot muizen die *Neat1_2*- en paraspeckle-incompetent zijn, en die een verlaging tonen in epidermale hyperplasie en een verhoging in DNA-schade zoals aangetoond in het eerste deel van dit werk. Op een gelijkaardige manier kunnen de fenotypes beschreven door ons lab en anderen die de borstklier en andere aspecten van de vruchtbaarheid van de muis beïnvloeden niet teruggevonden worden in muizen die enkel de korte vorm ontbreken, wat aantoont dat het effectief de lange vorm is die een groei- en overlevingsvoordeel geeft wanneer cellen moeten weerstaan aan interne en externe stressors.

Samengevat, het werk gepresenteerd in deze thesis toonde aan dat kankercellen de beschermende functie van *NEAT1_2*/paraspeckles kapen, maar dat de meer abundante en beter geconserveerde *NEAT1_1* isovorm niet blijkt bij te dragen aan dit proces. Dus, alhoewel *NEAT1* een populair studieobject is, blijven vele aspecten van *NEAT1* en paraspeckle biologie nog steeds een mysterie.

List of Figures and Tables

Table of Figures

Part I: Introduction

Chapter I: Long non-coding RNA

- Figure 1: Classes of lncRNAs.*
- Figure 2: Basic modules of lncRNA action.*
- Figure 3: The human exosome.*
- Figure 4: Simplified scheme of the p53 pathway.*
- Figure 5: Oligo-based strategies to target lncRNA.*
- Figure 6: Different Cre-LoxP modifications of the mouse genome.*

Chapter II: NEAT1 and paraspeckles

- Figure 7: Paraspeckle proteins and nucleoli.*
- Figure 8: Basic features of the NEAT1 locus.*
- Figure 9: 3' end processing of NEAT1_1.*
- Figure 10: 3' end processing of NEAT1_2.*
- Figure 11: NEAT1_2 oligo-A stretch.*
- Figure 12: Models for paraspeckle function.*

Part II: Objectives

Part III: Results

Chapter I

- Figure 1. p53 induces NEAT1 expression and paraspeckle formation.*
- Figure S1. A network of lncRNAs downstream of p53.*
- Figure 2. DNA damage induces NEAT1 paraspeckle formation.*
- Figure S2. NEAT1 is a direct p53 target gene.*
- Figure 3. Neat1 KO mice are resistant to chemically-induced skin cancer formation.*
- Figure S3. DNA damage induces NEAT1 expression and paraspeckle formation.*
- Figure 4. Neat1 prevents accumulation of DNA damage and p53.*
- Figure S4. Neat1 KO animals do not show proliferation defects and increased apoptosis after the first DMBA application.*
- Figure S5. RNase-H and RNAi-dependent targeting of NEAT1 causes G1 arrest.*
- Figure 5. NEAT1 paraspeckles modulate ATR-signaling and chemosensitivity.*
- Figure S6. Increased NEAT1 paraspeckle formation in human cancers compared to normal matching tissues.*
- Supplementary Table S2. Site of origin, histology and NEAT1 score of primary carcinomas included in tissue microarrays.*
- Figure 6. NEAT1 paraspeckles are detected in human cancers and NEAT1_2 levels correlate with response to platinum-based therapy.*
- Figure S7. Paraspeckles do not colocalize with DNA damage foci and paraspeckle proteins are excluded from DSBs.*
- Supplementary Table S1. List of putative p53-target lncRNA genes.*

Chapter II

- Figure 1: NEAT1 isoforms are induced differentially by different p53 activating stressors.*
- Figure S1: NEAT1_2 accumulations in HU conditions are genuine paraspeckles.*
- Figure 2: NEAT1 isoforms are alternatively regulated during cell cycle.*
- Figure S2: Paraspeckle and NEAT1_1 characteristics in different cell cycle phases.*
- Figure S3: RT-qPCR analysis of NEAT1 isoforms in different cell cycle stages.*
- Figure S4: NEAT1_1 is rapidly degraded by the RNA exosome.*
- Figure 3: NEAT1_1 is rapidly and specifically degraded by the RNA exosome.*
- Figure 4: Loss of NEAT1_1 in proliferating cancer cells does not impact on cell growth.*
- Figure S5: NEAT1_1 KO in HCT116 does not induce growth phenotypes.*
- Figure S6: NEAT1_1 KO cells make genuine paraspeckles.*
- Figure 5: NEAT1_1 KO only causes limited changes in gene expression.*
- Figure S7: Loss of Neat1_1 or both isoforms in untransformed P3 mouse embryonic fibroblasts (MEFs) does not affect cell growth.*
- Figure 6: Neat1 full KO phenotypes are due to lack of Neat1_2, not Neat1_1.*

Part IV: Discussion

Part V: References

Part VI: Addendum

List of abbreviations

µg	Microgram
µM	Micromolar
ABT-888	Veliparib, a PARP inhibitor
ANOVA	Analysis of Variance
ASO	Antisense Oligonucleotides
A,T,G,C	Adenine, Thymine, Guanine, Cytosine
bp	Base Pairs
c.i.	Confidence interval
cDNA	Complementary DNA
cGAS	Cyclic GMP-AMP synthase
CHART	Capture Hybridization Analysis of RNA Targets
ChIP	Chromatin Immunoprecipitation
CRISPR	Clustered Regularly InterSpaced Palindromic Repeats
d	Days
DAPI	4',6-diamidino-2-phenylindole
DBHS	Drosophila Behavior, Human Splicing
DDR	DNA Damage Response
DE	Differentially Expressed
dH2O	Demineralized water
DNA/RNA	(Deoxy)ribonucleic acid
dNTP	Nucleoside Triphosphate
Doxo	Doxorubicin
DSB	Fluorescent in-situ hybridization
ENCODE	Encyclopedia of DNA elements
FBS	Fetal Bovine Serum
FFPE	Formalin-fixed, paraffin embedded
FISH	Fluorescent In Situ Hybridization
g	Gravitational force
G0/G1/G2	Growth phase 1, 0, or 2
GEMMs	Genetically Engineered Mouse Models
GEO	Gene Expression Omnibus
GMP/AMP	Cyclic guanosine monophosphate–adenosine monophosphate
H3K27Ac	Histone 3 Lysine 27 Acetylation
H3K4me3	Histone 3 Lysine 4 trimethylation
HDP-RNP	HEXIM1-DNA-PK-paraspeckle components-ribonucleoprotein complex
HDV	Hepatitis Delta Virus
HET	Heterozygous
HF	Hair Follicle
Hg19	Human Genome 19
HIV	Humane Immunodeficiency Virus
HR	Hazard Ratio
HSV	Herpes Simplex Virus
HU	hydroxyurea
Id	Identity
IFE	Interfollicular epidermis
Kb	Kilobases
KD	Knockdown
LINE	Long Interspersed Nuclear Element

LNA	Complementary DNA
lncRNA/lincRNA	Long (intergenic) non-coding RNA
m6A	6-methyl-adenosine
m7G	7-methylguanylate cap
mESCs	Mouse Embryonic Stem Cells
mm	Millimeters
MMTV	Mouse Mammary Tumor Virus
mRNP	Messenger RiboNucleoProtein
N	Normal
n.s.	Not significant
NEXT	Nuclear Exosome Targeting Complex
NGP	Neuroblastoma Cells
NHEJ	Non-Homologous End Joining
NT	Non Treated
OCT	Optimal Cutting Temperature
Padj	Adjusted p-value
PARIS	Psoralen analysis of RNA interactions and structures
PAS	Polyadenylation Signal
PAXT	PolyA Tail Exosome Targeting
PCO	Protein-Coding Overlapping
PFS	Progression-free Survival
PI	Propidium iodide
PML body	Promyelocytic Leukaemia bodies
PROMPTs	Promoter Upstream Transcripts
PS	Paraspeckle
PWM	Position Weight Matrix
PYMT	Polyoma Virus middle T antigen
RIPA	Radioimmunoprecipitation assay
RISC	RNA-induced silencing complex
RNAi	RNA interference
ROS	Reactive Oxygen Species
Rot	Rotenone
RRMs	RNA recognition motifs
RT-qPCR	Quantitative reverse transcription Polymerase Chain Reaction
s.d.	Standard deviation
s.e.m.	Standard error of the mean
S/M	S phase / mitosis
SCC	Squamous Cell Carcinoma
SHAPE	Selective 2'-hydroxyl acylation analyzed by primer extension
sh-CTRL	Short-hairpin control
sh-P53	Short-hairpin p53
SINE	Short Interspersed Nuclear Element
siRNA	Small interfering RNA
SM/SM-like	A group of RNA binding proteins named after Stephanie Smith, who suffered from Systemic Lupus Erythematosus
ssDNA	Single stranded DNA
T	Tumor
t-/r-/sn-/mi-/m-/ce	Transfer/ribosomal/small nuclear/micro-/messenger/competing endogenous RNA
T-ALL	T-cell Acute lymphoblastic Leukaemia
TAM	Tamoxifen

TCGA	The Cancer Genome Atlas
TGF-β	Transforming Growth Factor beta
TMA	Tissue micro-array
TNBC	Triple Negative Breast Cancer
TPA	Tissue plasminogen activator
TRAMP	Trf4/Air2/Mtr4p Polyadenylation complex
TSS	Transcription Start Sites
Txn	Transcription
UCSC	University of California Santa Cruz
UV	Ultraviolet
WT/KO	Wild Type / Knock Out
FDA	Food and Drug Agency
FIGO	The International Federation of Gynecology and Obstetrics
FWO	Fonds Voor Wetenschappelijk Onderzoek
IWT	Instituut voor Wetenschap en Technologie
SCBT	Santa Cruz Biotechnology
VIB	Vlaams Instituut voor Biotechnologie
BJ	Human foreskin fibroblasts
HCT116	Colorectal Carcinoma cell line
HeLa	Ovarian Cancer Cell Line
HSMM	Human Skeletal Muscle Myoblasts
HUVEC	Human Umbilical Vein Endothelial Cell
MCF-7	Michigan Cancer Foundation-7
MEFs	RNA recognition motifs
MM118	Malignant Melanoma 118
NHEK	Normal Human Epidermal Keratinocytes
NHLF	Normal Human Lung Fibroblasts
U2OS	Human Bone Epithelial OsteoSarcoma cells
WI38	Diploid human lung cell fibroblasts
7SK	Small nuclear RNA 7SK
AGO1	Argonaute RISC component 1
AID	Activation-induced cytidine deaminase
AIRC	PAICS phosphoribosylaminoimidazole carboxylase and phosphoribosylaminoimidazolesuccinocarboxamide synthase
ALB	Albumin
APC	Adenomatous polyposis coli
ARF	Cyclin Dependent Kinase Inhibitor 2A Alternative Reading Frame
ATF2	Activating transcription factor 2
ATM	Ataxia telangiectasia mutated
ATR	ATR serine/threonine kinase
B23	NPM1; nucleolar phosphoprotein nucleophosmin 1
B2M	Beta-2 microglobulin
BAX	BCL2 Associated X
BRG1	SMARCA4 SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4
C9ORF72	C9orf72-SMCR8 complex subunit
CARM1	Coactivator associated arginine methyltransferase 1
Cas9	CRISPR-associated protein 9

CDC5L	Cell division cycle 5 like
CDK9	Cyclin Dependent Kinase 9
CDKN1A	Cyclin Dependent Kinase Inhibitor 1A
CPSF6	Cleavage and polyadenylation specific factor 6
CTCF	CCCTC-binding factor
Ctn-RNA	Slc7a2 solute carrier family 7
DANCR	Differentiation antagonizing non-protein coding RNA
DNAPKcs	DNA-dependent protein kinase catalytic subunit
dRRP6	Ribosomal RNA-processing protein 6
eGFP	Enhanced Green Fluorescent Protein
EXO9	Exosome component 9
EZH2	Enhancer of zeste 2 polycomb repressive complex 2 subunit
g-H2A.X	Serine 139 phosphorylated Histone 2 A.X
hDis3	Human chromosome Disjunction 3
HER2	Erb-b2 receptor tyrosine kinase 2
HEXIM1	HEXIM P-TEFb complex subunit 1
HIF2a	Hypoxia Inducible Factor 2a
hnRNP K	Heterogeneous nuclear ribonucleoprotein K
HOAIR	HOX transcript antisense RNA
HSP1	Heat Shock Protein 1
IRF3	Interferon Regulatory Factor 3
K1	Keratin 1
KAP1	TRIM28 tripartite motif containing 28
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
Ku70-80	XRCC6-X-ray repair cross complementing 6-5
lncRNA-HIT	HOXA transcript induced by TGFβ
Malat1	Metastasis Associated Lung Adenocarcinoma Transcript 1
MATR3	Matrin 3
mCAT	Malonyl-CoA-acyl carrier protein transacylase
MDM2	Mouse Double Minute 2
MDM4	Mouse Double Minute 4
MYC	Myelocytomatosis proto-oncogene
NEAT1	Nuclear Enriched Paraspeckle Transcript 1
NONO/p54nrb	Non-POU domain containing octamer binding
NOXA	PMAIP1 - Phorbol-12-myristate-13-acetate-induced protein 1
NUDT21	Nudix hydrolase 21
PARP	Poly (ADP-ribose) polymerase
p-H3	Phosphorylated Histone 3
PSF/SFPQ	PTB-associated splicing factor / splicing factor proline and glutamine rich
PSPC1	Paraspeckle Protein Component 1
p-TEFb	Positive Transcription Elongation Factor beta
PUMA	p53 upregulated modulator of apoptosis
PURPL	p53 Upregulated Regulator Of P53 Levels
RBM7/14	RNA Binding Motif protein7/14
RNR	Ribonucleotide Reductase
RPA32	Replication Protein A 32
RRP40	EXOSC3 - Exosome complex component RRP40
RRP6	Exosome nuclease subunit 6
SAMMSON	Survival Associated Mitochondrial Melanoma Specific Oncogenic ncRNA

SFTPB	Surfactant protein B
STAR-PAP	Non-canonical polyA polymerase
SWI/SNF	Switch/sucrose non fermenting
TBP	TATA Binding protein
TDP-43	TAR DNA binding protein
THOR	Testis-associated highly conserved oncogenic long non-coding RNA
TP53	Tumor protein p53
TP53LNC	TP53-long non-coding
UBC	Ubiquitin C
VINC	Virus inducible, non-coding
WDR5	WD repeat domain 5
Xist	Inactive X specific transcripts
XRN1/2	5'-3' Exoribonuclease 1/2

Part I: Introduction

Chapter I: Long non-coding RNA

The eukaryotic cell nucleus is compartmentalized into multiple functionally and compositionally distinct domains. In many cases, these discrete regions consist of mixes of DNA, RNA, histones, and other structural and regulatory proteins¹. In addition to the chromatin-rich compartments, some nuclear domains are rather devoid of chromatin and represent smaller, membraneless organelles often established through phase separation²⁻⁴. These entities are thought to physically and functionally separate transcriptional outputs and RNA processing events within the nuclear space through molecular crowding mechanisms⁵⁶, e.g. the PML bodies⁷, nucleoli⁸, or clusters of RNA Polymerase II and Mediator⁹. One of these nuclear domains is the paraspeckle suborganelle (PS), which is defined as a phase-separated mixture of RNA and proteins built upon the long non-coding RNA (lncRNA) *NEAT1_2*, the 22.7 kb isoform of Nuclear Paraspeckle Assembly Transcript 1¹⁰⁻¹⁵.

Here, I will first introduce some basic aspects of long non-coding RNA and, where relevant to this work, nuclear RNA degradation mechanisms. With a side-focus on the tumor suppressor p53, I will further briefly discuss the role of lncRNA in cancer as well as their potential as cancer therapeutic targets and how this can be studied in the lab *in vivo*. In a second chapter of the Introduction of this manuscript, I will lay out the current state-of-the-art on *NEAT1* research and its role in paraspeckle biology, definitively setting the stage for the results presented in the rest of this thesis. In the Results section, I will address a number of questions in the fields of paraspeckle and *NEAT1* biology. I will present data on its role in cancer formation, and talk about some of the more functional and compositional aspects of these nuclear bodies and lncRNA. Finally, in the last main section of this work, I will put the results of this thesis back into their broader context, and, with newly gained insights, discuss some aspects in the field of *NEAT1*/paraspeckles, in particular on their role in cancer, on their structure, and finally on their function as well.

Long non-coding RNA: basic features

After the completion of the Human Genome Project at the turn of the millennium, researchers came to the realization that although almost 80 % of the genome is transcribed, only ~ 2% of these transcripts encode proteins^{16,17}. Moreover, subsequent studies from the Encyclopedia of DNA Elements (ENCODE) project identified biochemically meaningful elements in an exceptionally large portion of the genome¹⁸. These observations combined suggest that much of our nuclear material is functioning, albeit often in a structural role, providing the right context for protein-coding and non-coding gene expression.

A large part of the functional genome is transcribed into non-coding RNA^{19,20}. Although several categories of non-coding RNAs are described beyond the canonical tRNA and rRNAs, e.g. microRNA, small nucleolar RNA, or piwiRNAs, here, we will focus mostly on long non-coding RNAs (lncRNAs).

LncRNAs are RNA molecules that, by definition, exceed 200 nucleotides in length and lack appreciable protein coding potential²¹. They can either be retained in various locations in the nucleus, or trafficked to other cellular compartments in the cytoplasm²². LncRNA classification is based on their relation to the nearest coding gene. For instance, they can be transcribed from intergenic regions (lincRNA), or overlap with other genes, and may be generated as sense or antisense transcripts²³ (figure 1). LncRNAs are conserved to varying degrees, but usually less so than protein coding genes²⁴. Interestingly, while the number of coding genes is generally not predictive of organismal complexity, the number of non-coding transcripts may be^{25,26}.

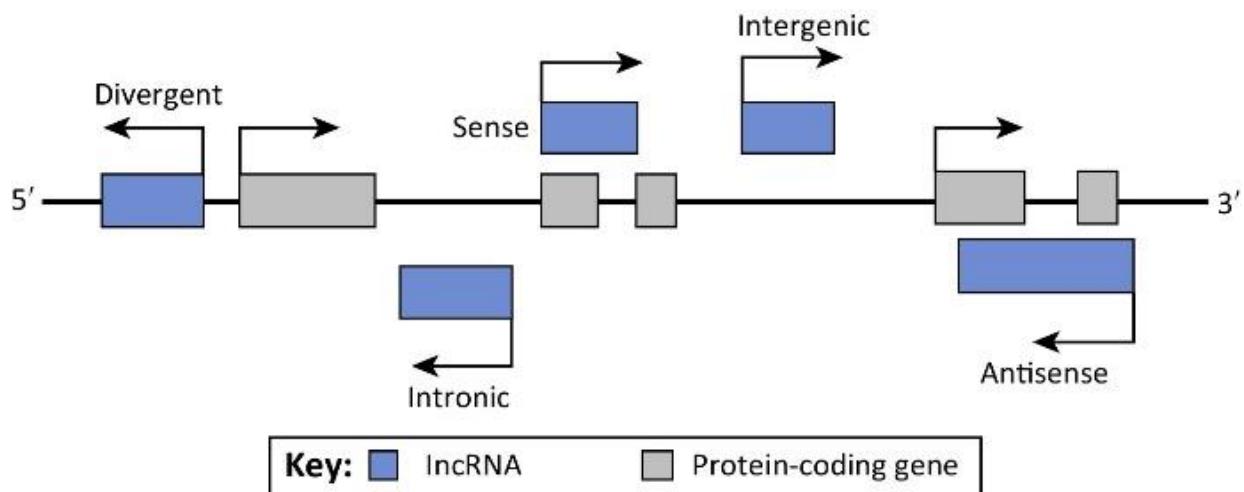


Figure 1: Classes of lncRNAs. Long non coding RNAs are classified based on their relation to the nearest protein coding gene, eg. when transcribed in the opposite away from the protein coding gene (divergent) or the same (sense) or opposite (antisense) direction, when transcribed from the intronic region or from a region in between two protein coding genes (intergenic). Adapted from Zhao & Lin, 2015²⁷.

LncRNAs function in the cell by interacting with chromatin, other proteins, RNA, and other macromolecules²⁸ (Figure 2). As they are thought to help to fine-tune complex organism biology, lncRNAs are precisely regulated to be highly cell type and biological process-specific²⁹. They take part in numerous cellular functions such as gene regulation³⁰, metabolism²⁷ and 3D genome organization³¹. For instance, they act both in cis and in trans to direct chromatin modifying proteins to individual loci in the genome in a context-dependent manner³², they can interact with transcription (co-)factors to modify their activity³³, or define

chromatin interactions by associating with, for example, the chromatin insulator protein CTCF³⁴ or Mediator³⁵ and their cofactors.

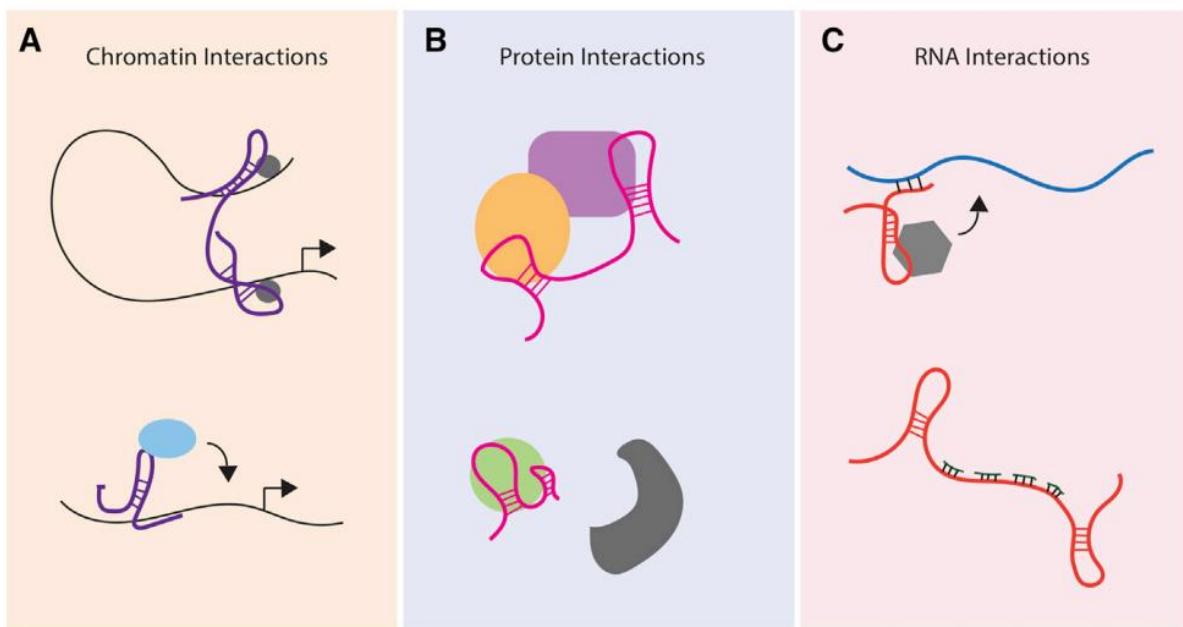


Figure 2: Basic modules of lncRNA action. LncRNAs function by interacting with other macromolecules, e.g. with chromatin (A), non-chromatin proteins (B) or other RNAs (C). Adapted from Schmitt & Chang, 2016²⁸.

The regulation of lncRNA, as all other RNA, does not end at transcription. Instead, lncRNA and mRNA transcripts undergo an extensive series of modifications, including splicing, 3' polyadenylation, 5' methylguanosine capping (m7G) and potentially several other post-synthesis changes (i.e. modifications of the chemical composition of RNA)^{36,37}. Transcripts are exported to the cytoplasm or retained in the nucleus in a controlled manner^{38,39}.

In the second part of the Results, the main topic of this thesis (the lncRNA *NEAT1*, discussed below) intersects with the RNA degradation pathway. Therefore, as the biology described here focuses on processes in the nucleus, I will highlight, albeit non-exhaustively, some of the most relevant players in nuclear RNA turnover in the next section.

RNA turnover

The versatility of RNA degradation pathways

About 80% of the genome is transcribed, but only a small portion of the transcripts is stabilized and potentially functional^{21,40}. In fact, a common hypothesis warranting rapid degradation of the (non-functional) RNA products is that it is not necessarily the produced RNA that is important, but rather the act of transcription itself⁴¹. For instance, transcription may help to

shape and maintain chromatin architecture and interactions⁴², recruit regulatory factors^{43,44}, or provide flexibility for the cell to respond quickly to stress⁴⁵. We can imagine that just like for proteins, quality control of RNA is essential to prevent aberrant protein production and cell metabolism⁴⁶. Thus, a highly active RNA surveillance pathway is necessary for normal RNA homeostasis⁴⁷. Moreover, because of the multistep nature of mature RNA processing, the several non-essential RNA pieces (i.e. introns, sequences of pre/pri-miRNA, etc.) should be efficiently cut away and degraded and also this process relies on the RNA degradation machinery. In addition to the fact that RNA homeostasis is a highly energy- and nutrient demanding process, keeping around unwanted RNA products involves risk: their tightly regulated degradation is necessary to prevent the formation of aberrant structures such as R-loops, which can, if not resolved, inflict DNA damage during replication⁴⁸.

RNA degradation in the cell can occur both through endonucleolytic cleavage (i.e. inside the RNA sequence), or by exonucleolytic degradation starting from the RNA's 3'- or 5'- end⁴⁹. Early on, however, researchers found that there might *not* be a generalized endonuclease mechanism for RNA degradation in the nucleus⁵⁰ except for the more recently discovered nuclear Argonaute-mediated RISC activity as part of a nuclear RNA interference (RNAi) pathway^{51,52}. Technically, this pathway contributes to RNA turnover, but as this process is rather important for post-transcriptional gene regulation, I will not further discuss nuclear RNAi. In contrast, nuclear decay and turnover of RNAs is thought to involve primarily exonucleolytic activity⁵⁰.

5'-to-3' exonucleases

Several complexes contribute to the degradation and turnover of nuclear RNA. A first mechanism consists of degradation from the 5' start of the transcript towards the 3' end, or, can occur by chasing the RNA polymerase after cleavage of the polyA tail to aid transcriptional termination⁵⁰. Although not very well described for higher eukaryotes, it seems that, similar to the process in the cytoplasm, 5'-to-3' degradation can occur in several steps⁴⁹, including binding and decapping of the to-be-degraded transcript by the RNA chaperone complex Sm- and Sm-Like (Lsm)⁵³. In contrast, nuclear mRNAs are not necessarily deadenylated prior to degradation⁵⁴. 5'-to-3' hydrolytic cleavage is performed by XRN1/2-category proteins, which can both occur in the cytoplasm and the nucleus^{49,50}. The cleaved-off nucleotides are then converted to nucleosides and can potentially be recycled for reincorporation via the nucleoside salvage pathway⁵⁵. mRNAs that are stabilized and who should not be targeted for degradation are protected by the co-transcriptional addition of the 5' cap and 3' poly-A tail or other modifications such as m6A³⁷.

3'-to-5' exonucleases

The second important nuclear RNA degradation pathway is led by another complex called the RNA exosome. It does not only regulate transcript levels of mRNAs^{56,57}, but also degrades defective RNA species⁴⁷, intronic sequences⁵⁸, intronless RNA both from self- and viral origins^{59,60}, rRNA⁵⁶, etc. In addition, it degrades nascent 3' unprotected transcripts to prevent their unwanted accumulation or translation⁶¹ and ensures proper 3' end maturation of rRNAs, tRNAs, telomeric RNAs and small nuclear RNAs⁶². In mammalian cells, the RNA exosome consists of nine core subunits (EXO-9) and two catalytic parts (RRP6 and hDis3, both of which have 3'-to-5' activity)⁶³. The core components form a hexameric ring structure and a protein “cap”, to which the catalytic subunits and several cofactors are attached (figure 3A)⁶². hDis3 functions primarily to degrade RNA in the nucleoplasm, whereas RRP6 is predominantly active in the nucleoli⁶². In general, after an initial deadenylation step, the RNA is threaded through the central channel of the ring, where it can both be decapped and degraded^{64,65}.

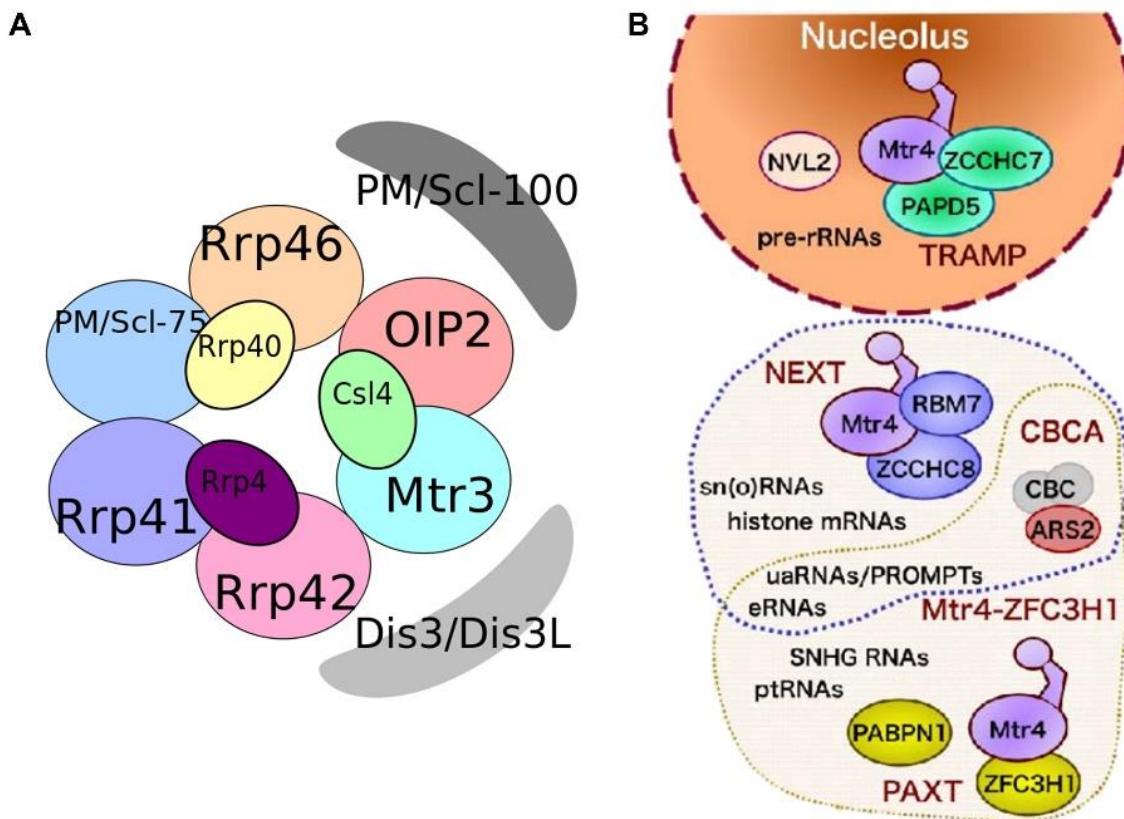


Figure 3: The human exosome. A. Human RNA exosome components, consisting of a six protein core ring and three S1 RNA binding domain containing proteins on top, which in turn associate with two catalytic subunits, Dis3/Dis3L and PM/Scl-100. Adapted from RNApathwaysDB⁶⁶; B. Human RNA exosome targeting complexes TRAMP (top), NEXT (middle) and PAXT (bottom) with their respective targets. Note that certain components and RNA targets are shared between NEXT and PAXT. Adapted from Ogami et al., 2018⁶².

Substrate specificity of the RNA exosome is established by the targeting of the different RNA species via specific binding to several RNA exosome-associated targeting complexes. These include NEXT (Nuclear Exosome Targeting)⁶⁷, PAXT (poly(A) tail exosome targeting)^{68,69}, and TRAMP-like (Trf4/Air2/Mtr4p polyadenylation) complexes^{67,70}, which, although they each target their own set of RNAs in the nucleus, share the RNA helicase Matrin 4⁵⁶ (figure 3B). The TRAMP-like complex functions in the cells by oligo-adenylation of its targets, usually rRNAs and snRNA^{71–73}, and is, in human cells, restricted to the nucleolus⁶⁹. On the other hand, the nucleoplasmic NEXT guides the degradation of many short-lived RNAs such as Promoter-Upstream Transcripts (PROMPTs)⁷⁴, enhancer RNAs⁶¹, and other nuclear transcripts⁵⁶, whereas PAXT assists in the exosome targeting of transcripts with long poly-A tails⁶⁹.

Physiological relevance of the RNA exosome

In yeast, most subunits of the RNA exosome are essential for growth and survival, indicating that this complex is extremely important to prevent aberrant cellular homeostasis⁶². In line with this, changing the expression levels and sequence of the mammalian exosome components can also impact on cellular physiology. For instance, mutations that disrupt hDIS3 endo-and exonuclease activity are synthetic lethal in HeLa cells⁷⁵ and disruption of other mammalian RNA exosome components causes premature differentiation, indicating that the RNA exosome may be important to maintain stemness, in particular in skin⁷⁶ and erythroid cells^{77,78}. Similar observations were made for Matrin 4, the shared RNA helicase of the different RNA exosome targeting complexes. Indeed, negative modulation of Matrin4 levels in stem- and progenitor cells caused mitotic defects, decreased viability, and premature differentiation^{79,80}. The RNA exosome and the NEXT complex are further thought to contribute to degradation of lowly expressed RNAs that change expression levels in this process⁸¹.

Furthermore, knockdown of RBM7, a central RNA binding component of the NEXT complex, sensitizes cells to chemotherapeutic drugs, suggesting an essential role for RBM7 targeting to maintain cellular homeostasis and the response to stress⁸². In fact, it was shown that the RNA exosome participates directly in the DNA damage response pathway by helping to degrade aberrant nucleic acid structures such as R-loops⁸³ and to promote activation-induced cytidine deaminase (AID) access to the to-be-recombined loci during class switch recombination in B lymphocytes⁸⁴. Importantly, mutations in the non-catalytic exosome component RRP40 cause spinal muscular dystrophy and hypoplasia in the pons and cerebellum⁸⁵, but, at least in Drosophila, loss if this core component does not contribute to the cell cycle defects observed in knockdown conditions of the catalytic subunit dRrp6⁸⁶. In conclusion, although more work is needed to clarify the roles of individual RNA exosome components in mammalian cells, it is clear that at least part of the subunits have essential roles in cellular homeostasis.

In the next section, we will return to lncRNA biology and discuss how lncRNAs can, just like the products of protein coding genes, become hijacked by cancer cells. We will consider their altered regulation in cancer with a special focus on the transcription factor p53, and examine their potential as cancer therapeutic targets. Finally, before moving on to the second chapter of this introduction, we will discuss how we can study lncRNAs in an *in vivo* setting relevant to human cancer biology.

LncRNA in cancer

LncRNAs are hijacked by cancer cells

Although *in vivo* phenotypes are described for some highly conserved and highly expressed lncRNAs (e.g. Kcnq1ot1⁸⁷, Fendrr⁸⁸, Tug1⁸⁹, Nest⁹⁰), many lncRNAs can be deleted in animal models with no striking consequences for normal tissue development and homeostasis⁹¹. Rather, they are in some cases suggested to become crucial only in distinct (patho-) physiological conditions⁹². Although they have diverse and broad-ranging functions that often define cell type-specificity and contribute to organismal complexity, their expression can be leveraged by the cell to fine-tune stress responses^{93–95}.

One potential stress the cell may have to deal with is the (replicative) stress associated with the continuous and uncontrolled division characteristic of cancer cells. LncRNAs contribute to many of the typical hallmarks of cancer including uncontrolled growth, evading programmed cell death and acquisition of invasion and metastasis potential⁹⁶. In these conditions, they are often differentially regulated in response to the changes taking place upon transformation. These changes in lncRNA expression levels can, for example, help to protect against cell death⁹⁷ or regulate levels of tumor suppressors by acting as competing endogenous RNAs (ceRNAs)⁹⁸. Furthermore, they can prevent the accumulation of DNA damage via direct interactions with the DNA damage repair machinery^{99–101}, or are produced at the damaged loci and might guide and ensure proper function of repair factors^{102,103}. The transformed cell can, just like with protein-coding genes, “hijack” the normal lncRNA function and use it to its own advantage for survival and growth²⁸.

P53 and downstream lncRNAs

Several cancer-related transcription factors induce non-coding transcription which can either prevent or stimulate the tumorigenic process¹⁰⁴. For instance, Myc, a transcription factor known to increase cell proliferation, targets several lncRNAs that influence gene regulation of cancer pathways including the cell cycle^{105,106} and tumor metabolism¹⁰⁷. On the other side of the spectrum we find p53, a master regulator of genomic integrity, who also transcriptionally regulates lncRNA, both in a context of cell survival and cell death¹⁰⁸. As this project started

with a search for novel lncRNAs targeted by p53, I will discuss in the following paragraphs some major aspects of p53 biology and its downstream (non-coding) targets.

Four decades of research into p53, the biology of the protein proves tremendously complex. In particular, several feedback loops, both positive and negative, fine-tune its levels and activity. For instance, the MDM2 E3 ligase targets the continuously produced p53 protein for degradation by the proteasome, but is also a downstream target of p53, as such contributing to a negative feedback loop¹⁰⁹. Thus, although p53 has a clear tumor suppressor function, its downstream targets may act as oncogenes, for instance by participating in these feedback loops common to the p53 transcriptional network¹⁰⁸.

As there is only a fine line in determining the outcomes of p53 activation, the p53 response needs to be tightly regulated¹¹⁰. When the damage inflicted on a cell can possibly be repaired, a cell needs to ensure its survival, but it also needs to halt its activities so the repair can occur without hurdles. The prime example of this pathway choice is mediated by the transcriptional activation of p21 (*CDKN1A*), which, when produced, arrests the cells in the Growth 1 (G1) phase of the cell cycle and allows for repair^{111,112}. Cells in which the damage is irreparable or in which too many lesions occur simultaneously might, in a normal physiological system, need to die, either via apoptosis or via other pathways¹¹³. A major example of the former is mediated by the p53 targets *BAX*, *NOXA* and *PUMA*, which, through their own cellular effectors, lead to Caspase cleavage and cell death^{114,115}.

Which pathway is chosen by p53 depends not only on the amount of damage, but also on the cellular context, the type of cell, the type of DNA stress, and the cofactors present to fine-tune p53 responses¹¹³ (figure 4). It was therefore not surprising that long non-coding RNAs, which naturally frequently display fine-tuning characteristics and cell-type specific expression, play a role in p53 biology that cannot be neglected^{104,108}. For instance, p53 has been shown to target several lncRNAs for transcriptional upregulation, but also can itself be regulated by them^{104,108}. This network is, anno 2019, well documented. For a detailed database of the p53-lncRNA network, see e.g. TP53LNC¹¹⁶ (<http://www.trihpph.net/TP53LNC/index.php>). Finally, the p53 downstream responses are modulated by lncRNAs as well, and they can help to determine the p53 pathway choice e.g. lincRNA-p21 regulates hundreds p53-target genes upon its transcriptional activation in trans¹¹⁷ or the p53-targeted lncRNA PURPL regulates its levels as part of a suppressive negative feedback loop¹¹⁸. Thus, it is clear that, just like for protein coding genes, p53 can give rise to lncRNAs that either promote or prevent cell survival, and thus its downstream lncRNA effectors can both have tumor suppressing and tumor promoting characteristics¹⁰⁸.

As a major tumor suppressor, p53 is inactivated in virtually all cancers. However, although an avenue extensively explored, it may in itself not be a great cancer therapeutic target. The main reason for this is because p53 is such an important protein to maintain homeostasis in virtually all cells of the body, and targeting p53 in a non-specific way has proven very toxic to normal cells¹¹⁹. Instead, targeting its downstream effectors may be a more interesting therapeutic strategy, as they are often either more cancer cell-specific or less detrimental to normal cells¹²⁰.

One of the options here, is targeting downstream lncRNA, which will be discussed below.

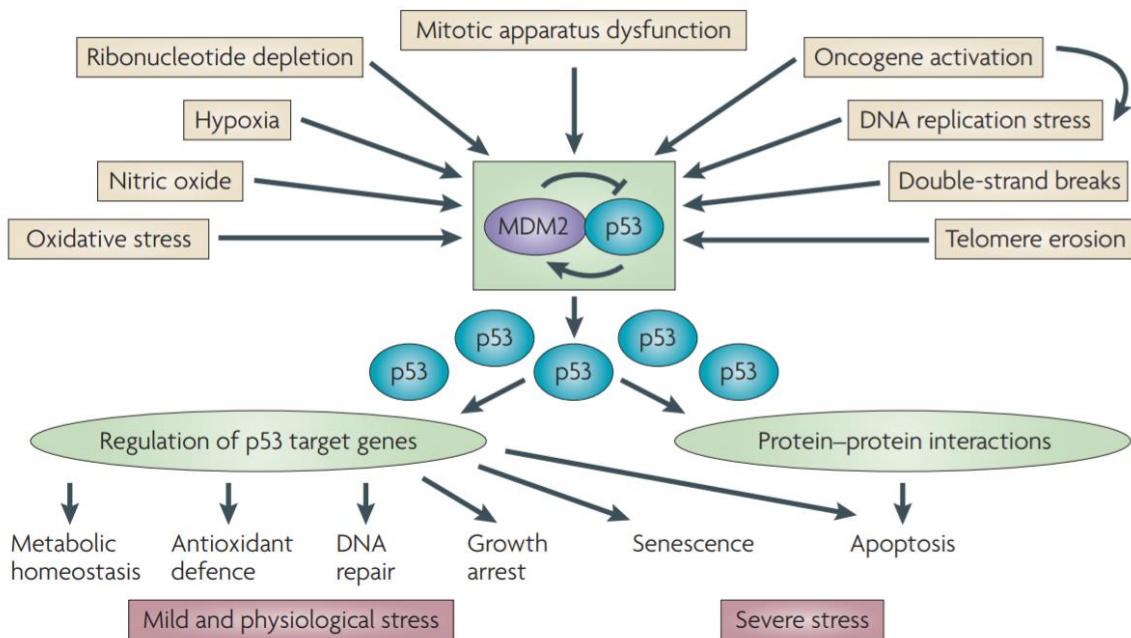


Figure 4: Simplified scheme of the p53 pathway. At the core of the pathway is the feedback interaction of p53 with its negative regulator and transcriptional target, MDM2. P53 in turn exerts its function in the cell (bottom) depending on the nature and the amount of stress inflicted (top). Adapted from Levine & Oren, 2009¹¹⁰.

LncRNAs as therapeutic targets in cancer

In the field of cancer therapy, a relatively recent development is the use of antisense oligonucleotides to modulate RNA^{121,122}, opening an avenue for non-coding RNAs to become therapeutic targets. In fact, as the sequence of RNA is easily known, the design and preclinical testing of a potential anticancer oligodrug are usually straightforward and rather inexpensive. The versatility that comes with targeting RNA contrasts greatly with the difficulty of design and cost of production for small molecules and therapeutic biologics against proteins^{123,124}.

Although cancer is broadly classified according to the tissue of origin and the major genetic alterations it carries, each tumor is its own unique disease with its own mechanism of transformation and growth. As conventional cancer therapy often fails or results in resistance

and for most cancer types no specific (“targeted”) drug strategies exist as of yet, the potential of modulating RNA may tremendously increase the number of anti-cancer targets¹²⁵.

There are several arguments to consider lncRNAs as potentially good candidates for cancer therapeutics development. First of all, lncRNAs have been shown to play important roles in tumor cell survival, metabolism, proliferation and metastasis capacity⁹⁶, and their removal often reduces cell viability or sensitizes cancer cells to existing therapies¹²⁵. Second, since their regulation is in many cases more cell type restricted, a potential therapy aimed against the lncRNA, rather than the proteins it might interact with, may increase the specificity of targeting and thus reduce unwanted side effects¹⁰⁴. Third, many tools exist to manipulate lncRNAs in vitro, which makes routine preclinical research relatively easy to set up. These reasons combined make that lncRNAs can be a great therapeutic target category, and indeed several are currently being developed and tested, at least in a preclinical setting, including in melanoma¹²⁶, prostate-¹²⁷, liver¹²⁸ and pancreatic cancer¹²⁹. An overview of oligo-based and other lncRNA targeting strategies can be found in figure 5.

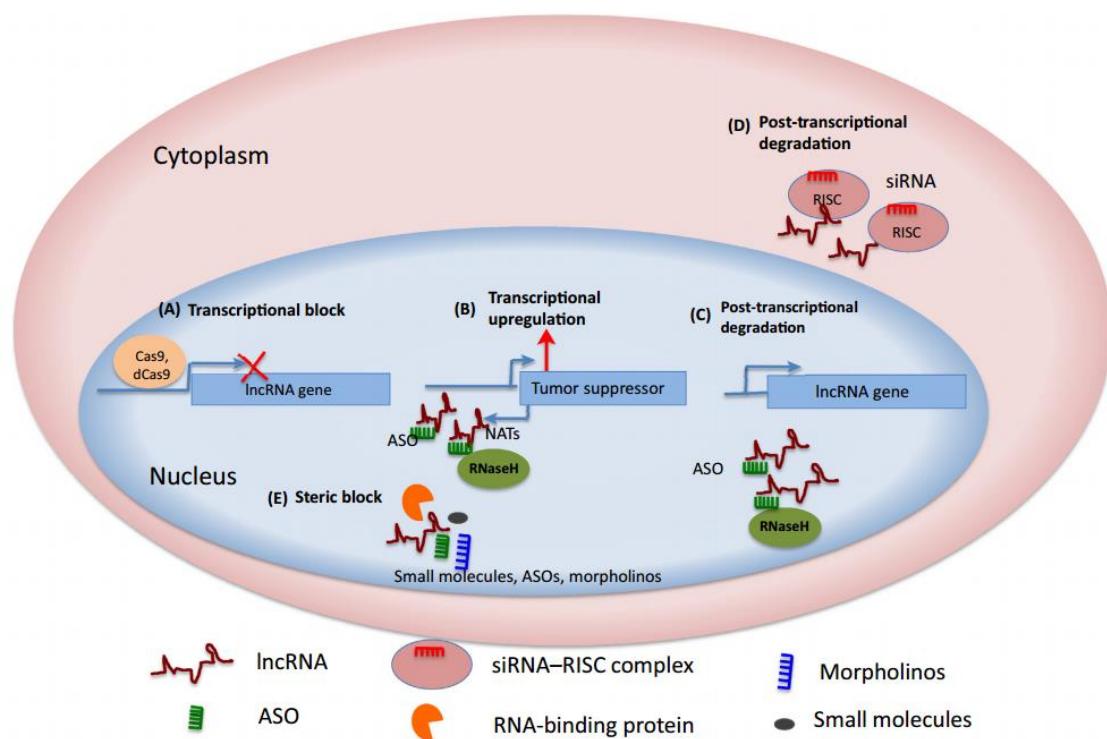


Figure 5: Oligo-based strategies to target lncRNA. Different strategies can be used to target lncRNA function in the cell, e.g. by blocking or stimulating their transcription (A&B), stimulating their degradation (C&D) or sterically blocking their function (E). Adapted from Arun et al., 2018¹²⁵.

It is however important to also consider several downsides to the use of nucleotide-based drugs, something that has contributed to the fact that despite early discovery of antisense therapy and many years of clinical development, only few have reached FDA approval^{121,122}.

Among these, the main problem for ASO-based therapy is their crossing of the cell membranes and their intracellular reaching of the target RNA¹²⁵, especially because most cells have important innate immune mechanisms to fight the entrance of foreign DNA and RNA. However, recent advances in ASO chemistry and delivery have significantly improved the targeting efficiency and paved the way for advanced oligodrug development¹²⁵.

Another potential problem consists of toxicity of the nucleotide based drug, which could emerge both from a generalized immune response, or from unwanted off-target effects, including on non-target RNA or in cells in which the drugs was not intended to work. A potential solution to the latter consists of local delivery; e.g. injection of the oligodrug into the tumor directly or topical application of a cream onto the skin¹³⁰. In addition, despite straightforward design of oligos based on target sequence, difficult-to-predict secondary and tertiary structures as well as the potential protection of the sequence by other intracellular components may cause reduced efficiency of targeting, and thus finding the best molecule for the target can remain a challenge. Finally, although lncRNAs are relatively easy to study in human cells *in vitro*, their lower conservation may complicate finding the right preclinical models *in vivo*¹²⁵. As these models, both in a fundamental (academic) research setting and for novel drug discovery in a preclinical (pharmaceutical) setting are crucial, they will be discussed below.

Mammalian *in vivo* cancer models to study lncRNA

Although several non-mammalian models to study lncRNA in cancer exist including in the fruitfly (*Drosophila melanogaster*) (e.g. lncRNA *cherub* is required for brain tumor development in *Drosophila melanogaster*¹³¹) or in Zebrafish (e.g. loss of human-cancer upregulated lncRNA THOR causes fertility defects in *Danio rerio*¹³²), I will focus here only on preclinical cancer models of lncRNA in the mouse (*Mus musculus*).

Genetic models

Genetically Engineered Mouse Models (GEMMs) have been extensively used to study the function of genes both in normal and cancer physiology. For instance, upon insertion of loxP sites at a specific locus in the mouse genome (e.g. flanking an exon or regulatory sequence) and crossing of these mice with a general or tissue-specific Cre-expressing mouse, the knockout or modification of the gene of interest can be obtained (figure 6).

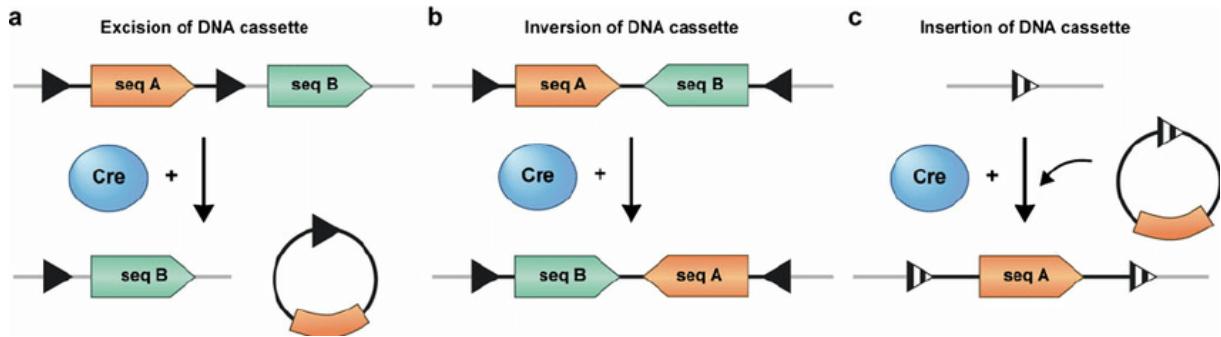


Figure 6: Different Cre-LoxP modifications of the mouse genome. DNA modified with LoxP sites can be excised (a), inverted (b), or used to insert another piece of DNA (c). Adapted from Renninger et al., 2011¹³³.

To model cancer formation, the expression of tumor suppressors and oncogenes can be modulated. For instance, the deletion of the tumor suppressor APC using a Lgr5+ specific Cre-line to target intestinal crypt stem cells causes intestinal cancer¹³⁴; or the expression of an epidermal compartment-specific mutated, constitutively active KRas (KRas^{G12D}) is used to generate squamous cell carcinomas¹³⁵. Often, these mice are intercrossed with floxed p53 (TP53^{flox/flox}) or TP53 null mice to generate fast, clinically relevant tumor models¹³⁶. Another strategy consists of putting oncogenes under the control of activating regulatory elements. For instance, Myc placed under the control of the IgH enhancer invariably leads to lymphoid tumors of B-cell origin¹³⁷ and is a popular model to study lymphoid- or leukemia cancers. Note that, since the advent of CRISPR editing of the genome, the generation of genetic alterations in mice has been greatly facilitated and accelerated¹³⁸.

In addition, when the lncRNA of interest is conserved, cancer GEMMs can be crossed with non-coding RNA knock out mouse models to find their contribution to tumor formation¹³⁹. Furthermore, non-cancer knockout models of lncRNAs can be used to study the contribution and therapeutic potential of these molecules when combined with radiation- or chemically induced oncogenesis systems, for instance in lung cancer¹⁴⁰, or cancer of the skin^{141,142}, in addition to the use of these models to elucidate their functions in normal physiology⁹¹.

Finally, GEMMs are not only valuable for the study of lncRNAs by e.g. profiling of their expression and modulation of their levels using the cellular RNAi machinery in established cancer mouse models, but can also be used, albeit only when the lncRNA is functionally conserved, as a preclinical model to study therapy responses¹²⁵. As a proof-of-concept example, researchers have previously used antisense oligonucleotides to discover the effect of mouse Malat1 on breast cancer growth in the MMTV (mouse mammary tumor virus)-PyMT¹⁴³. In conclusion, GEMM systems, either inducing cancer genetically or via exogenous stimuli in combination with genetically modulated levels of the lncRNA can be extremely valuable to find function and purpose for these molecules in a preclinical setting.

Non-genetic models

Partially to circumvent conservation issues and often to simplify experimental procedures, it is also possible to study cancer *in vivo* without modifying the host genome. These studies are usually based on grafts, and can either be syngeneic (mouse-to-mouse), or xenotransplantations (human-to-mouse). In the former, a cultured mouse cell line is transplanted into the tissue-of-origin in a host with the same genetic background as where the cells originate from. Due to the high degree of inbreeding and thus low degree of genetic variation of the mice within a same strain, this does not usually trigger immune responses and can serve as a valuable model to study mouse lncRNA in immunocompetent animals. For instance, lncRNA-HIT reduces primary tumor growth and metastasis in an orthotopic model of breast cancer induced by TGF- β ¹⁴⁴.

Apart from syngeneic cancer models, human cells can be transplanted into an immunocompromised mouse either subcutaneously or orthotopically. Here, both the classical transplantations of cell lines and more recently, also patient-derived xenografts can prove extremely useful to study and predict therapy response of the lncRNA therapy under study¹²⁵. For instance, targeting DANCR in a xenograft model was sufficient to induce considerable tumor shrinkage¹⁰⁶, or ASO-mediated knockdown of the lncRNA SAMMSON in melanoma¹²⁶ and HOTAIR in breast tumors can be highly effective anti-cancer strategies¹⁴⁵ and small interfering (si)RNA mediated knockdown of linc-Ceruloplasmin decreases growth and glycolysis in an orthotopic model of ovarian cancer¹⁴⁶.

In conclusion, a plethora of *in vivo* models exist to study the expression, roles and therapeutic potential of lncRNA in cancer. The fast evolution of novel model development and an increasingly large toolkit to study and target lncRNAs will help to expand our knowledge of these molecules, and help develop effective strategies to modulate their levels in cancer therapy¹²⁵.

Chapter II: NEAT1 and Paraspeckles

Paraspeckle structure, components, and definition

Paraspeckles: a nuclear domain built on long non-coding RNA

Paraspeckles were discovered in a proteomic screen for novel nucleolar components by the groups of A. Lamond and M. Mann in 2002¹⁴⁷. The newly identified nucleolus-associated components PSPC1, NONO/p54nrb and PSF/SFPQ, all belonging to the Drosophila Behavior, Human Splicing (DBHS) protein family, did however not functionally relate to the nucleolus itself. Rather, when fluorescently tagged and tracked, they formed a distinct body often found adjacent to but not overlapping with nuclear splicing speckles¹⁴⁸. The authors named this novel nuclear domain based on this spatial feature: the para (παρα, Gr., “beside, near”) -speckle. In unstressed cells, paraspeckle proteins, apart from broad nucleoplasmic localization^{12,149}, appear to shuttle continuously between paraspeckles and the nucleolus¹⁵⁰. In specific circumstances such as upon transcriptional inhibition, these proteins relocate more permanently (until the inhibitory factor is removed) to the perinucleolar caps^{148,150} (figure 7). Furthermore, the functionally redundant^{149,151} and biochemically interacting^{152,153} DBHS proteins localized to paraspeckles in an RNA-dependent manner¹⁵⁰. Indeed, paraspeckles are transcription-inhibitor and RNase sensitive¹⁴⁸, disassemble in the transcription-low telophase of the cell cycle and their main protein components all contain RNA Recognition Motifs (RRMs) essential for their localization to the nuclear bodies¹⁵⁰. Thus, it was hypothesized that paraspeckles also contain an important RNA component¹⁵⁰.

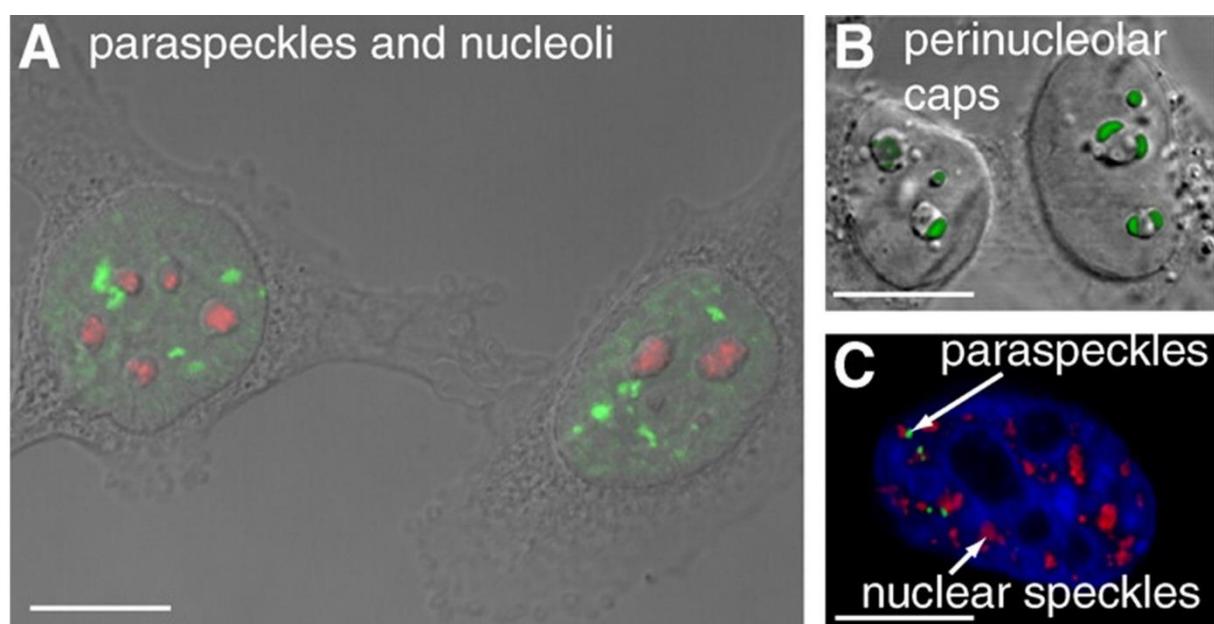


Figure 7: Paraspeckle proteins and nucleoli. A. Combined differential interference contrast and fluorescence micrograph of HeLa cells stained with anti-PSPC1 to show paraspeckles (green) as nucleoplasmic foci distinct from nucleoli (stained with B23 antibody; red). (B) HeLa

cells showing reorganization of the DBHS protein PSPC1 (green) to perinucleolar caps after treatment with actinomycin D to inhibit RNA Pol II transcription. (C) HeLa cell stained with anti-PSPC1 (green), anti-SC35 (red), and DAPI (blue) to show the relationship between paraspeckles abutting nuclear speckles in the interchromatin space. Adapted from Bond & Fox, 2009¹⁵⁴.

In 2009, three papers described how the long non-coding RNA (lncRNA) NEAT1 (also called MENε/β, referring to its localization in the Multiple Endocrine Neoplasia locus on chromosome 11 (11q13.1)) is the architectural backbone of the paraspeckle nuclear body both in mouse and in human cells^{12,155,156}. NEAT1 is a highly conserved, abundant mono-exonic lncRNA. Two isoforms are transcribed from the NEAT1 locus, demonstrating an identical sequence for the first 3765 nucleotides (3185 nucleotides in mouse), which is also the full genomic length of the short polyadenylated isoform NEAT1_1 (figure 8).

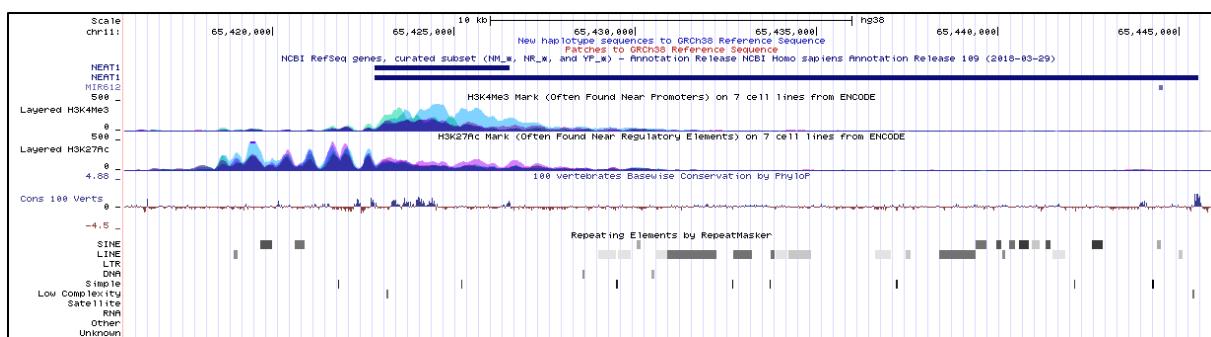


Figure 8: Basic features of the NEAT1 locus. Screenshot from UCSC genome browser hg38 assembly. From top to bottom: Scale indicating size in nucleotides; RefSeq genes; H3K4me3 mark in 4 cell lines (HSMM, HUVEC, NHEK, NHLF); H3K27Ac in these same lines; conservation in 100 vertebrates; Repeats.

NEAT1_1 is polyadenylated by the polyadenylation machinery, including through binding of CPSF6-NUDT21 to regulatory regions upstream of the canonical polyadenylation signal (PAS) AAUAAA at the NEAT1_1 RNA's 3' end¹⁴ and recruitment of the polyadenylate polymerase, in particular Star-polyadenylate polymerase (Star-PAP)¹⁵⁷ (figure 9A,B).

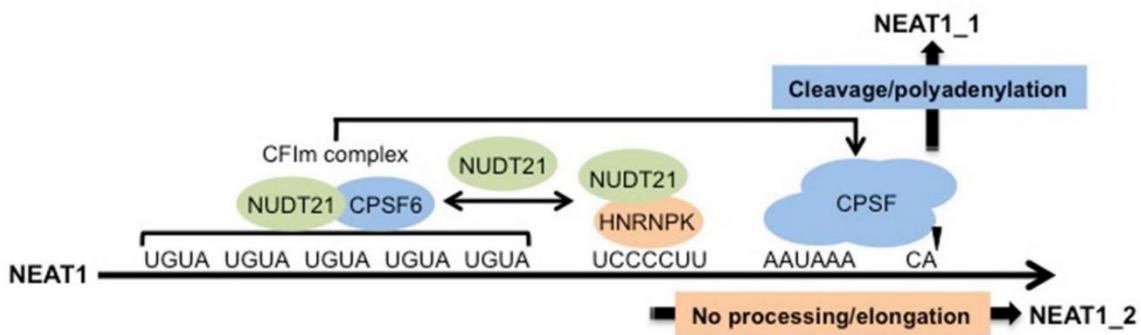
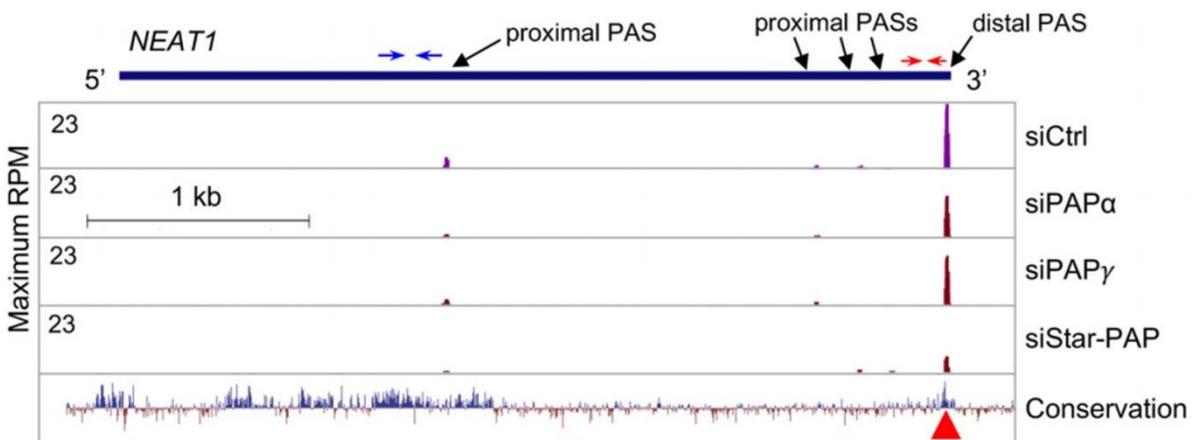
A**B**

Figure 9: 3' end processing of NEAT1_1. A. Factors involved in NEAT1_1 polyadenylation, including members of the Cleavage and Polyadenylation Family (CPSF6), NUDT21 and hnRNP K. B. NEAT1_1 is polyadenylated by Star-PAP; as evidenced by intact polyadenylation in knockdown conditions of the canonical PAP α and PAP β proteins, but loss of its polyadenylation upon STAR-PAP KD. Top panel (A) adapted from Yamazaki et al., 2015¹⁵⁸. Lower panel (B) adapted from Li et al., 2017¹⁵⁷.

The long, non-polyadenylated ~22.7 kb NEAT1_2 isoform is produced through masking of the 3'-end processing signals for the short via a mechanism involving hnRNP K, which competes for binding to NUDT21 with CPSF6¹⁴. To terminate NEAT1_2, a tRNA-like structure is generated by RNase P cleavage^{156,159} (figure 10). NEAT1_2 further forms a triple helix at its 3' end which has been implicated in stabilization of the transcript and protection from 3'-5' exonucleases^{159–161}. In mouse, both isoforms display an exceptionally high turnover (*Neat1_2*: half-life of only +/- 63 minutes; *Neat1_1*: half-life of only +/- 19.6 minutes), which may indicate stress-responsive functions¹⁶². In contrast, in humans both NEAT1_1 and NEAT1_2 are degraded at a significantly slower pace (>8 hours)¹⁶², suggesting that, despite relatively high levels of conservation, the mechanisms of RNA turnover and function in these organisms might differ.

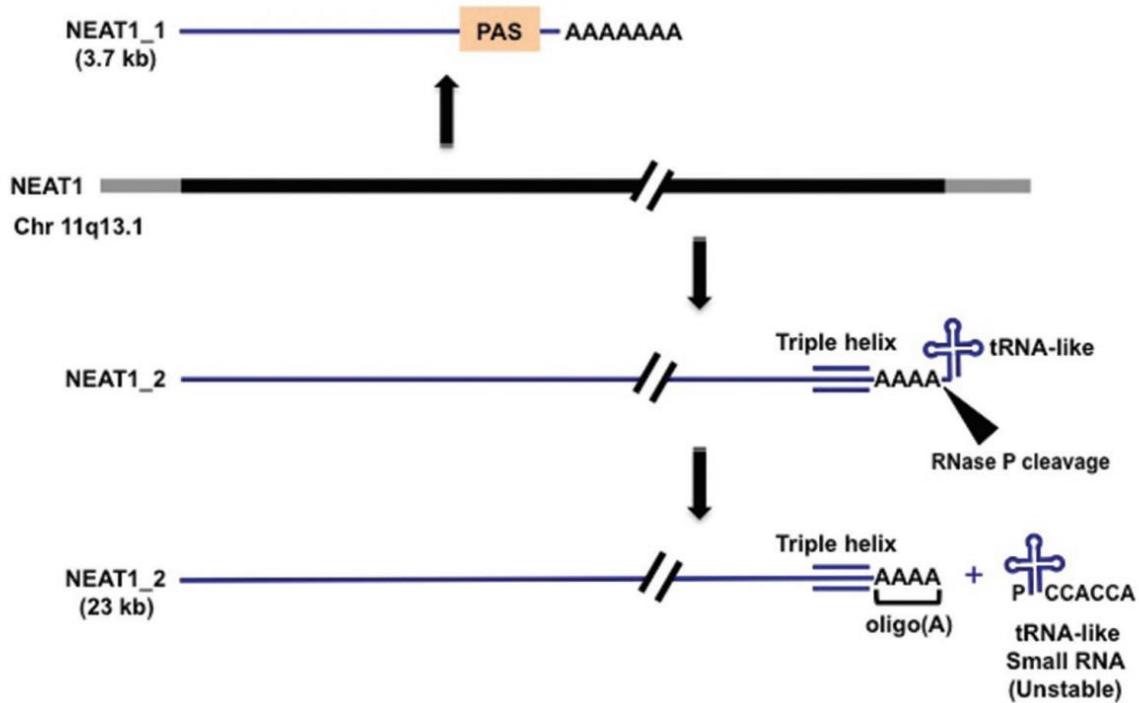


Figure 10: 3' processing of NEAT1_2. NEAT1_2 forms triple-helix and tRNA-like structures at its 3' end. The tRNA-like structure is cleaved by RNaseP to obtain fully matured NEAT1_2 and a highly-unstable tRNA-like small RNA. Adapted from Yamazaki et al., 2015¹⁵⁸.

NEAT1 and paraspeckles: a complicated relationship

Overexpression and correlation studies, particularly during differentiation, have suggested that the more abundant NEAT1_1 isoform is recruited into paraspeckles^{13,163}, but only NEAT1_2 is required as the architectural backbone of the nuclear body^{14,15,161,164}. In support, embryonic and undifferentiated cells in culture expressing only the short isoform do not display paraspeckles^{15,156,165}, and overexpression of NEAT1_1 alone in these cells is not sufficient to trigger their formation¹⁵. Inversely, mouse tissues lacking NEAT1_2 do not form detectable paraspeckle protein foci^{165,166}, and paraspeckle integrity is not affected by deletion of NEAT1_1 through deletion of the polyadenylation domain^{161,167}. As the bodies assemble co-transcriptionally, paraspeckles require active transcription of the long isoform, an observation strengthened by active transcription being required for its maintenance¹⁶². Upon reaching a certain size, presumably by the combined accumulation of transcripts and paraspeckle proteins, paraspeckles “bud off” to become multiple smaller units via a mechanism that involves liquid-liquid phase separation (see below)^{15,156}. Furthermore, the essential paraspeckle proteins PSF/SFPQ and NONO/p54nrb associate preferentially with the long isoform of NEAT1^{155,161} in the paraspeckle core¹⁶³. More recently, it was shown that NEAT1_1 transcripts can occur outside of paraspeckles even when the nuclear body is present¹⁶⁷. However, no function so far has been proposed for these ‘microspeckles’, consisting of NEAT1_1 transcripts and a potentially unknown protein component¹⁶⁷.

Since *NEAT1_1* and the 5' end of *NEAT1_2* are experimentally indistinguishable, there is currently no compelling evidence supporting the abovementioned recruitment hypothesis: any experimental approach detecting the short isoform also detects the first 3756 nucleotides of the long. In addition, the overlap between the two isoforms has hampered the dissection of isoform-specific roles and their individual contributions to paraspeckle biology. For instance, assays probing the proteomic composition, interacting RNA and genomic localization of *NEAT1* transcripts via biochemical pulldowns using Capture Hybridization Analysis of RNA Targets (CHART) followed by DNA/RNA sequencing and mass spectrometry concluded that paraspeckles are enriched at transcriptional start sites, interact with AG-rich RNAs and harbor over 200 nuclear proteins^{163,168}. However, the RNA region assayed by CHART was within the shared sequence of the two isoforms, therefore, it remains unclear whether these observations pertain to paraspeckles via *NEAT1_2* binding or reflect the behavior of *NEAT1_1*, independently of the nuclear body¹⁶⁷. Similarly, in most reported RNA-FISH or RT-qPCR experiments throughout the *NEAT1/PS* literature, the used primers or probes target the overlapping region of *NEAT1* and thus, as the individual isoforms are not assayed, uncertainty remains on whether the observations are *NEAT1* short isoform- or paraspeckle related. Independent measurements of *NEAT1* short transcripts are further hindered by the fact that although the long isoform is not polyadenylated, it contains a 22nt long A-rich stretch toward its 3' end (figure 11). Thus, cDNA preparation using oligo-dT primers does not reliably omit reverse transcription of the long isoform, which, in theory, could have been leveraged experimentally to address *NEAT1_1* and *NEAT1_2* independent functions.



Figure 11: NEAT1_2 oligo-A stretch. *NEAT1_2* is targeted, albeit maybe not reliably, by poly-dT based experimental approaches because of an A-rich stretch at its 3' end.

NEAT1_1 and NEAT1_2 are distinct entities in the cell

The expression patterns of the short and the long isoforms in adult normal individuals are very dissimilar in different tissues. For example in mice, paraspeckles, as probed by RNA-FISH for *Neat1_2*, seem to appear only upon physiological or pathological stress conditions^{141,165,169,170}, or are present in only a very distinct subset of cells within a given tissue such as the terminal lining of the gut^{165,169,170}. In addition, perturbation experiments affecting uniquely the long isoform are sufficient to induce phenotypes¹⁴¹. This idea is further bolstered by the low estimated number of *NEAT1_1* molecules (in HeLa, 6,5 molecules) per paraspeckle¹⁷¹. These observations point to potentially independent roles for paraspeckle nuclear bodies and *NEAT1* short isoform transcripts¹⁴¹ and warrant a more careful dissection of individual contributions to

the different phenotypes observed both in mice and human cells¹⁰. For example, the levels of individual *NEAT1* isoforms can now be manipulated with CRISPR-Cas9 technology^{161,167}, providing valuable insights into *NEAT1* and paraspeckle structure and regulation.

Paraspeckle protein phase separation: a framework to explain its nucleation?

Recent evidence has prompted an understanding of the nucleus as an assembly of physically separated mixtures of proteins and nucleic acids in which different functions are structurally separated through liquid-liquid demixing of its constituents¹⁷². Phase separation of nuclear bodies occurs via the low-complexity, disordered domains of their protein components (prion-like domains) through an accumulation of polar, positively charged amino acids², often potentiated by and affecting the nucleic acids present in the cell¹⁷³. Paraspeckles belong to this category of membraneless, functionally isolated organelles within the nuclear space¹⁰. A model for their formation can be found in the analysis of their structure combined with the physics of intracellular hyper-crowding of the paraspeckle molecular players¹⁰.

The localization and topology of the protein and RNA components within paraspeckles has been described in great detail. For instance, immunogold electron microscopy experiments showed that the 5' and 3' ends of the long *NEAT1* isoform were located on the outer side, whereas the middle part of the transcript was found internally^{13,161,163}. In normal culturing conditions, each paraspeckle is estimated to contain approximately 50 *NEAT1_2* molecules¹⁷¹. Furthermore, also the localization of other paraspeckle proteins is specific. For example, superresolution microscopy determined that the DBHS proteins NONO/p54rb, SFPQ/PFS and PSPC1 are all found on the inside (the ‘core’) of the nuclear body, whereas TDP-43 can be found in the shell, and the SWI/SNF factor BRG1, which has been shown to associate with paraspeckles¹⁷⁴, together with RBM14, exists in patches on its outside¹⁶³.

Combined with size calculations, the data on *NEAT1* topology has led to a U-shaped model of the transcripts within the nuclear body¹³. RNA secondary structure analyses have further corroborated these observations: psoralen analysis of RNA interactions and structures (PARIS) found that *NEAT1* displays a relatively limited secondary structure dominated by small hairpins¹⁷⁵ and SHAPE-analysis showed that the 3' and 5' ends of the long transcript could potentially interact through long-range basepairing¹⁷⁶.

Paraspeckles: a working definition.

In summary, paraspeckles are currently most commonly defined experimentally by either *NEAT1* or paraspeckle protein accumulations. Most studies have probed the overlapping part of the *NEAT1* isoforms (figure 7, 9, 10), or used *NEAT1* transcripts and paraspeckle proteins interchangeably to determine paraspeckle function. Given that *NEAT1_1* is found at least also outside of the paraspeckle nuclear bodies¹⁶⁷, if not existing as a completely distinct nuclear

compartment (i.e. microspeckles), most of the RNA-focused studies need to be interpreted with caution. Although it is clear that the interactions of *NEAT1* RNA with its protein binding factors leads to the nuclear body defined as the paraspeckle¹⁰, much of the structural and functional information may need experimental verification in light of this new concept.

Thus, here, we will define paraspeckles clearly and unequivocally, consistent with the definition proposed by Fox. et al. (2018)¹⁰:

“[The paraspeckle is] a nuclear body in which one of the essential paraspeckle proteins co-localizes with the longer isoform of NEAT1”.

Proposed functions and paraspeckle mechanisms of action

In the past ~16 years, numerous functions have been proposed for the nuclear body based on observations in the mouse, in cellular models, and in human tumor tissue. They can broadly be divided into two main categories: RNA retention and protein sequestration (figure 12). Below, we will discuss some of these functions in light of the knowledge that paraspeckle proteins and the *NEAT1* RNA may also have independent roles, and highlight how some of the data point toward potentially different functions for differently composed PS-like nuclear bodies in the cell.

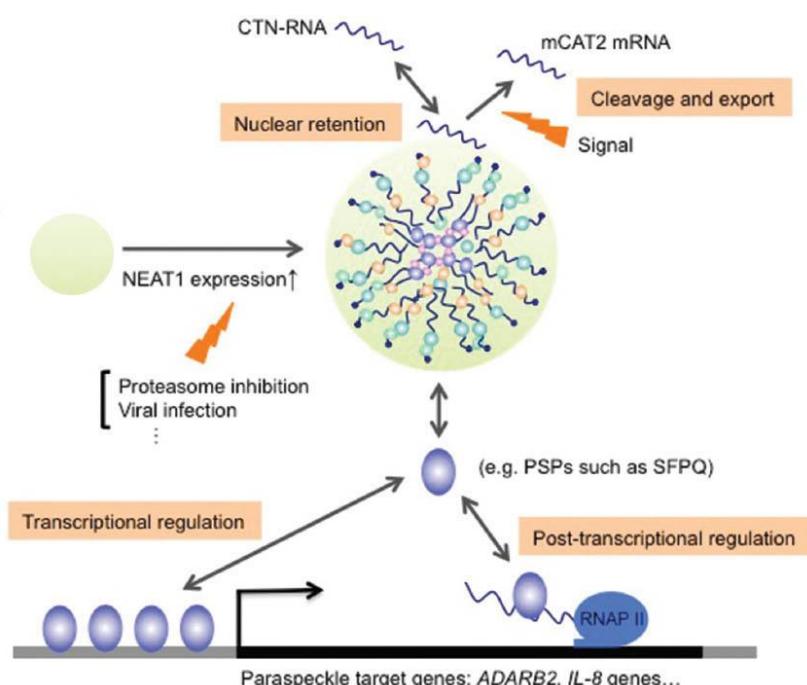


Figure 12: Models for paraspeckle function. Paraspeckles are thought to retain RNA in the nucleus, aid in RNA processing and modulate (post-) transcriptional regulation. Adapted from Yamazaki et al., 2015¹⁵⁸.

RNA retention

A role for RNA retention in PS was initially suggested by the finding that paraspeckles retained inverted-repeat containing A-to-I edited mRNAs when they appeared during differentiation, and supported by the fact that all major PS proteins contained RNA binding domains (RRMs)¹⁷⁷. As such, they were suggested to facilitate a post-transcriptional layer of rapid gene regulation mediated by RNA-editing^{15,177}. In example, as paraspeckles were thought to be stress-related nuclear bodies, they would retain Ctn-RNA in the nucleus. When warranted in response to stress; the A-to-I edited 3' UTR of Ctn-RNA would be post-transcriptionally cleaved, giving rise to the coding mRNA mCAT, its export to the cytoplasm, and translation¹⁷⁸. Comparably, *NEAT1*, upregulated upon mitochondrial stress via ATF2, modulates the response to this stress by retaining nuclear encoded mRNAs of mitochondrial proteins¹⁷⁹. Furthermore, retention functions of paraspeckles have been proposed in circadian rhythm regulation¹⁸⁰, adipogenesis¹⁸¹, and hypoxia¹⁸², as well as the response to viral infection either through retention of human-encoded transcripts¹⁸³ or potentially even viral transcripts¹⁸⁴. In an RNA-pulldown study, *NEAT1* preferentially associated with AG-rich RNA sequences from both specific introns and spliced mRNAs at the outer ‘shell’-like region of the paraspeckles¹⁶³.

While paraspeckles are enriched for RNA-binding proteins, the protein components of paraspeckles rather than the paraspeckle body itself appear necessary for the nuclear retention of RNA¹⁴⁹. For instance, the abovementioned Ctn-RNA, which is commonly used as an example in paraspeckle RNA retention studies, remains nuclear and forms residual foci in the absence of *Neat1*, relocates to perinucleolar caps as do the members of the DBHS family¹⁸⁵ and is associated with the PS-mimicking nuclear bodies built on phosphorothioate-ASOs in cells which lack *NEAT1*¹⁸⁶. Paraspeckle proteins further form phase-separated nuclear bodies with other RNAs, including the G₄C₂-repeat-containing aberrant C9ORF72 in motor neurons¹⁸⁷. In addition, several other observations point to a scaffolding role for RNA in paraspeckle biogenesis irrespective of sequence. For instance, phosphorothioate oligonucleotides can form paraspeckle-like foci with the paraspeckle proteins in the absence of *NEAT1* in cultured cells¹⁸⁶, and in marsupials, an orthologous *NEAT1* with very little sequence conservation nucleates paraspeckle-like foci *in cellulo*¹⁸⁸.

More generally, the RNA recognition motifs of the DBHS family broadly bind RNA, and their particular preference for specific RNA species, including A-to-I edited mRNAs, remains unclear¹⁴⁹. These observations strongly suggest that retention is not a paraspeckle-specific mechanism, but rather may be the result of a local interaction of RNA-binding proteins

independently of *NEAT1* RNA¹⁸⁵. In summary, although these studies report a common interaction and retention mechanism of RNA by paraspeckle proteins, a distinct role for the nuclear body per se and not just of its ubiquitous proteinaceous components remains poorly defined.

Protein sequestration

In addition to RNA retention, paraspeckles have been proposed to dynamically regulate protein availability of their components in the nucleoplasm and at the DNA in response to stress^{168,189,190}. For instance, proteasome inhibition enlarges them, resulting in decreased levels of SFPQ and NONO in the rest of nucleus¹⁹⁰, and they regulate SFPQ localization to immune responsive genes upon viral infection¹⁸⁹. However, paraspeckle components also bind other nuclear factors that may regulate their nucleoplasmic concentrations and activities. For instance, *NEAT1* nucleates the non-homologous end joining factors Ku70, Ku80 and DNAPKcs and the transcription regulation factor HEXIM1 to form an alternative paraspeckle-like nuclear body, the HEXIM1-DNA-PK-paraspeckle components-ribonucleoprotein complex (HDP-RNP), which serves as a platform to assemble the components necessary for the response to the presence of foreign DNA in the cell¹⁹¹. Similarly, the paraspeckle component NONO regulates Cyclic GMP-AMP synthase (cGAS)-regulated immune activation by recognizing the viral capsid proteins of HIV¹⁹².

Some components of the microprocessor complex, involved in the generation of miRNAs, associate with *NEAT1* and several paraspeckle proteins¹⁹³ which may explain the frequent interactions found between *NEAT1* and various miRNAs¹⁹⁴. *NEAT1* is also associated with AGO1, another protein of the endogenous RNA interference machinery, which is suggested to contribute to the maintenance of 3D chromatin architecture¹⁹⁵. Furthermore, *NEAT1* is enriched at the DNA around transcription start sites¹⁶⁸ and binds to chromatin associated factors such as WDR5¹⁹⁶, EZH2¹⁹⁷, BRG1^{163,174} and CDC5L¹⁹⁸, providing sequestration away from or guidance to their DNA targets. Finally, paraspeckle proteins interact with nuclear import and export factors during differentiation and development^{181,199}, and it was hypothesized that their presence in the paraspeckles may modulate their availability for other processes in the cell.

Conclusion

In summary, *NEAT1* and paraspeckle biology prove tremendously complex, and although many studies elucidate structural aspects and potential mechanisms for their function, it appears from the literature that still a lot can be learned. In the following parts of this thesis, I will line up the objectives of this work, and present the data that I generated, together with my colleagues and collaborators, since the start of this project in our lab. Note that an important part leading to the results in this thesis has already been presented by Dr. Laura Standaert,

with whom I, at least for first part of my work, have tightly collaborated. Part of these results can be found in her Doctoral Thesis (Laura Standaert, 2015: The functional relevance of lncRNA downstream of p53²⁰⁰).

Part II: Objectives

The work described in this thesis emerged from a previous screen in our lab (described in Results Chapter I), in which we searched for novel long non-coding RNA targets of the tumor suppressing transcription factor p53 by ChIP-seq and RNA-seq. With the work presented here, we hope to provide an advancement in our understanding of the transcriptional network downstream of p53, as well as a more detailed account of one of these downstream targets, *NEAT1*, which is a lncRNA identified as a target induced upon Nutlin-3a treatment.

We chose to follow up on *NEAT1* for several reasons, both practical and biological. First, *NEAT1* is the architectural component of paraspeckles and it is highly evolutionarily conserved. Second, a *Neat1* KO mouse model existed, and, just like the KO for *Trp53* itself, this mouse model was viable. Being a p53 target gene, we could formulate the main hypothesis of this work, that *NEAT1* is a modulator of a stress response downstream of p53. Third, although highly studied, the role of *NEAT1* and its exact mechanism of action in the nucleus remained elusive, in particular in the context of cell growth, cancer, and survival. Because this RNA is so abundant, and because it partakes in the formation of a nuclear body, we considered this an interesting target, shaping the work in this thesis.

For the first part of this work, we formulated the following aims:

1. Study the role of *NEAT1* downstream of p53 in *in vitro* and *in vivo* cancer models.
2. Understand its contribution to disease both in human and mouse.

As many, if not all of the phenotypes we described for loss of both *NEAT1* isoforms could be phenocopied by the loss of the long isoform *NEAT1_2* and paraspeckles alone, we decided to follow up on isoform-specific roles in the second part of this thesis. Thus, we defined the following aims:

3. Understand the dynamic regulation of *NEAT1* isoforms upon treatment with different p53-activating agents.
4. Dissect the individual contributions of each isoform to the observed phenotypes.

As long non-coding RNAs remain underdogs often disregarded in the field of cancer research, this study also aims to highlight the importance of non-coding RNA as major molecular players in the cell, and emphasizes the importance of research on specific RNA isoforms to discern functional transcripts from potential transcriptional “junk”. We hope to make a case for the idea that an eye for detail when designing experiments with regard to individual functions of the players in a multi-component system is of utmost importance.

Part III: Results

Chapter I

p53 induces formation of *NEAT1* lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity.

Carmen Adriaens^{1,2*}, Laura Standaert^{1,2*}, Jasmine Barra^{1,2}, Mathilde Latil³, Annelien Verfaillie⁴, Peter Kaleb⁵, Bram Boeckx^{6,7}, Paul Wijnhoven⁸, Enrico Radaelli⁹, William Vermi^{10,11}, Eleonora Leucci^{1,2}, Gaëlle Lapouge³, Benjamin Beck³, Joost van den Oord¹², Shinichi Nakagawa¹³, Tetsuro Hirose¹⁴, Anna Sablina⁵, Diether Lambrechts^{6,7}, Stein Aerts⁴, Cédric Blanpain³ & Jean-Christophe Marine^{1,2,#}

¹Laboratory for Molecular Cancer Biology, Center for the Biology of disease, VIB, 3000 Leuven, Belgium;

²Laboratory for Molecular Cancer Biology, Center of Human Genetics, KU Leuven, 3000 Leuven, Belgium;

³Université Libre de Bruxelles, IRIBHM, Brussels B-1070, Belgium; ⁴Laboratory of Computational Biology, Center for Human Genetics, University of Leuven, Belgium; ⁵Laboratory for Mechanisms of Cell transformation, Center for the Biology of disease, VIB, 3000 Leuven, Belgium;

⁶Vesalius Research Center, VIB, 3000 Leuven, Belgium; ⁷Laboratory for Translational Genetics, Department of Oncology, KULeuven, Belgium; ⁸The Wellcome Trust/Cancer Research UK Gurdon, Gurdon Institute, University of Cambridge, Cambridge, UK; ⁹Mouse histopathology core facility, Center for the biology of disease, VIB - KU Leuven, Belgium; ¹⁰Department of Molecular and Translational medicine, Section of Pathology, University of Brescia, Italy; ¹¹Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA; ¹²Laboratory of Translational Cell and Tissue Research, Department of Pathology, KU Leuven and UZ Leuven, Leuven, Belgium;

¹³RNA Biology Laboratory, Riken, Japan; ¹⁴Institute for Genetic Medicine; Hokkaido University; Sapporo, Japan. [#]jeanchristophe.marine@kuleuven.vib.be

*Equally contributed to this work

Abstract

In a search for mediators of the p53 tumor suppressor pathway, which induces pleiotropic and often antagonistic cellular responses, we identified the long noncoding RNA (lncRNA) *NEAT1*. *NEAT1* is an essential architectural component of paraspeckle nuclear bodies, whose pathophysiological relevance remains unclear. Activation of p53, pharmacologically or by oncogene-induced replication stress, stimulated the formation of paraspeckles in mouse and human cells. Silencing *Neat1* expression in mice, which prevents paraspeckle formation, sensitized preneoplastic cells to DNA-damage-induced cell death and impaired skin tumorigenesis. We provide mechanistic evidence that *NEAT1* promotes ATR signaling in response to replication stress and is thereby engaged in a negative feedback loop that attenuates oncogene-dependent activation of p53. *NEAT1* targeting in established human cancer cell lines induced synthetic lethality with genotoxic chemotherapeutics, including PARP inhibitors, and nongenotoxic activation of p53. This study establishes a key genetic link between *NEAT1* paraspeckles, p53 biology and tumorigenesis and identifies *NEAT1* as a promising target to enhance sensitivity of cancer cells to both chemotherapy and p53 reactivation therapy.

Introduction

The p53 tumor suppressor is inactivated in most cancers through mutations in the *TP53* gene or via alternative mechanisms, such as overexpression of mouse double minute 2 (MDM2) or MDM4, which promote p53 degradation and/or inactivation of function²⁰¹. Tumors are addicted to loss of p53 function²⁰², and pharmacological reactivation of p53 has become a potential therapeutic strategy. The observation that Nutlin-3a, a small-molecule inhibitor of the p53–MDM2 interaction, triggers p53 activation and suppresses tumor growth in a *TP53* wild-type (WT) osteosarcoma murine xenograft model²⁰³ paved the way to the development of more potent and selective small-molecule inhibitors of MDM2, some of which are now in clinical development²⁰⁴.

However, the pleiotropic effects of p53 complicate p53 reactivation strategies. p53 triggers antagonistic cellular responses such as apoptotic cell death, senescence, reversible cell cycle arrest, DNA repair and autophagy. In contrast to activation of apoptosis, which helps to eliminate defective cells, induction of cell cycle arrest and the associated DNA repair program leads to cell survival²⁰⁵. The molecular events that initiate p53-dependent cell cycle arrest (life) or apoptotic (death) transcriptional programs remain elusive. As MDM2 antagonists induce mainly G1 cell cycle arrest and apoptosis or senescence only rarely²⁰⁶, a clear understanding of the mechanisms that control cell fate determination by p53 may inform development of therapies to kill cancer cells more efficiently.

A putative contribution of the noncoding portion of the human genome in cell fate decision downstream of p53 has been largely underexplored. The noncoding genome is transcribed in vast numbers of RNA species, including lncRNAs²⁰. Many lncRNAs have been suggested to regulate tumorigenesis on the basis of transfection studies and/or correlative expression analyses, although *Xist* (encoding X-inactive specific transcript) and *Malat1* (encoding metastasis-associated lung adenocarcinoma transcript 1) are the only two lncRNAs for which a clear genetic link with tumorigenesis has been established^{143,207}. Relevant to p53, several lncRNAs, including *NEAT1* (encoding nuclear paraspeckle assembly transcript 1), have already been identified as p53 targets^{117,208–210}. Here we provide genetic evidence that *NEAT1* is engaged in a negative feedback loop with p53 and thereby modulates cancer formation in mice by dampening oncogene-dependent activation of p53. Consistent with this finding, *NEAT1* targeting sensitized established human cancer cells to both chemotherapy and p53 reactivation therapy. These data further illustrate the relevance of lncRNAs in tumorigenesis and their potential as therapeutic targets.

Results

***NEAT1* is a p53 target gene**

To gain a comprehensive view of the p53 downstream signaling network, we established the transcriptome (RNA-seq) and genome-wide occupancy of p53 (chromatin immunoprecipitation sequencing (ChIP-seq)) in TP53 WT MCF-7 breast cancer cells exposed to Nutlin-3a to stabilize p53 (ref. ²¹¹). We identified 110 and 147 annotated lncRNAs that were downregulated and upregulated, respectively, in cells treated with Nutlin-3a ($|\log_2 \text{fold change (FC)}| > 1$, adjusted $P < 0.05$). To determine which of these transcripts are bona fide p53 targets, we searched for p53-binding sites within 10 kb upstream and downstream of their transcription start sites (TSS) by adapting the i-cisTarget method²¹². The p53-responsive element was the most significantly enriched motif within the upregulated genes, whereas it was undetectable among the downregulated loci. Among 73 loci in which at least one p53-binding motif was identified, 16 exhibited p53 ChIP-seq peak(s) within 10 kb of their TSS (Supplementary Table 1 at the end of this chapter and Supplementary Fig. 1). We curated this list for reliability of the annotation and identified five lncRNA genes likely to be direct p53 targets (Supplementary Table 1 and Supplementary Fig. 1). One of these, *NEAT1*, was previously reported to be induced by p53 (ref. ²¹⁰) and is required for the assembly of subnuclear bodies known as paraspeckles¹². Indeed, we observed a p53 ChIP-seq peak upstream of the *NEAT1* promoter and significant ($P = 5.314 \times 10^{-11}$) induction of the 3.7-kb polyadenylated *NEAT1* transcript (*NEAT1_1*) in Nutlin-3a-treated cells (Fig. 1a and Supplementary Fig. 2a), indicating that *NEAT1* is a direct p53 target gene.

Figure 1

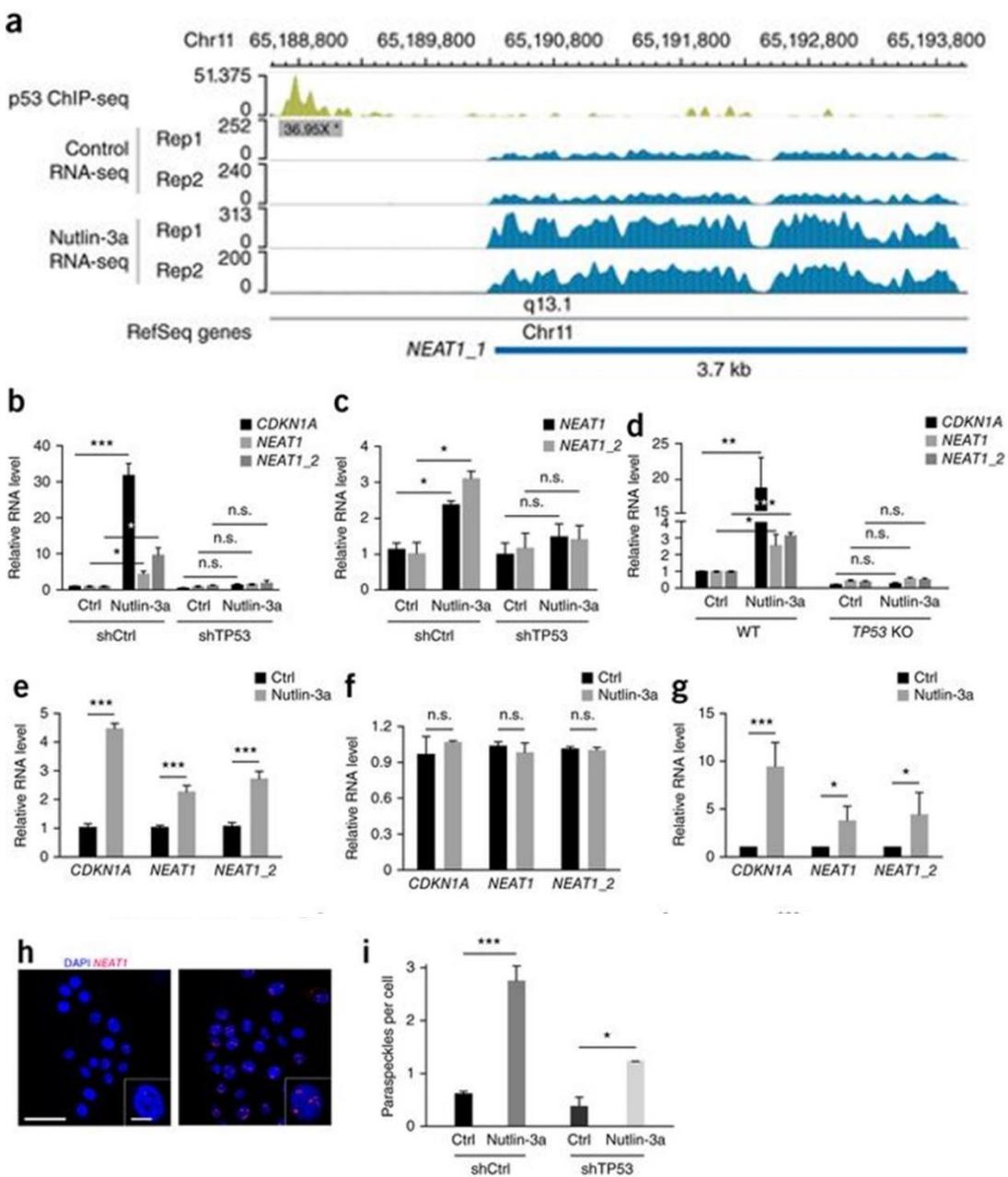


Figure 1. p53 induces NEAT1 expression and paraspeckle formation. (a) NEAT1 locus showing the p53 ChIP-seq signal (green) and the RNA-seq signal (blue) for control (untreated) or Nutlin-3a-stimulated MCF-7 cells (24 h) in two biological replicates (rep1 and rep2). Track heights represent raw read depth. (*) Significant fold change compared to input. (b–g) Relative RNA levels in shCtrl and shTP53 MCF-7 ($n = 3$; b) and NGP neuroblastoma cells ($n = 2$; c), WT and TP53 KO isogenic colon carcinoma HCT116 cells ($n = 5$; d), TP53 WT bone osteosarcoma cell line U2OS ($n = 6$; e), a TP53 mutant short-term melanoma culture ($n = 3$; f) and human embryonic stem cells (hES; $n = 4$; g) exposed to Nutlin-3a (5 μ M) for 24 h. Values are normalized to the levels of RNA expression in non-treated cells, which was set to 1. Error bars represent s.e.m. (h) Confocal images of NEAT1 RNA-FISH (red) with nuclear DAPI counterstain (blue) in vehicle treated (left panel) or MCF-7 cells exposed to Nutlin-3a (right panel). Scale bar, 50 μ m; inset 12.5 μ m. (i) Quantification of RNA-FISH in shCtrl and shTP53 MCF-7 cells. Error bar, s.d. of two biological replicates. Statistical significance was calculated using unpaired two-sided t-tests (b, c, e, f, g, h) or Mann–Whitney U test (d) with *, $p < 0.05$; **, $p < 0.01$ and ***, $p < 0.001$.

Supplementary Figure 1

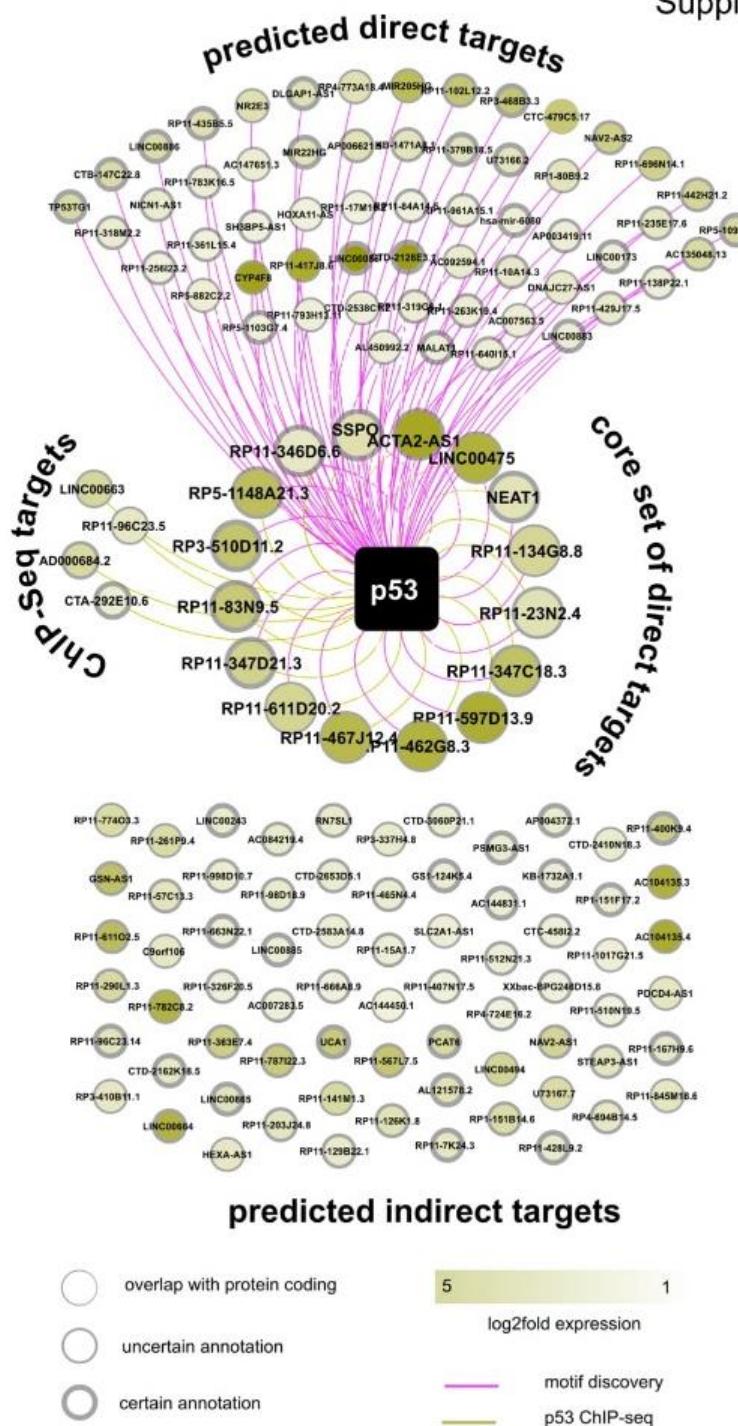
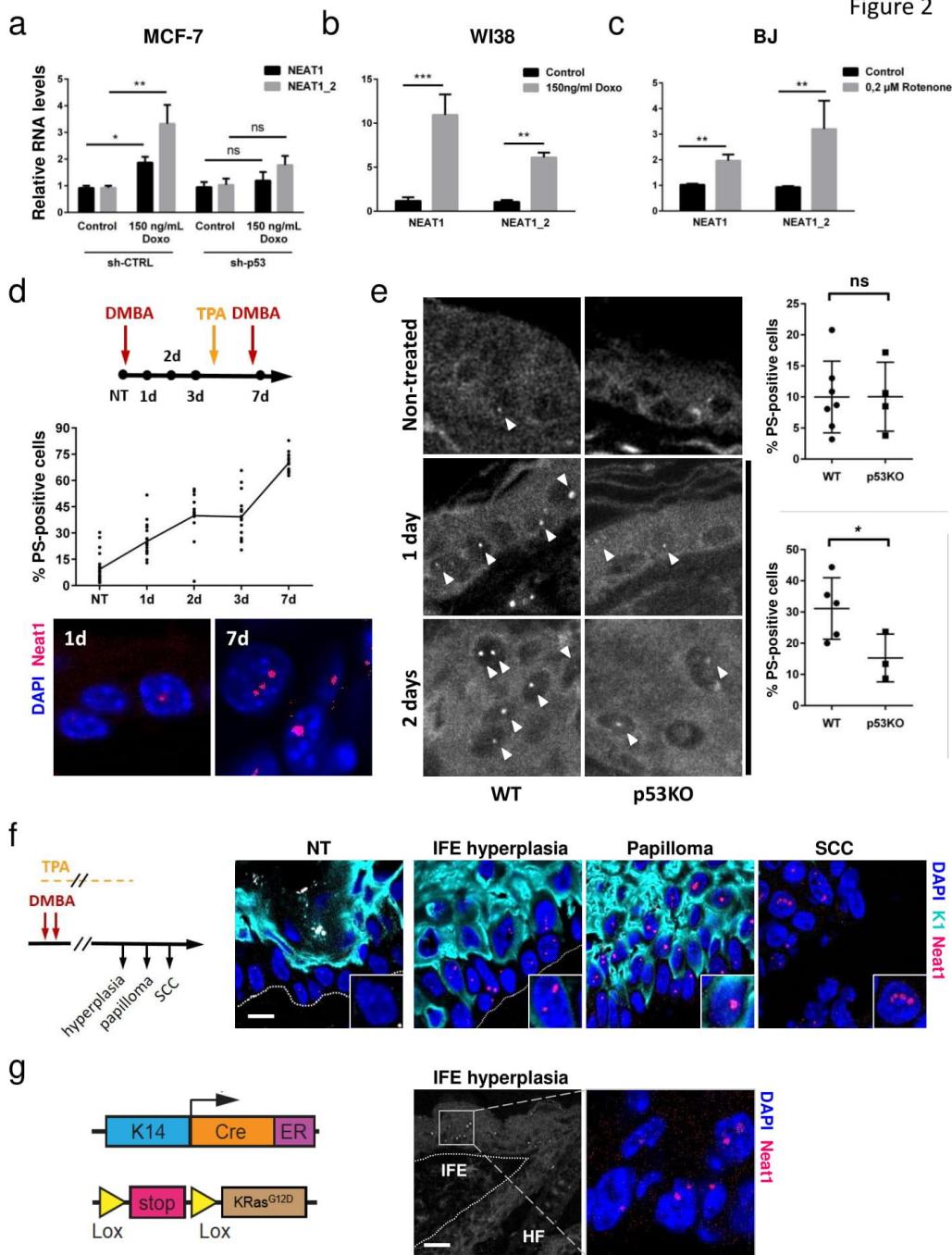


Figure S1. A network of lncRNAs downstream of p53. A regulatory network of 147 upregulated putative lncRNAs in MCF-7 cells exposed to Nutlin-3a. Targets are presented in a gradient of yellow according to the log₂ fold change expression values. Edges show targets predicted based on motif discovery (pink) or p53 ChIP-peaks (yellow) within a search space of 10 kb around the TSS. Node borders represent the reliability of the predicted lncRNAs based on a manual curation of their annotation. Candidate lncRNAs either have a confident annotation (thick), uncertain annotation (medium) or overlap with protein coding genes (thin).

In addition to the 3.7-kb transcript, *NEAT1* produces a long (22.7 kb) isoform (*NEAT1_2*). This transcript is not polyadenylated¹⁵⁹ (Supplementary Fig. 2b) and, accordingly, was not detected in our poly(A)+ RNA-seq data sets. Although both *NEAT1_1* and *NEAT1_2* are present in paraspeckles, their formation is strictly dependent on *NEAT1_2* (refs. ^{12,155,156,177}). To assess whether *NEAT1_2* is also induced by p53, we quantified its expression by RT-qPCR in cells exposed to Nutlin-3a. We generated cDNA using random hexamers (instead of oligo-dTs) to amplify both *NEAT1* isoforms. Because these overlap completely at the 5' end, primer pair A detects both isoforms, whereas primer pair B detects *NEAT1_2* specifically (Supplementary Fig. 2b). Both total *NEAT1* and *NEAT1_2* were induced in human TP53 WT MCF-7, NGP, HCT116 and U2OS cells but not in MCF-7 or NGP cells expressing small hairpin RNAs (shRNAs) targeting TP53 (shTP53), TP53 knockout (KO) HCT116 cells or TP53 mutant MM118 cancer cells (Fig. 1b–f). Induction of the p53-target gene *CDKN1A* confirmed activation of p53 in all *TP53* WT Nutlin-3a-treated cells (Fig. 1b–f). Expression of *NEAT1* and *NEAT1_2* was also induced in normal human embryonic cells (hES), immortalized fibroblasts (BJ) and low-passage mouse embryonic fibroblasts (MEFs) exposed to Nutlin-3a (Fig. 1g and Supplementary Fig. 2c,d).

Figure 2. DNA damage induces NEAT1 paraspeckle formation. (a–c) RT–qPCR analysis in shCtrl and shTP53 MCF–7 cells ($n = 4$) (a) and diploid fibroblasts WI38 ($n = 4$) (b) exposed to Doxo (24 h) and in skin fibroblast BJ cells ($n = 5$) exposed to Rotenone (Rot, 24 h) (c). Values are normalized to the levels of expression in non–treated cells, which was set to 1. Error bars represent s.e.m. Statistical significance was calculated using an unpaired two-sided t-test with *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. (d–e) Representative images and quantification of Neat1 RNA–FISH with DAPI counterstain in the skin of mice WT and Trp53 KO exposed to DMBA (and TPA). Quantification of percentage of paraspeckle–positive cells (% PS) is described in the Methods section. Error bars represent s.d. Statistical significance was determined using an unpaired two-sided t-test with *, $p < 0.05$. Scale bar, 10 μm . (e) White arrows indicate paraspeckles. (f–g) RNA–FISH of Neat1 (red) with immunostaining for cytokeratin 1 (K1, cyan), and nuclear counterstain (DAPI, blue) in mouse skin of mice exposed to DMBA and TPA (f) or expressing oncogenic Kras (g). (f) Scale bar, 10 μm ; inset, 5 μm . (g) Scale bar, 40 μm ; inset, 10 μm . IFE, interfollicular epidermis; SCC, squamous cell carcinoma. HF, hair follicle.

Figure 2



p53 induces paraspokele formation

Assembly of paraspokeles, which is dependent on *NEAT1_2* expression, is thought to widely affect gene expression via distinct mechanisms¹⁵⁸. Because p53 stimulates *NEAT1_2* expression, we asked whether p53 can promote paraspokele formation using *NEAT1_2* RNA–fluorescence *in situ* hybridization (RNA–FISH). We observed a marked increase in the *NEAT1_2* RNA–FISH signal, which was distributed in a characteristic punctate pattern, in MCF-7 cells after treatment with Nutlin-3a. This increase was attenuated in MCF-7 cells expressing TP53-targeting shRNAs (Fig. 1h,i) but was not

completely abrogated, as silencing of *TP53* was not complete (data not shown). Likewise, human diploid fibroblasts (BJ cells) and *TP53* WT HCT116 cells showed increased numbers of paraspeckles upon Nutlin-3a exposure, but BJ cells expressing sh*TP53* and isogenic *TP53* KO HCT116 cells did not (Supplementary Fig. 2e,f). We conclude that pharmacological activation of p53 stimulates paraspeckle formation.

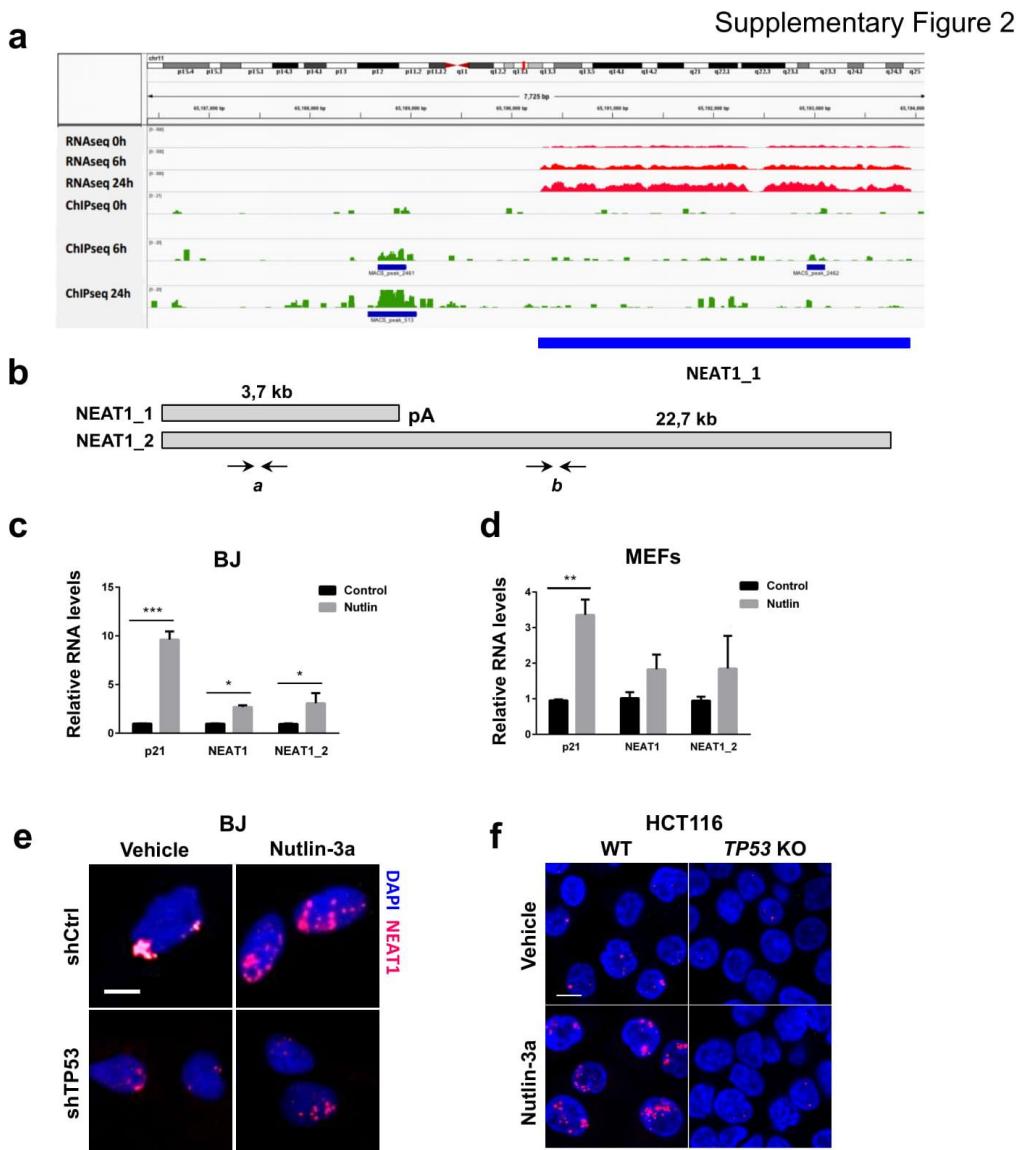


Figure S2. NEAT1 is a direct p53 target gene. (a) The NEAT1 locus with p53 ChIP-seq (green) and RNA-seq (red) tracks from MCF-7 cells exposed to Nutlin-3a for 6 and 24 h. (b) Schematic representation of the NEAT1 transcripts. The primer pairs used to detect both ("a", NEAT1) and the long form specifically ("b", NEAT1_2) by RT-qPCR are indicated by the arrows. (c-d) Relative RNA levels of NEAT1 and the p53 target gene CDKN1A (p21) in normal human diploid fibroblast BJ cells and in MEFs exposed to 5 μM Nutlin-3a for 24 h. Values are normalized to levels in untreated cells, which are set to 1. Error bars represent s.e.m. n = 3 and 5. (e-f) NEAT1 RNA-FISH (red) with nuclear DAPI counterstain (blue) in BJ (e) and HCT116 (f) cells exposed to Nutlin-3a. Scale bars represents 10 μm.

Oncogenic stimuli induce NEAT1 paraspeckle formation *in vivo*

p53 is induced by a variety of stress signals, many of which activate, either directly or indirectly, the DNA-damage response (DDR)²¹³. Such signals include exposure to the DNA-damaging agent doxorubicin; increases in reactive oxygen species (ROS) owing to treatment with the mitochondrial uncoupler rotenone or passaging in nonphysiological O₂levels; and telomere attrition upon long-term culturing. We observed induction of *NEAT1* and *NEAT1_2* expression in all these experimental conditions, in both cancer cell lines (i.e., MCF-7) and immortalized diploid fibroblasts (WI38 and BJ cells) and this induction was accompanied by an increase in paraspeckle formation (Fig. 2a–c and Supplementary Fig. 3a–c). *NEAT1* and *NEAT1_2* were induced in parental MCF-7 cells exposed to doxorubicin, but this increase was attenuated in cells expressing shTP53 (Fig. 2a). Hence, expression of *NEAT1* and *NEAT1_2* is induced upon DNA damage, and this induction is at least partly p53 dependent.

The specific physiological conditions under which paraspeckles form *in vivo* are still unclear. We assessed whether paraspeckles are assembled in the epidermis of mice exposed to DMBA, a carcinogen known to induce p53. Whereas paraspeckles were barely detectable in the skin of untreated mice, we detected them after one dose of DMBA in WT (Fig. 2d,e) but not *Neat1* KO mice (Supplementary Fig. 3d), and the number of paraspeckle-positive cells increased after the second DMBA application (Fig. 2d). Mice were also exposed to the proinflammatory agent TPA at day 4. However, TPA exposure alone did not induce paraspeckle assembly (Supplementary Fig. 3e), indicating that the latter is a consequence of DMBA-induced genotoxic stress and not TPA-induced inflammation. To test whether DMBA-induced paraspeckle formation depends on the presence of functional p53, we exposed *Trp53* KO mice to the DMBA–TPA protocol. The number and intensity of the *Neat1* foci were substantially higher in *Trp53* WT than in *Trp53* KO tissues (Fig. 2e). Genotoxic stress, therefore, induces *Neat1* expression and paraspeckle formation *in vivo* in a p53-dependent manner.

Oncogene activation also induces p53 by engaging the ARF–MDM2 pathway and a DNA replication stress response²¹³. In the two-stage DMBA and TPA carcinogenesis protocol, after the initial and transient DMBA-induced DDR, TPA administration results in a progressive selection of cells carrying oncogenic *Ras* mutations²¹⁴. Paraspeckle-positive cells were absent from untreated epidermis but readily detectable in hyperplastic interfollicular epidermis 1 week after DMBA–TPA administration and in differentiated K1-positive cells from benign skin tumors (papillomas) (Fig. 2f). Paraspeckle-positive cells were also observed in roughly half of the lesions that progressed into malignant squamous cell carcinoma (SCC) (Fig. 2f).

To rule out the possibility that paraspeckle formation is specific to carcinogen-induced tumorigenesis, we measured paraspeckles in a genetically induced skin cancer mouse model¹³⁵. We crossed mice bearing an oncogenic *Kras*^{G12D} allele silenced by an upstream *loxP*-flanked (floxed) transcriptional stop cassette (*Kras*^{LSL-G12D}) with mice bearing a tamoxifen-inducible Cre allele driven by the keratin 14 (*K14*) promoter. Analysis of skin from TAM-exposed *Kras*^{G12D} knock-in, but not wild-type (WT), mice showed the presence of paraspeckles in skin hyperplasias and papillomas (Fig. 2g and Supplementary Fig. 3f). Paraspeckle assembly is therefore induced in response to oncogene activation and is maintained as tumors are established and progress.

***Neat1* promotes skin tumor formation**

The presence of paraspeckles during the early stages of carcinogenesis prompted us to determine whether genetic inactivation of *Neat1* modulates tumor formation. We used previously described mice in which the *Neat1* locus is silenced by insertion of a transcriptional stop cassette upstream of its TSS¹⁶⁵ (Supplementary Fig. 4a). We exposed these *Neat1* KO mice, which are viable and healthy¹⁶⁵, to DMBA and TPA and quantified benign papillomas and invasive SCC (Fig. 3). Fewer and smaller papillomas were observed before the onset of SCC formation in *Neat1* KO and heterozygous mice compared with WT controls (Fig. 3a–c). When more than 50% of the WT mice had developed invasive skin lesions, none of the *Neat1* KO mice showed SCC (Fig. 3d–f), indicating that *Neat1* is required for tumor progression. *Neat1* haploinsufficiency also decreased the number and size of DMBA- and TPA-induced SCC (Fig. 3d–f). DMBA–TPA-induced paraspeckle formation was abolished in *Neat1* KO mice and compromised in *Neat1* heterozygous mice (Fig. 3h).

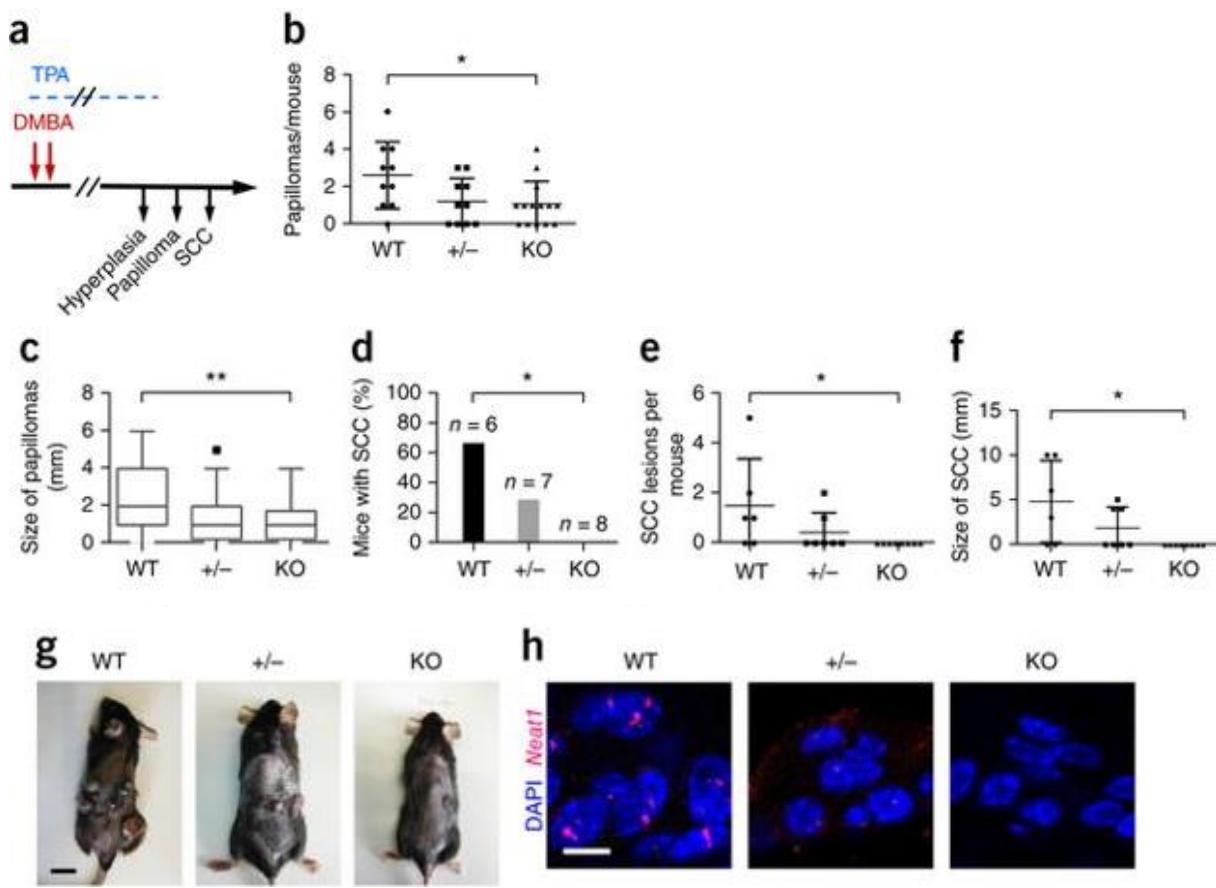


Figure 3. Neat1 KO mice are resistant to chemically-induced skin cancer formation. (a) DMBA-TPA treatment scheme. (b,c) Number (b) and size (c) of papillomas induced upon DMBA and TPA treatment in WT ($n = 10$), Neat1 heterozygous (+/−; $n = 10$) and Neat1 KO ($n = 14$) mice before onset of SCC. * $P < 0.05$; ** $P < 0.01$, unpaired two-sample two-sided t-test. Error bars represent mean ± s.d. (b); box boundaries represent 25th and 75th percentiles; center line, median; whiskers, last data point within ±1.5 interquartile range (c). (d) Percentage of mice with SCCs, evaluated when >50% of WT mice had developed SCCs (on average, 29 weeks). * $P < 0.05$, chi-squared test. Average number (e) and size (f) of SCC lesions, evaluated as in d. Data represent mean ± s.d. * $P < 0.05$, unpaired two-sided t-test. Data points represent individual mice ($n = 6$ (WT), 7 (+/−) or 8 (KO)). (g) Representative images of WT, Neat1 KO and Neat1 heterozygous mice at the time point described in d. Scale bar, 1 cm. (h) Representative Neat1 RNA-FISH with nuclear DAPI counterstain in interfollicular epidermis of DMBA- and TPA-treated mice. Scale bar, 10 μm.

Supplementary Figure 3

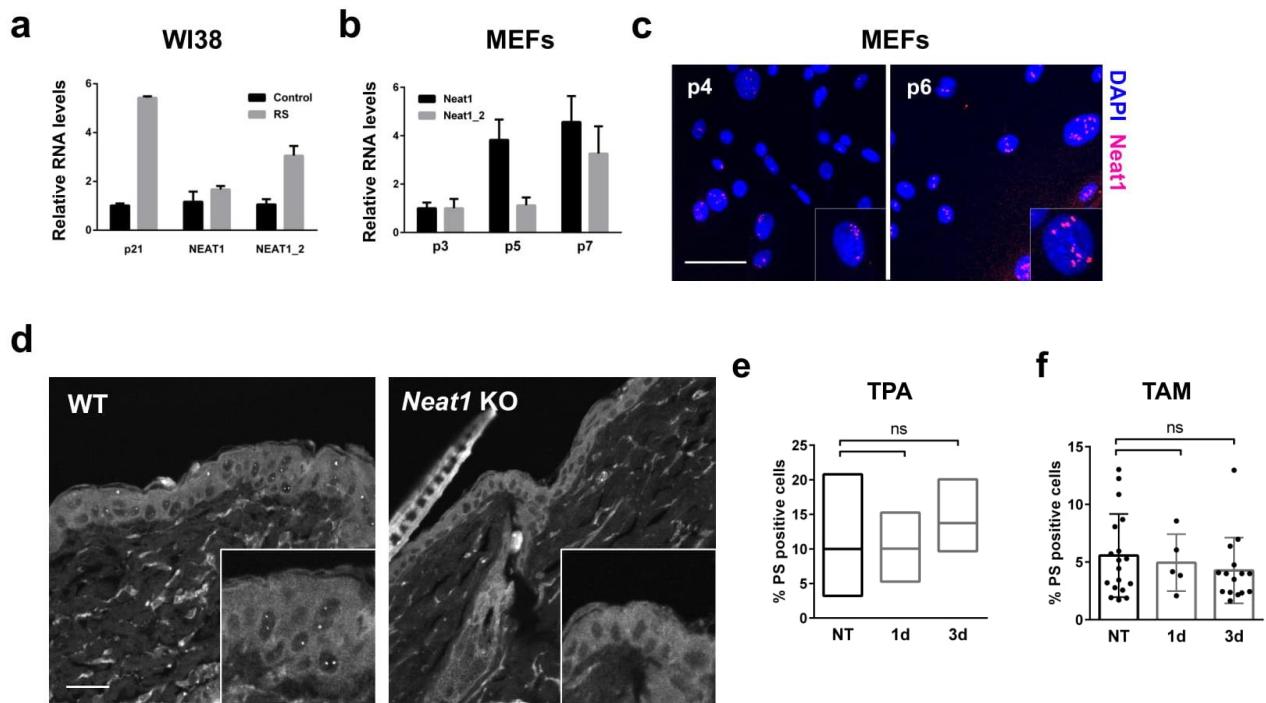


Figure S3. DNA damage induces NEAT1 expression and paraspeckle formation. (a-b) RT-qPCR for NEAT1/Neat1 and CDKN1A (p21) levels (as positive control for p53 upregulation/stabilization) in early and/or late passaged WI38 cells and MEFs. Values are normalized to levels in untreated cells, which are set to 1. Error bars represent s.e.m. of three technical replicates. (c) Paraspeckle numbers increase upon continued passaging of MEFs. Scale bar, 25 μm. (d) Paraspeckles are detected in DMBA-treated WT but not in Neat1 KO animals. Representative images at 48h post-DMBA exposure. Scale bar, 50 μm. (e) Quantification of paraspeckle- (PS-) positive cells in mouse interfollicular epidermis upon TPA exposure for 1 and 3 days. (f) Quantification of paraspeckle-positive cells in mouse interfollicular epidermis upon 5-OH-tamoxifen (TAM) exposure. For (e) and (f), NT, non-treated, n = 7; 1 d, n = 6; 3 d, n = 5 mice.

Neat1 prevents accumulation of DNA damage

WT mice treated with DMBA–TPA showed diffuse hyperplasia on day 7 and severe hyperplasia on day 11, after two applications of DMBA. In contrast, *Neat1* KO mice showed no hyperplasia on day 7 and only mild focal hyperplasia on day 11 (Fig. 4a,b). Keratinocyte proliferation was lower after DMBA exposure in *Neat1* KO mice than in WT mice (Fig. 4c,d). In addition, we observed a transient apoptotic response in keratinocytes of WT mice after the first DMBA application until the first TPA administration (Fig. 4a). In contrast, apoptosis was sustained for up to 11 d after the start of the treatment in *Neat1* KO mice (Fig. 4e). Differences in cell proliferation and viability were significant only after the second DMBA application (Supplementary Fig. 4b–e).

p53 stabilization after DMBA administration was more pronounced and persistent in *Neat1* KO mice than in WT mice (Fig. 4f). Similarly, the number of cells expressing the DNA

damage marker histone 2A.X phosphorylated at S139 (γ -H2A.X) decreased in WT mice 2 to 3 d after DMBA exposure but continued to increase in *Neat1* KO mice 3, 7 and 11 d after treatment (Fig. 4g). These results indicate that *Neat1* prevents accumulation of excessive DNA damage and concomitant activation of p53 in cells after DMBA-induced oncogenic stress.

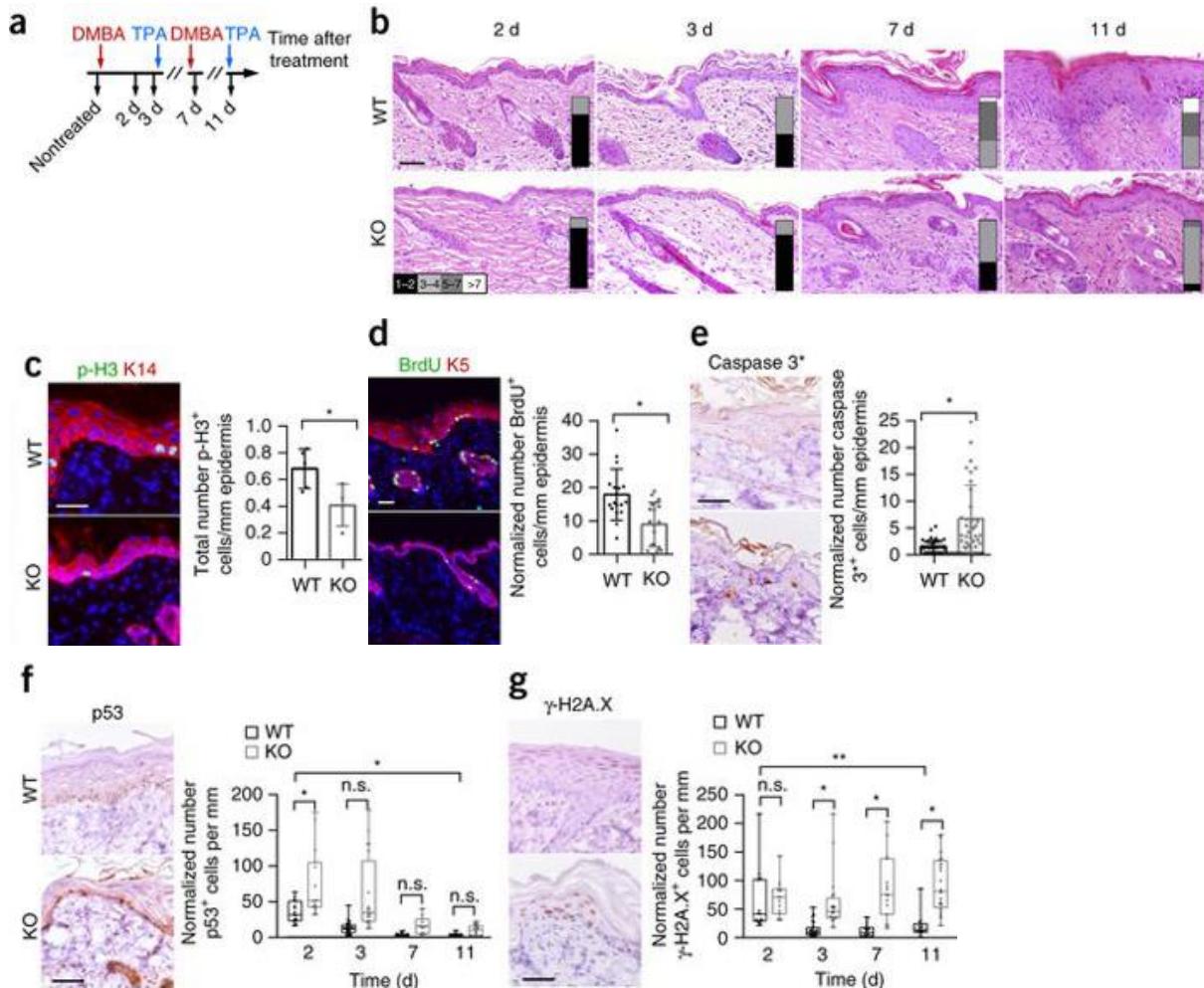


Figure 4. *Neat1* prevents accumulation of DNA damage and p53. (a) DMBA and TPA protocol. (b) H&E of skin sections from *Neat1* KO and WT mice treated as in (a). Quantification of numbers of cell layers in the IFE ($n = 3$) is shown. Scale bars, 50 μ m. (c–g) Immunostaining for phosphorylated histone H3 (p-H3) and cytokeratin 14 (K14) at 11 days and its quantification at 7–11 days ($n \geq 4$) (c), BrdU and cytokeratin 5 (K5) and quantification at 11 days ($n \geq 5$) (d), cleaved caspase 3 at 11 days and its quantification at 7–11 days ($n \geq 7$) (e). Significance was determined using unpaired two-sided t-test (BrdU and p-H3) or Mann–Whitney U test (Caspase 3*) with *, $p < 0.05$. (c) Dots represent the average number of positive cells per mm epidermis. (d–e) dots represent individual data points, the height of the bars the mean of all data points. Error bars are s.d. Scale bars, 30 μ m (c and e); 50 μ m (d). (f–g) Immunostaining for p53 (f) and γ -H2A.X (g). Dots indicate the number of positive cells per cell layer and mm of epidermis ($n \geq 3$). Bars represent 10–90 c.i. Significance on individual time points was calculated by the Holm–Sidak method. Significance between genotypes was calculated using two-way ANOVA with *, $p < 0.05$; **, $p < 0.01$. Representative images of immunostaining at 11 days are shown. Scale bars, 50 μ m.

Supplementary Figure 4

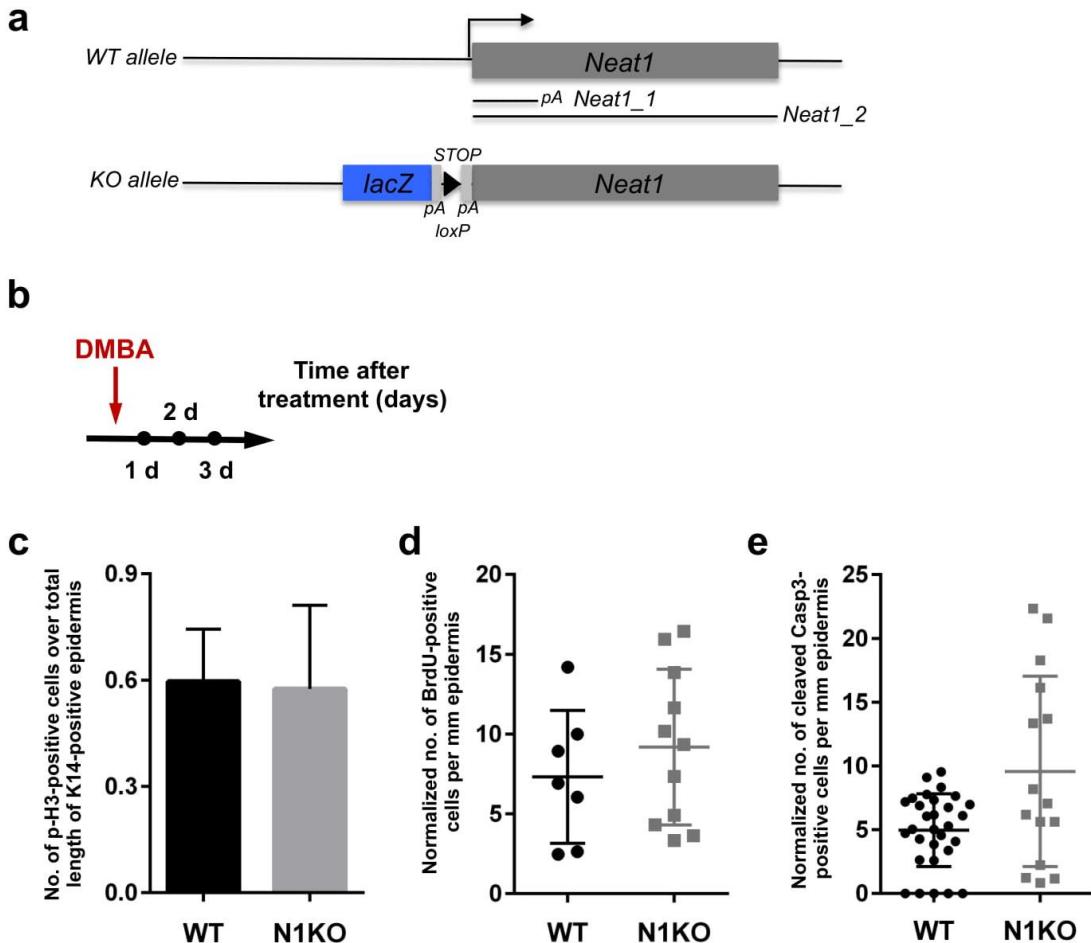


Figure S4. *Neat1* KO animals do not show proliferation defects and increased apoptosis after the first DMBA application. (a) Schematic representation of the *Neat1* wild-type (WT) and targeted (referred to as KO) alleles. Silencing of expression of both *Neat1_1* and *Neat1_2* isoforms was achieved by inserting a *lacZ*-transcriptional stop cassette immediately upstream of the TSS. See Nakagawa et al., J Cell Biol, 2011 for details. (b) Treatment scheme for short-term DMBA treatment of mouse back skin. (c-e) Quantification of phospho-Histone 3 (c; $n = 2$), BrdU- (d; $n > 3$) and cleaved Caspase 3-positive cells (e; $n > 3$) between day 1 and 3 following a single DMBA application.

NEAT1 depletion causes replication stress

We next asked whether *NEAT1* silencing leads to accumulation of DNA damage in established human cancer cells, which are by definition subject to oncogenic stress. To minimize confounding off-target effects, we silenced the *NEAT1_1* and *NEAT1_2* isoforms together or the long isoform (*NEAT1_2*) alone using both small interfering RNAs (siRNAs) and locked nucleic acid-modified antisense oligonucleotides (ASOs), which trigger RNase-H-mediated degradation of the target transcript upon ASO binding. We observed similar knockdown efficiencies and biological responses between these approaches (Supplementary Fig. 5b,c). Silencing of both *NEAT1* isoforms or of *NEAT1_2* alone in MCF-

7 cells led to progressive increases in γ -H2A.X levels, foci formation and DDR signaling as indicated by an increase in phosphorylation of the ATM substrate KAP1 at S824 (Fig. 5a).

Oncogenes induce DNA hyper-replication, which causes replication fork stalling²¹⁵. Sophisticated mitotic and S-phase checkpoint pathways have evolved to ensure replication completion and prevent replication fork collapse, which leads to double-strand break (DSB) formation and possibly cell death. Activation of the kinase ATR prevents formation of DSBs in response to replication stress²¹⁵. To assess whether NEAT1paraspeckles modulate this critical survival pathway, we knocked down NEAT1 in U2OS cells in which replication fork stalling was exacerbated by hydroxyurea (HU), an agent that blocks DNA synthesis²¹⁶. ATR signaling was compromised in *NEAT1* and *NEAT1_2* knockdown (KD) cells exposed to HU, as evidenced by a decrease in ATR-mediated phosphorylation of checkpoint kinase CHK1 at S345 and replication protein RPA32 at S33 (Fig. 5b).

Supplementary Figure 5

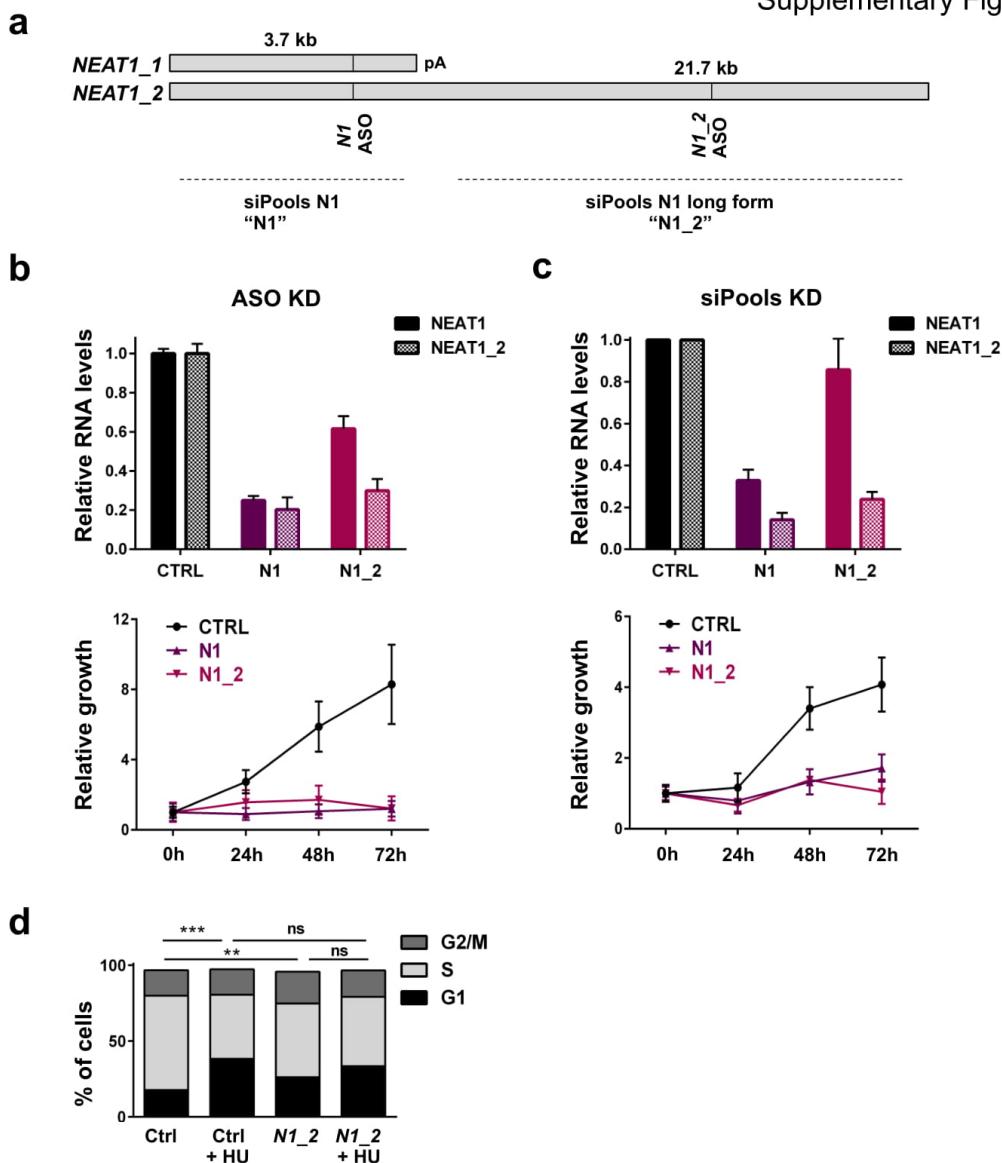


Figure S5. RNase-H and RNAi-dependent targeting of NEAT1 causes G1 arrest. (see previous page) (a) Schematic representation of the NEAT1 isoforms with the location of the LNA™ GapmeRs and siRNAs which are pooled to target either both isoforms (N1) or the long form only (N1_2). (b-c) Relative RNA levels of NEAT1 isoforms in MCF-7 cells 24 h after transfection of the indicated LNA™-GapmeRs or siRNA pools. Values are normalized to the levels in cells transfected with the non-targeting LNA™-GapmeR or siRNA pool control (CTRL), which are set to 1. Error bars represent s.e.m. of three technical replicates. Lower panels show cell growth as evaluated using WST1-assays following transfection of LNA™-GapmeR ASOs or the siRNA pools described in (a). Data are relative to growth at day 0, which is set to 1. Error bars represent SD. (d) Quantification of cell cycle distribution in U2OS cells exposed to 1mM hydroxyurea for 4.5 h. Statistical significance was determined by two-way ANOVA with multiple comparisons using the Tukey method.

Engagement of the ATR–CHK1 pathway activates G2–M and intra-S-phase checkpoints; failure of either step leads to replication fork collapse and formation of DSBs²¹⁷. The cell cycle profile of NEAT1_2 KD cells was only moderately affected in the absence of HU (Fig. 5c,d). In contrast, progression through S phase was severely compromised in cells expressing control siRNA (siCtrl), whereas NEAT1_2 KD cells continued to progress through S phase in the presence of HU (Fig. 5c,d). NEAT1_2 is therefore essential for activation of ATR signaling and checkpoint activation in response to replication stress.

NEAT1 targeting sensitizes cancer cells to chemotherapy

Many chemotherapeutic agents cause DNA damage in replicating cells and thereby enhance replication stress²¹⁵. The above results suggest that NEAT1_2 targeting may increase the sensitivity of human cancer cells to genotoxic chemotherapeutics. Consistently, we detected significantly higher amounts of γ-H2A.X foci in NEAT1 KD and NEAT1_2KD U2OS cells exposed to a 1-h pulse of bleomycin than in cells treated with control ASOs (Fig. 5e). Similarly, the increase in phosphorylated KAP1 and γ-H2A.X levels and foci formation observed in NEAT1 KD and NEAT1_2 KD MCF-7 cells (Fig. 5a) was exacerbated upon doxorubicin exposure (Fig. 5f). Knockdown of NEAT1 caused a marked decrease in cell growth and viability (Fig. 5g,h).

Poly(ADP-ribose) polymerase (PARP) inhibitors also enhance replication stress²¹⁸. Consistently, knockdown of NEAT1 or NEAT1_2 increased cancer cells' vulnerability to the potent PARP inhibitor ABT-888 (Fig. 5g,h). We obtained similar results in MCF-7 cells treated with shTP53 and in the TP53 mutant SCC cell line SQD9 (data not shown).

Silencing of NEAT1 or NEAT1_2 also increased Nutlin-3a-induced cytotoxicity (Fig. 5g,h), consistent with the idea that increased DNA damage sensitizes cells to p53 reactivation therapy²¹⁹. NEAT1 targeting is therefore a promising strategy to enhance the effectiveness of DNA-damaging chemotherapeutics and p53 reactivating agents.

Figure 5

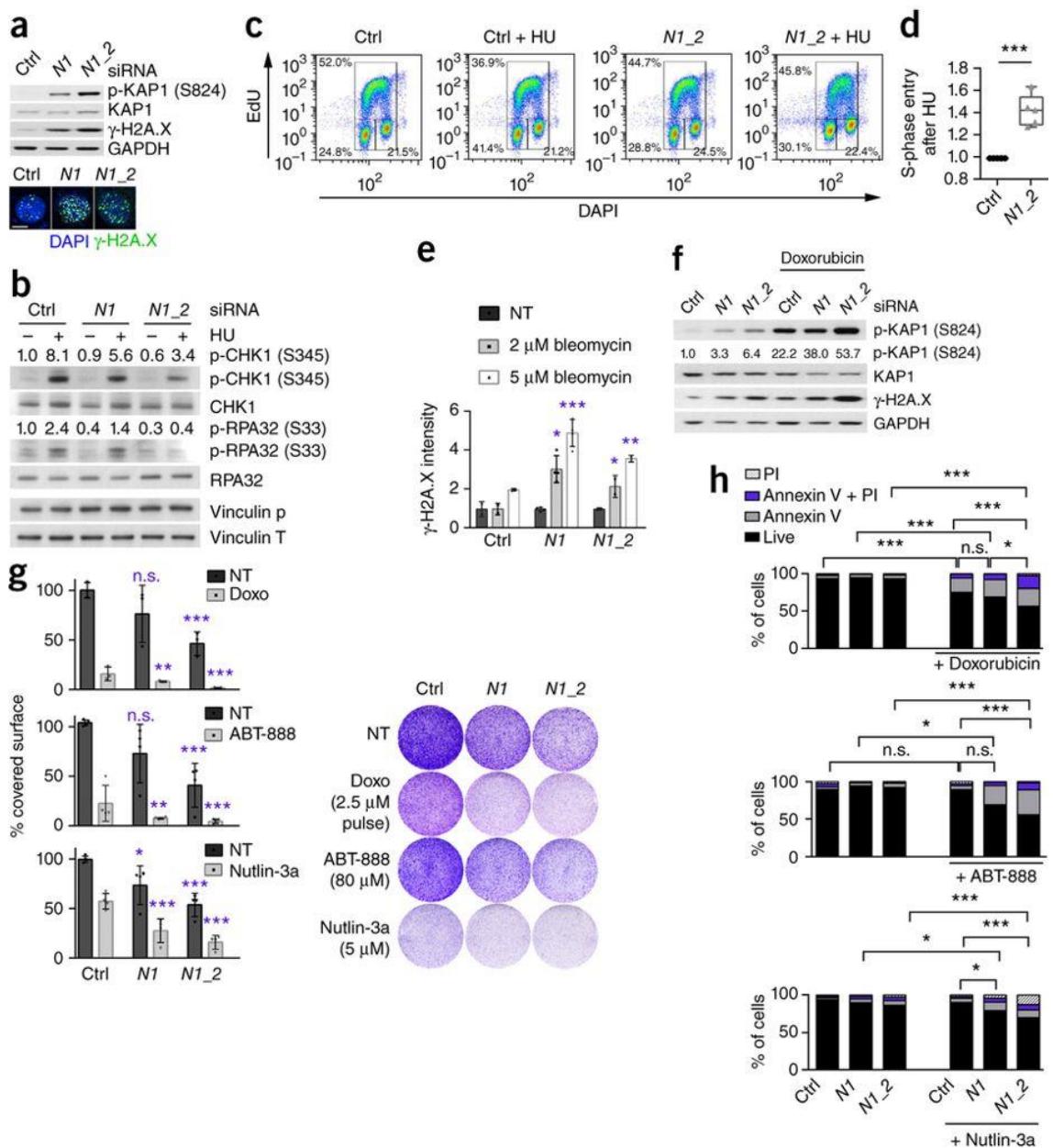


Figure 5. NEAT1 paraspckles modulate ATR-signalling and chemosensitivity. (a–b and f) Immunoblotting of MCF-7 (a), U2OS cells exposed to hydroxyurea (HU) (b) and MCF-7 to Doxorubicin for 1 h followed by a 24 h recovery period (f) in which both (N1) or only the long NEAT1 isoform (N1_2) were silenced. GAPDH and VINCULIN were loading controls. (a) Lower panel, immunofluorescence of γ -H2A.X and nuclear DAPI (blue) in NEAT1 KD MCF-7 cells. (c–d) Representative FACS analysis of U2OS cells exposed to HU, pulsed with EdU and stained with DAPI (c), quantification of S-phase entry relative to control (Ctrl) cells, which was set to 1 (d). Significance was determined by unpaired two-sided t-test ($n = 5$). (e) Quantification of γ -H2A.X staining in NEAT1 KD U2OS cells exposed to bleomycin for 1 h and a 24 h recovery period. Error bars represent s.d. Significance was determined by two-way ANOVA with Sidak post-testing correcting for multiple comparisons. (g) Colony formation assays using cells described in (a) and exposed to ABT-888, Nutlin-3a or 1h pulse of Doxo. Graphs show the average and s.d. ($n \geq 3$). Significance was calculated by Holm–Sidak after two-way ANOVA with *, $p < 0.05$; **, $p < 0.01$ and ***, $p < 0.001$. (h) FACS-based cell viability assay using Annexin V and propidium iodide (PI) staining of cells described in g. Significance was determined by Tukey after two-way ANOVA ($n = 3$).

NEAT1_2 expression levels predict chemotherapy response

To assess whether paraspeckles are assembled in human primary cancers, we performed *NEAT1* RNA–FISH on a tissue microarray (TMA) containing 97 specimens of diverse origins (Supplementary Table 2). Consistent with the absence of paraspeckles in the majority of normal adult tissues in mice¹⁶⁵, the non-neoplastic component in the TMA core, which includes normal resident and stromal cells, systematically showed no paraspeckle formation (Fig. 6a and Supplementary Fig. 6a). We confirmed the absence of paraspeckles in additional normal skin and breast tissue samples (Fig. 6a and Supplementary Fig. 6b). In contrast, we detected paraspeckles in about 65% of the human carcinomas analyzed, including skin SCC and ovarian carcinomas (Fig. 6a, Supplementary Fig. 6a,b and Supplementary Table 2).

Supplemental Figure 6

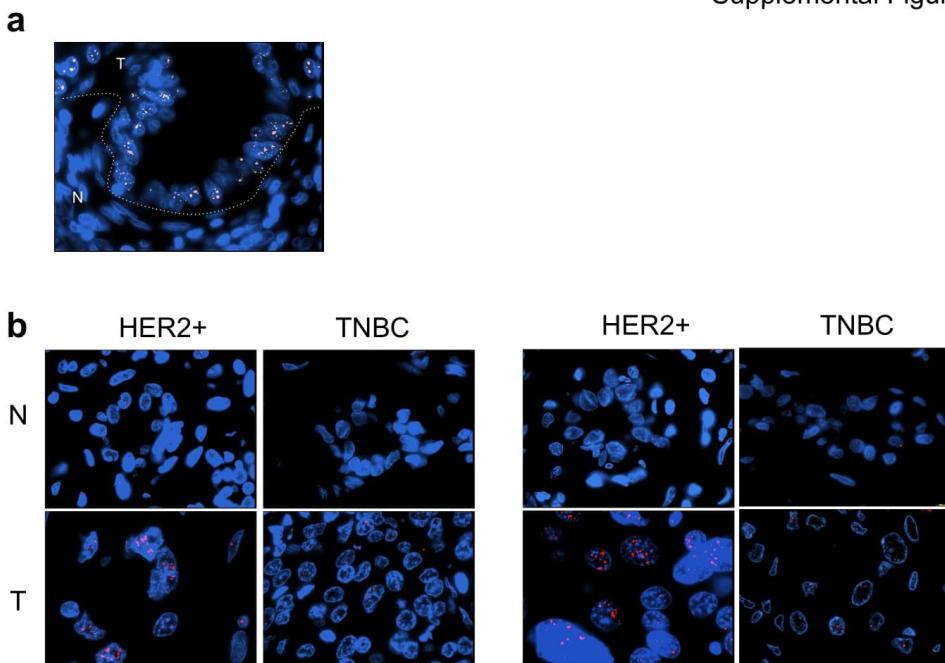


Figure S6. Increased NEAT1 paraspeckle formation in human cancers compared to normal matching tissues. (see previous page) (a-b) Representative images of *N* *NEAT1* RNA FISH in sections from a case of lung adenocarcinoma (a), HER2-positive (HER2+) and triple negative (TNBC) human breast tumors (T) and normal (N) matching breast tissues (b). In contrast to the neoplastic component (T), both the normal (N) peri-tumoral lung (a) and breast (b) tissues scored negative for paraspeckles. In blue: DAPI nuclear staining. In red: *NEAT1* RNA-FISH signal.

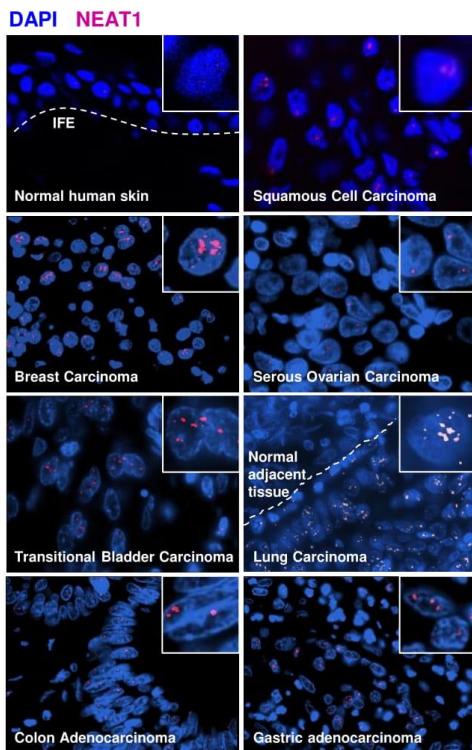
To evaluate whether *NEAT1* expression levels correlate with response to chemotherapy, we chose a cohort of primary high-grade ovarian carcinoma patients²²⁰ (GEO GSE30161), because all patients included in the cohort had been exposed to a first-line platinum-based chemotherapy, and expression data were available (owing to the presence of specific Affymetrix probes) for both *NEAT1_1* and *NEAT1_2*. When we correlated the expression

Figure 6

a

Primary site	Carcinoma Histology	Number of cases	Negative cases (score 0)	Positive cases (score +)	Positive cases (score ++)
Lung	squamous	4	0	4	0
	adenocarcinoma	4	2	2	0
Head and Neck	squamous	8	0	8	0
	ductal	8	4	4	0
Pancreas	neuroendocrine	2	2	0	0
	medullary	1	1	0	0
Thyroid	papillary	4	1	2	1
	follicular	3	0	2	1
	squamous	2	0	2	0
Cervix	adenosquamous	2	0	1	1
	ductal	6	3	2	1
Breast	lobular	2	1	1	0
	mixed	2	0	2	0
Skin	squamous	4	2	1	1
	basal cell carcinoma	4	4	0	0
Stomach	diffuse	3	2	0	1
	intestinal	5	4	1	0
Colo-rectal	adenocarcinoma	8	0	5	3
	clear cell	4	2	2	0
Kidney	cromophobe	2	1	1	0
	papillary	4	2	2	0
Bladder	transitional	8	1	7	0
	serous-papillary	5	2	3	0
Ovary	endometrioid	2	0	2	0
Total		97	34 (35.05%)	54 (55.67%)	9 (9.37%)

b



c

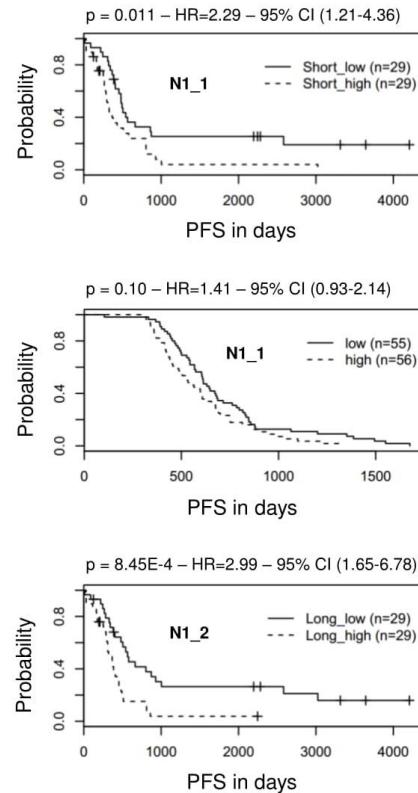


Figure 6. NEAT1 paraspeckles are detected in human cancers and NEAT1_2 levels correlate with response to platinum-based therapy. (a) Table with quantifications of cases in TMA with negative (score 0), moderately positive (score +) and very positive (score++) cases per tissue of origin. (b) Representative images of a series of paraspeckle-positive cases; normal control skin (upper left panel) is shown as a negative control. In blue: DAPI nuclear staining. In pink: NEAT1 RNA-FISH signal obtained with a probe set that detects both NEAT1 isoforms. Scale bar, 20 μ m; inset 8 μ m. (c–e) Kaplan-Meier curves of progression free survival (PFS) in ovarian carcinoma patients from the Ferris et al.²²⁰ (c and e) and TCGA (d) cohorts treated with platinum-based antineoplastic agents. P-values, Hazard Ratio (HR) and 95% confidence intervals (c.i.) are indicated above each panel.

levels of *NEAT1_1* with progression-free survival (PFS), correcting for race, disease stage, age and histology, we found that the correlation was not significant when considering a significance threshold of 0.01 ($P = 0.011$; Fig. 6b). We performed a similar analysis using the ovarian cancer cohort from the Cancer Genome Atlas²²¹ (TCGA), for which expression data were available only for the polyadenylated *NEAT1_1* transcript. To obtain a patient population comparable to the high-grade ovarian carcinoma cohort (Fig. 6b), we included only patients who had grade 2–3 serous ovarian cancers treated with first-line platinum-based chemotherapy and for whom recurrence data were available. This analysis confirmed that *NEAT1_1* expression is a poor predictor of chemotherapy response (Fig. 6c). In contrast, we observed a significant correlation between expression levels of *NEAT1_2* and PFS in the high-grade ovarian carcinoma data set (hazard ratio (HR) = 2.99, $P = 8.45 \times 10^{-4}$; Fig. 6d). On the basis of the analysis of this data set, it appears that expression of *NEAT1_2* but not *NEAT1_1* reliably predicts response of ovarian cancer to platinum-based chemotherapy.

SUPPLEMENTARY TABLE 2

Supplementary Table 2. Site of origin, histology and *NEAT1* score of primary carcinomas included in tissue microarrays

Primary site	Carcinoma Histology	Number of cases	Negative cases (score 0)	Positive cases (score +)	Positive cases (score ++)	Total % positive cases score + & ++
Lung	squamous	4	0	4	0	100.0
	adenocarcinoma	4	2	2	0	50.0
Head and Neck	squamous	8	0	8	0	100.0
	ductal	8	4	4	0	50.0
Pancreas	neuroendocrine	2	2	0	0	0.0
	medullary	1	1	0	0	0.0
	papillary	4	1	2	1	75.0
Thyroid	follicular	3	0	2	1	100.0
	squamous	2	0	2	0	100.0
Cervix	adenosquamous	2	0	1	1	100.0
	ductal	6	3	2	1	50.0
Breast	lobular	2	1	1	0	50.0
	mixed	2	0	2	0	100.0
Skin	squamous	4	2	1	1	50.0
	basal cell carcinoma	4	4	0	0	0.0
Stomach	diffuse	3	2	0	1	33.3
	intestinal	5	4	1	0	20.0
Colo-rectal	adenocarcinoma	8	0	5	3	100.0
	clear cell	4	2	2	0	50.0
Kidney	chromofobe	2	1	1	0	50.0
	papillary	4	2	2	0	50.0
Bladder	transitional	8	1	7	0	87.5
	serous-papillary	5	2	3	0	60.0
Ovary	endometrioid	2	0	2	0	100.0
Total		97	34 (35.0%)	54 (55.7%)	9 (9.4%)	65.1 %

Discussion

We demonstrate here that the lncRNA *NEAT1* enables tumorigenesis *in vivo* by promoting the survival of oncogene-targeted cells. Although recent studies have indicated that *NEAT1* can promote the growth and survival of *in vitro*-cultured cancer cells^{182,222} and that high *NEAT1* levels correlate with poor prognosis in various types of cancer^{223–225}, our data establish the first clear genetic link between *NEAT1* and carcinogenesis.

Consistent with published reports²¹⁰, we provide evidence that *NEAT1* is a bona fide p53 target gene. Moreover, we show that activation of p53 stimulates the formation of *NEAT1* paraspeckles. Our work therefore establishes a direct functional link between p53 and paraspeckle biology. There is growing evidence that paraspeckles regulate gene expression through a number of mechanisms, including sequestration of specific transcription factors and/or hyperedited mRNAs¹⁵⁸. Consequently, activation of this regulatory program downstream of p53 may profoundly influence p53-mediated biological responses and, importantly, influence cell fate decisions downstream of p53.

The role of p53 as the 'guardian' of the genome has long been recognized. p53 can modulate virtually all DNA repair processes by both transcription-dependent and independent mechanisms²²⁶. The observation that p53 induces *NEAT1* paraspeckle formation, which prevents the accumulation of excessive DNA damage in cells undergoing replication stress, reveals yet another route through which p53 preserves genomic integrity. Activation of the p53 network sets in motion an elaborate process of autoregulatory positive and negative feedback loops²²⁷. Our findings that p53 regulates *NEAT1* expression to stimulate paraspeckle formation and that *NEAT1* paraspeckles, in turn, dampen replication-associated DNA damage and p53 activation reveal an autoregulatory negative feedback loop that attenuates p53 activity in DNA-damaged cells.

Mechanistically, we provide evidence that *NEAT1_2* preserves genomic integrity and viability of preneoplastic cells by modulating ATR signaling. This pathway is crucial for ensuring replication completion and in preventing replication fork breakage in cells undergoing replication stress. Accordingly, ATR or CHK1 deficiency compromises the viability of cells undergoing oncogene-induced replication stress²¹⁵. We therefore propose a model in which paraspeckle assembly is one of the mechanisms employed by preneoplastic cells to cope with oncogene-induced replication fork stalling and, ultimately, prevent accumulation of replication fork breakage and the deleterious consequences associated with excessive DSBs.

The precise molecular mechanisms underlying the role of *NEAT1* in ATR–CHK1 signaling remain to be elucidated. We show that specific silencing of the *NEAT1_2* isoform is sufficient to reduce ATR signaling, which indicates that proper ATR activation may depend on *NEAT1_2*-mediated paraspeckle assembly. It is possible that paraspeckles promote the retention or sequestration of proteins directly involved in the termination of the ATR–CHK1 response (such as WT p53-induced phosphatase 1 or protein phosphatase 2A family phosphatases)^{228–230}. Alternatively, paraspeckles may act upstream of ATR–CHK1 signaling by decreasing the ability of oncogenes to induce hyper-replication by, for instance, controlling the availability of key replication factors and thereby limiting the firing of additional replication forks. Paraspeckles may also function as a site of activation (for example, through induction of specific post-translational modifications) of proteins that contribute to ATR–CHK1 activation.

In addition to its role in ATR signaling, *NEAT1_2* and/or paraspeckles may also contribute to DSB repair. However, paraspeckles do not colocalize with sites of ionizing radiation (IR)-induced DSBs (Supplementary Fig. 7a), suggesting that any influence of paraspeckles on DSB repair is indirect. Notably, and in contrast to previous reports^{231–236}, we found that all three RNA-binding proteins of the *Drosophila* DBHS family (SFPQ, NONO-p54nrb and PSPC1), which are all recruited to and essential components of paraspeckles, are excluded from sites of laser-induced DSBs (Supplementary Fig. 7b). On the basis of these observations, we propose that paraspeckles may contribute indirectly to DSB repair by sequestering DBHS, and possibly other, RNA-binding proteins away from DNA breaks to enable efficient recruitment of the DNA-repair machinery. Consistent with this possibility, it has been observed that transcription is interrupted at sites of damage and that RNA-binding proteins are rapidly excluded from these sites to allow efficient repair^{237–239}.

Neat1 KO mice do not show the phenotypes typically associated with deficiency in genes required for programmed DSB repair, such as growth retardation, progeria or defects in lymphocyte development²⁴⁰. Paraspeckles are therefore required for cell survival only under specific physiological conditions^{169,241} and, as shown here, in cells undergoing replication stress. This finding offers a unique opportunity to develop therapeutic modalities that are cancer-cell specific. The deficiency of cancer cells to respond to S-phase checkpoint activation could provide a basis for development of anticancer therapies that increase this vulnerability without affecting normal cells. *NEAT1* targeting exacerbates the sensitivity of cancer cells to DNA-damaging agents such as doxorubicin or platinum compounds.

Supplemental Figure 7

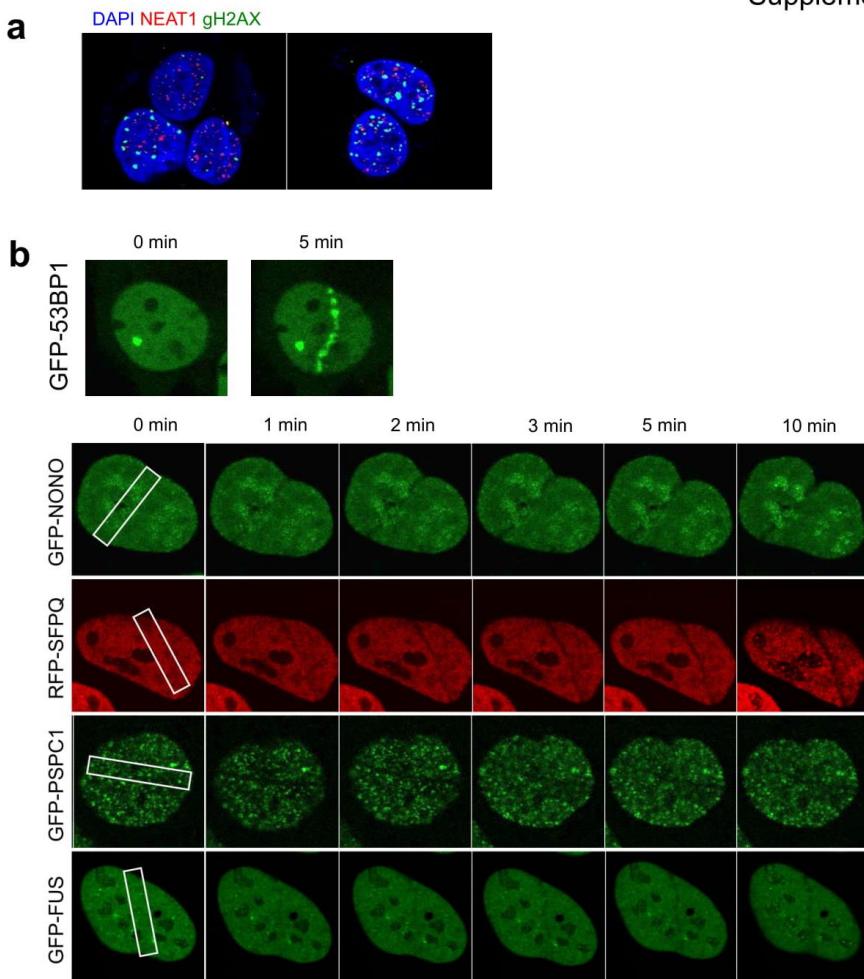


Figure S7. Paraspeckles do not colocalize with DNA damage foci and paraspeckle proteins are excluded from DSBs. (a) Combined NEAT1 RNA-FISH and immunofluorescence for phospho-g-H2AX in MCF7 cells exposed to 2Gy of radiation for 2 hours. (b) U2OS cells were transfected with the indicated constructs, subjected to laser micro-irradiation (400 μ W) and the localization of the indicated fluorescently-tagged proteins monitored at the indicated time points. (b) The position of the laserline is indicated by a white rectangle in the first panel. For (b) and (c), a single representative cell is shown for at least 3 independent experiments.

Moreover, and as expected if NEAT1 and p53 are engaged in a negative feedback loop, NEAT1 targeting is also synthetic lethal with nongenotoxic reactivation of p53. These data indicate that a wide range of human cancers may benefit from NEAT1 targeting, as this approach can sensitize tumors expressing WT or mutant p53 to specific combination therapies. Given the recent surge of interest in RNA-targeting therapeutics and antisense drugs in particular, this therapeutic strategy may be rapidly amenable to the clinic.

Materials and methods

Cell culture

All cell lines were acquired from the LCG ATCC Cell Biology collection and kept in culture at 37 °C and 5% CO₂ in medium supplemented with 1% penicillin and streptomycin (Invitrogen) and 10% FBS (Invitrogen). None of the cell lines used were reported in the ICLAC database of commonly misidentified cell lines. All cell lines were tested monthly for mycoplasma contamination and found negative. After their initial purchase, cell lines were not further authenticated. For knockdown experiments with locked nucleic acid (LNA)-GapmeRs (Exiqon), cells were transfected using Lipofectamine 2000 (Thermo Fisher Scientific) with a final concentration of 25 nM GapmeR and transfection reagent according to the manufacturer's instructions. NEAT1-specific sequences were 5'-TAAGCACTTGGAAAG-3' (N1) and 5'-CTCACACGTCCATCT-3' (N1_2), and a validated nontargeting oligonucleotide 5'-TCATACTATGACAG-3' was used as a control. For knockdown experiments with siPOOLs (siTOOLS Biotech), 30 nM siRNA was transfected using Lipofectamine RNAiMAX (Thermo Fisher Scientific).

Mice

Neat1 KO, p53 KO and WT mice were maintained on a pure C57BL/6J background in a certified animal facility. All animal experiments were carried out in accordance with the guidelines of the University of Leuven Animal Care and Use ethical Committee under project license 089/2013. *Neat1* and *Tp53* KO mice were described previously^{165,242}. K14 CreER KRas^{G12D} have been previously described²⁴³. For chemically induced carcinogenesis with DMBA (D3254, Sigma-Aldrich) and phorbol 12-myristate 13-acetate (TPA) (P8139, Sigma-Aldrich), cohoused 4- to 8-week-old mice were treated as previously described²⁴⁴. For the short-term experiments, mice were sacrificed at the indicated time points. For the long-term tumor formation assay, mice were sacrificed when the tumors exceeded a diameter of 10 mm or showed significant cachexia (weight loss ≥ 20%), or when > 50% of WT mice had reached those end points. WT and *Neat1* KO littermates were randomly assigned to each of the experiments irrespective of their gender. Owing to limited numbers of animals, no power testing was performed before the start of the experiments. The size of each group was determined by the number of animals at our disposal at the time of the experiment; *n* ≥ 3 mice per group for all experiments.

Quantitative real-time reverse transcription PCR

Total RNA was isolated using the NucleoSpin RNA/Protein kit (740933.50, Machery-Nagel) or the RNeasy Mini kit (74104, Qiagen) according to the manufacturer's instructions. RNA

was quantified with a NanoDrop 1000 (Thermo Scientific) and RNA was reverse transcribed using the High-Capacity cDNA reverse transcription kit (4368814, Applied Biosystems). Quantitative reverse-transcription PCR (RT-qPCR) was performed using Fast SYBR Green 2× master mix according to the manufacturer's instructions (4385612, Applied Biosystems). For normalization, the geometric mean of the two most stable reference genes out of at least three was used, calculated using geNorm. Real-time PCRs were carried out in a Roche LightCycler-480-384. RT-qPCR primer sequences were as follows: *NEAT1_1* fw: 5'-GGAGAGGGTTGGTTAGAGAT-3'; *NEAT1_1* rev: 5'-CCTTCAACCTGCATTCCTA-3'; *NEAT1_2* fw: 5'-GGCCAGAGCTTGTTGCTTC-3'; *NEAT1_2* rev: 5'-GGTGCAGGGCACTTACTTACT-3'; *UBC* fw: 5'-ATTGGGTCGCGGTTCTG-3'; *UBC* rev: 5'-TGCCTTGACATTCTCGATGGT-3'; *TBP* fw: 5'-CGGCTGTTAACCTCGCTTC-3'; *TBP* rev: 5'-CACACGCCAAGAACAGTGA-3'; *B2M* fw: 5'-TGCTGTCTCCATGTTGATGTATCT-3'; *B2M* rev: 5'-TCTCTGCTCCCCACCTCTAAGT-3'; *HPRT1* fw: 5'-TGACACTGGCAAAACAATGCA-3'; *HPRT1* rev: 5'-GGTCCTTTACCAGCAAGCT-3'; m*Neat1_2* fw 5'-GCTCTGGACCTTCGTGACTCT-3'; m*Neat1_2* rev 5'-CTGCCTGGCTTGGAAATGTAA-3'; m*Neat1* fw 5'-TTGGGACAGTGGACGTGTGG-3'; m*Neat1* rev 5'-TCAAGTGCCAGCAGACAGCA-3'; m*Hmbs* fw 5'-GCGGAGTCATGTCCGGTAA-3'; m*Hmbs* rev 5'-GTGGTGGACATAGCAATGATT-3'; m*Gapdh* fw 5'-AGGTTGTCTCCTGCGACTTCA-3'; m*Gapdh* rev 5'-GGTGGTCCAGGGTTCTTACTC-3'.

Immunoblotting

Cells were scraped on ice in RIPA buffer containing protease and phosphatase inhibitor cocktails (Thermo Fisher, 78442). The cell lysates were five times pushed through a 22-gauge needle with syringe and vortexed, incubated on ice for 10 min and then centrifuged at 21,000 × g for 15 min at 4 °C. For western blotting, 20 µg total protein lysate was loaded on NuPAGE Novex 4–12% Bis-Tris Protein Gels (Invitrogen) and probed with the following antibodies: phosphoKAP-1 (S824) (Bethyl Laboratories, A300-767A, 1/1,000); KAP1 (BD Biosciences, 610334, 1/1,000); histone H2A.X (p-Ser139) (Millipore, 05-636, 1/1,000) and GAPDH (Abcam, ab9485, 1/1,000), ATR (N-19) (Santa Cruz, sc-1887, 1/200), p-S345 Chk1 (Cell Signaling, 133D3, 1/1,000), total Chk1 (G4) (Santa Cruz, sc-8408, 1/1,000), p-S33 RPA32 (9H8) (Bethyl Laboratories, A300-246A, 1/1,000), total RPA32 (9h8) (Abcam, ab2175, 1/1,000), total H2A.X (Abcam, ab11175, 1/5,000).

Cell growth assays

Cells were seeded and transfected in 96-well plates at a concentration of 3,000 cells per well. At different time points, WST1 reagent was added according to the manufacturer's

instructions (Roche, 05 015 944 001) and luminescence was measured with VICTOR X3 Multilabel Plate Reader (PerkinElmer).

Long-term growth assays.

24 h after transfection, cells were counted and seeded for the long-term growth assay. 24 h after seeding, cells were treated with a 1-h pulse of 2.5 µM doxorubicin (Sigma-Aldrich), or with a constant dose of either 80 µM ABT-888 PARP inhibitor (Enzo Life Sciences, ALX-270-444-M001) or 5 µM Nutlin-3a (Roche). After 4 or 7 d in culture (5–8 d after transfection), colonies were stained with crystal violet (Sigma-Aldrich) and quantified using ImageJ software.

FACS analysis

To detect cell death, cells were stained for 15 min with Annexin V and PI using the FITC Annexin V Apoptosis Detection Kit II (BD Biosciences) according to the manufacturer's instructions. Cell death was detected on a BD FACSCanto (BD Biosciences), and data were analyzed with FlowJo software (Tree Star). For cell cycle distribution analysis in *NEAT1_2* KD conditions, cells were treated with vehicle or 1 mM HU for 5 h 48 h after transfection and pulsed for 30 min with 10 µM EdU. After trypsinization and washing once in cold PBS, cells were fixed for 20 min in 2% PFA in PBS. They were washed with PBS + 1 mg/ml BSA (PBS-B) and permeabilized with PBS + 0,02% Triton X-100 for 20 min then washed again with PBS-B. Subsequently, the cells were spun down and resuspended in the Click-reaction cocktail (per sample: 43.75 µl PBS, 1 µl 100 mM CuSO₄, 0.25 µl 200 µM azide dye and 5 µl 100 mM sodium ascorbate), incubated for 40 min, washed in PBS-B, resuspended in 0.5 ml PBS containing 0.5 µg/ml DAPI and analyzed on a BD FACS Canto. Data were analyzed using FlowJo software (Tree Star).

RNA-seq

MCF-7 cells were plated onto 24-well plates (60,000 cells/well). The next day, cells were treated with either 5 µM Nutlin-3a or ethanol vehicle. After 24 h, cells were washed in PBS (Gibco) and prepared for RNA extraction according to the RNeasy protocol (Qiagen), which yielded ~2 µg of total RNA per sample. The quality of the RNA samples was checked using a Bioanalyzer 1000 DNA chip (Agilent), after which libraries were constructed according to the Illumina TruSeq RNA sample preparation guide. Final libraries were pooled and sequenced on a HiSeq 2000 (Illumina), generating approximately 30 million reads of 50 bp in length. After removing adaptor sequences, reads were mapped to the human reference genome (hg19) using TopHat v1.3.3 (ref. ²⁴⁵) with default settings. Reads were aggregated with HT-Seq²⁴⁶ (-str = no parameter, version 0.5.3p3) using the human Gencode annotation

V18 (ref. ²⁴⁷), lncRNAs only. DESeq²⁴⁸ was used to normalize and to calculate differential expression between Nutlin-3a-stimulated and unstimulated samples. A final list of differentially expressed genes was obtained using adjusted *P* value < 0.05 and $|\log_2\text{FC}| > 1$. The RNA-seq data are available in the GEO database (GEO GSE47043).

ChIP-seq

MCF-7 cells were seeded at a density of 5 million cells per 15-cm dish to ~80% confluence and grown overnight at 37 °C. Cells were stimulated with 5 µM Nutlin-3a for 24 h. ChIP samples were prepared following the Magna ChIP-Seq preparation kit using the p53 antibody (DO-1, SCBT, 1 µg/sample). Per sample, 5–10 ng of precipitated DNA was used to perform library preparation according to the Illumina TruSeq DNA sample preparation guide. In brief, the immunoprecipitated DNA was end-repaired, A-tailed, and ligated to diluted sequencing adapters (dilution of 1/100). After PCR amplification with 15–18 cycles and gel size selection of 200- to 300-bp fragments, the libraries were sequenced using the HiSeq 2000 (Illumina). Cleaned reads were mapped to the human reference genome hg19 (UCSC) using bowtie (v2.0.0-beta3) with the addition of the parameter -local, allowing for further soft clipping of the reads. Reads with a mapping quality <4 were removed. Peak calling was performed using MACS²⁴⁹ (version 1.4.2) with the default *P* value threshold yielding 3,634 peaks. The ChIP-seq data are available in the GEO database (GEO GSE47043).

Motif discovery and network analysis

To detect putative p53 binding sites we adapted the i-cisTarget method²¹². In short, a region of 10 kb flanking the TSS of each lncRNA (GENCODE annotation V18) was scored for 6,863 different position weight matrices (PWM), representing a large collection of transcription factors. This led to a ranking of all lncRNAs for each of these PWMs. Next, for each PWM we retrieved the rank position of all upregulated lncRNAs and calculated an enrichment score for this set using the normalized area under the recovery curve, as described²¹². This score recapitulates how many of the upregulated lncRNAs are enriched at the top of the ranking for a particular PWM and whether this enrichment is significantly higher than expected by chance. From this analysis we selected the PWM with the highest enrichment score across all 6,863 PWMs, namely TP53. Next, we determined the optimal set of p53 targets as the rank position in the p53 ranking where the cumulative recovery of upregulated lncRNAs is highest compared to the background (i.e., the average recovery across all PWM rankings). For each predicted target lncRNA, we searched for p53 ChIP-peaks within 10 kb of the TSS of each lncRNA. If multiple peaks were detected, the one with the highest score was selected. The final network was generated using Cytoscape v2.8.2 (ref. ²⁵⁰).

Kaplan–Meier curve construction

The correlation of *NEAT1* expression levels with response to chemotherapy was evaluated in two data sets of ovarian tumors (TCGA ovarian data set and GSE30161). We selected ovarian carcinoma because all the patients were treated with platinum-based chemotherapy. The TCGA ovarian data set was generated by poly(A)-based RNA sequencing, in which specific expression of the non-polyadenylated long *NEAT1* isoforms is not available. This data set was extended with in-house-sequenced ovarian tumors and primary, high-grade, serous, platinum sensitive tumors with a recurrence, excluding the patients with Fédération Internationale de Gynécologie et d'Obstétrique (FIGO) 1, were selected. In addition, we assessed the Affymetrix Human Genome U133 Plus 2.0 Array data set to discriminate the effect of the long and short isoform of *NEAT1* on the response of chemotherapy. This microarray has probes (227062_at, 234989_at, 238320_at, 225239_at and 239269_at) that specifically detect the long *NEAT1* isoform, whereas probes 224565_at, 224566_at and 214657_s_at detect both the short and long isoforms of *NEAT1*. Data set GSE30161 consists of 58 high-grade ovarian tumors that all received platinum therapy. A Cox survival analysis of low and high *NEAT1* expression levels was performed on both data sets including race, stage, age and histology as covariates (for GSE30161) and grade, residual disease, FIGO and race for the extended TCGA data set. Kaplan–Meier curves were generated and Cox regression analyses were performed using the survival package in R 3.2.1.

Tissue staining and image quantification

For histological analysis, mouse back skin was dissected, spread on Whatman cellulose filter paper, fixed for 2 or 24 h in 4% PFA in PBS and processed for OCT or paraffin embedding, respectively. Samples were sectioned at 5–7 µm and routinely stained with hematoxylin (C0302, Diapath) and eosin (C0362, Diapath) (H&E). Serial sections obtained from the same samples were used for immunohistochemical analysis using the following primary antibodies: rabbit polyclonal anti-pH3 (Calbiochem, 382159; 1/200), chicken polyclonal anti-keratin 14 (Abcam, 13970; 1/1,000), rabbit polyclonal anti-p53 (Santa Cruz, sc-6243; 1/1,500), rabbit polyclonal anti-Ki67 (Thermo Scientific RM-9106-S, clone SP6; 1/300), rabbit polyclonal anti-γ-H2A.X (Cell Signaling, 2577; 1/1,400), rabbit polyclonal anti-cleaved caspase 3 (Cell Signaling, Asp175; 1/600), rat monoclonal anti-BrdU (Abcam, ab6326; 1/1,500), rabbit polyclonal anti-keratin 5 (Covance, PRB-160P-0100; 1/1,000). For immunofluorescence, anti-rabbit or rat IgG-A488 or A594 (Life Technologies, 1/500) and anti-chicken-IgG-RXX (Elitech, 703-295-155, 1/1,000) secondary antibodies were used. For immunohistochemical (IHC) staining, sections were deparaffinized in xylene and rehydrated using an ethanol series. Endogenous peroxidase activity was blocked by 15 min

incubation with 3–4% H₂O₂ (Perdrogen, 31642, Sigma-Aldrich) in dH₂O. Epitope retrieval was performed in citrate buffer (pH 6) or EDTA buffer (pH 8.5, p53 antibody) using a 2100 Retriever (Aptum Biologics). The sections were blocked in 1% BSA and 10% NGS for 40 min at room temperature (RT) and incubated overnight at 4 °C with the primary antibody in 1% BSA in T-TBS. EnVision reagent (K400311, Dako) was then applied on the sections for 45 min at RT. Immunoreactivity was revealed via a diaminobenzidine chromogen reaction (Peroxidase Substrate Kit, DAB, SK-4100, Vector Laboratories). Next, slides were counterstained in hematoxylin, dehydrated in an ethanol series, cleared in xylene, and permanently mounted with a resinous mounting medium (60200, Micromount Diapath). A 0.1% Tween-20 in TBS solution was used as washing buffer in between steps. For immunofluorescence staining, the same procedure was followed but with blocking in 10% NDS, use of fluorescently labeled anti-rabbit, mouse and chicken secondary antibodies (Life Technologies, 1/500), nuclear counterstaining with DAPI and mounting with VectaShield (H-1000, Vector Laboratories). For proliferation analysis with BrdU, 5 µl/g of a 20 mg/ml BrdU solution was injected intraperitoneally 4 h before sacrifice.

To quantify IHC staining, 5 hotspot brightfield images per mouse were taken with a Leica DM 2500 microscope at 20x magnification, and positive cells in the interfollicular epidermis and infundibulum of hair follicles were counted and normalized to the total length of the epidermis measured at the basal layer in the image. Counting of total number of individual cells in the same area gave near-identical normalized results. For quantification of fluorescent images, a similar approach was used employing 20x magnified images taken with a Leitz DMRB microscope. LAS software (Leica Microsystems) was used for all IHC and IF image acquisition. Quantification of hyperplasia in Figure 4b was done in H&E-stained sections (6 images per mouse at 10x magnification) by calculating the ratio of the total length of each category of epidermal thickness (1–2, 3–4, 5–7 and >7 cell layers) to the total length of the epidermal tract analyzed. All measures or quantifications and analyses were performed by two observers blind to genotype and treatment. Exclusion was based on the Grubbs test for outlier calculation with $\alpha = 0.05$; however, no significant outliers were detected during the analysis. To determine differences in variance of the groups, an F-test was performed with $\alpha = 0.05$. If the variances showed to be different, the Welch-corrected *P* value was represented. Validation of all antibodies used in this study is available on the manufacturer's websites.

RNA–fluorescence in situ hybridization

NEAT1 RNA–FISH was performed using Stellaris FISH probes (human, SMF-2036-1; mouse, SMF-3010-1) according to the manufacturer's protocol. In brief, after deparaffinization and rehydration, the tissue was immersed in 70% ethanol for >1 h at room

temperature, washed in PBS, digested with 10 µg/ml proteinase K in PBS, preincubated in wash buffer and hybridized overnight at 37 °C. The next day, the tissue was washed extensively, counterstained with DAPI and mounted with Vectashield Anti-Fade mounting medium (Vector Laboratories). For cryosections and human cells, no proteinase K digestion was performed. For simultaneous immunofluorescence, the following primary antibodies were used together with the FISH probes: rabbit polyclonal anti-keratin 5 (Covance, PRB-160P, 1/1,000) or rabbit polyclonal anti-keratin 1 (Covance, PRB-165P, 1/1,000). Secondary antibodies coupled to Alexa 488 (1/500, Life Technologies) were added to the first washing step after hybridization for 40 min at 37 °C. Images were taken using a Nikon A1 confocal microscope acquired through a Hercules grant type 1 AKUL/09/037 and processed for overlay and brightness and contrast adjustments using ImageJ. Red-channel images (Fig. 2e) were smoothed once for aesthetical enhancement of the paraspeckles. Paraspeckle quantification in cell lines was automated and used IN Cell Analyzer 2000 (GE Healthcare Life Sciences) and IN Cell Developer software (GE Healthcare Life Sciences). Paraspeckle quantification in tissues was done by taking 6 images per mouse at a 60× magnification and counting the numbers of nuclei with paraspeckles in proportion to the total number of nuclei in all layers of the interfollicular epidermis.

Tissue microarray (TMA)

Formalin-fixed paraffin-embedded (FFPE) tissue blocks were retrieved from the tissue bank of the Department of Pathology, Spedali Civili di Brescia, Brescia, Italy. Human tissues included normal tonsils as controls, whereas pathological samples included multitumor tissue microarrays (TMA) of mixed carcinomas. TMA blocks were obtained from a series of 97 primary carcinomas (PC) from different sites and constructed using an automated tissue microarrayer (TMA Master, 3DHistech). Three representative tumor areas were identified on H&E-stained sections. For each area, a 1-mm core was obtained by punching the original tissue block. 4-µm-thick tissue sections were H&E stained and checked for tumor cell content. Sections and stainings on TMA were performed as described above. Slides were examined using an epifluorescence microscope (Nikon, Eclipse 90I) with appropriate filters for SpectrumOrange and a UV filter for the DAPI nuclear counterstain.

Acknowledgements

We thank O. Van Goethem, G. Bervoets and S. Peeters for excellent technical assistance; E. Bonomi and M. Bugatti (Fondazione Beretta) for ISH analysis of human cancer samples; and S. Jackson for helpful comments on the manuscript and for imaging facility access. Imaging was done on a Nikon A1 confocal microscope acquired through a Hercules grant type 1 (AKUL/09/037) to W. Annaert. This work was supported by Interuniversitaire Attractiepolen (IUAP), an Institute for Science, Innovation and Technology (IWT) scholarship to C.A. and a Fund for Scientific Research Flanders (FWO) scholarship to L.S. W.V. is supported by Associazione Italiana per la Ricerca sul Cancro (AIRC, IG 15378) and by Ministero Salute (RF-2010-2315888).

Contributions

C.A., L.S. and J.B. designed and conducted experiments and acquired, analyzed and interpreted the data. M.L. performed DMBA and TPA treatments and monitored tumor development. A.V. and S.A. performed p53 ChIP-seq and RNA-seq experiments and data analysis. P.K. assessed DNA damage repair efficacy with bleomycin. B. Boeckx and D.L. analyzed expression data and constructed KM curves. E.R. and J.v.d.O. provided mouse and human pathology support. J.v.d.O. provided clinical samples. A.A.S., C.B. and E.L. designed research studies and contributed to interpretation of the data. G.L. and B. Beck provided reagents from mouse skin tumors. S.N. and T.H. provided *Neat1* KO mice. W.V. performed the TMA. P.W.G.W. provided reagents and contributed to interpretation of the data. All authors read and edited the manuscript. J.-C.M. designed research studies and wrote the manuscript.

Personal Contributions (CA)

The parts of the work I contributed to are as follows, partly together with L.S.: Figure 1: panel b-h. Figure S1: N/A. Figure 2: all panels. Figure S2: panel b-f. Figure 3: panel h-o. Figure S3: all panels. Figure 4: all panels. Figure S4: all panels. Figure 5: panel a-d. Figure S5: all panels. Figure 6: N/A. Figure S6: N/A. Figure S7: all panels. Made all figures. Designed experiments and analyzed data. Contributed to writing part of the paper.

Competing interests

JCM and LS own a patent to target *NEAT1_2* in cancer (US9783803B2). JCM and EL are the cofounders of NewCo, a company aiming to develop oligo-based therapeutics to target cancer. The other authors declare no competing interests.

References

For readability of this thesis, the references in this chapter are numbered in sequence with the other references used, and are integrated in the References part at the end of the manuscript.

Appendix: Supplementary Table I

SUPPLEMENTARY TABLE 1

Supplementary Table 1. List of putative p53-target lncRNA genes

id	log2Fold Change	padj	gencode subtypes	iRegulon	ChIP-peak	ChIP-peak score	annotation
RP11-83N9.5	2,664124648	4,8379E-17	lincRNA	motif	YES	1611,36	certain
RP3-510D11.2	2,816413601	4,43431E-09	lincRNA	motif	YES	785,61	certain
RP11-347D21.3	2,270496902	3,39627E-08	lincRNA	motif	YES	730,37	certain
NEAT1	1,578035665	5,31455E-11	lincRNA	motif	YES	380,91	certain
RP11-346D6.6	1,400718674	3,07563E-07	lincRNA	motif	YES	123	certain
SSPO	1,738305627	4,68873E-12	processed_transcript	motif	YES	61,47	certain
RP5-1148A21.3	3,046453069	5,95671E-29	antisense	motif	YES	273,58	PCO
RP11-347C18.3	2,939053969	7,13393E-07	sense_intronic	motif	YES	266,51	PCO
ACTA2-AS1	3,989922001	1,0479E-29	antisense	motif	YES	198,31	PCO
RP11-23N2.4	1,646014501	1,1624E-16	antisense	motif	YES	54,67	PCO
RP11-134G8.8	2,12377017	1,4055E-18	lincRNA	motif	YES	1211,39	PCO
RP11-611D20.2	2,193096151	7,00905E-13	antisense	motif	YES	833,5	uncertain
LINC00475	3,574234991	9,01892E-37	processed_transcript	motif	YES	157,54	uncertain
RP11-467J12.4	3,293785874	8,34553E-07	antisense	motif	YES	146,29	uncertain
RP11-462G8.3	3,538655192	1,71639E-08	antisense	motif	YES	125,48	uncertain
RP11-597D13.9	3,729388752	2,10326E-62	antisense	motif	YES	70,74	uncertain
LINC00086	4,56999279	7,72E-102	lincRNA	motif			certain
CTD-212E3.1	4,325899304	1,06612E-16	sense_overlapping	motif			certain
RP3-468B3.3	2,724353356	0,001919272	lincRNA	motif			certain
RP5-1092A3.4	1,981942332	1,25384E-07	antisense	motif			certain
TP53TG1	1,917132938	1,63325E-27	processed_transcript	motif			certain
CTB-147C22.8	1,8915064	9,84488E-07	antisense	motif			certain
LINC00886	1,872196262	1,9033E-12	lincRNA	motif			certain
RP11-435B5.5	1,727033499	0,001116576	lincRNA	motif			certain
DLGAP1-AS1	1,686047399	0,008859722	antisense	motif			certain
MIR22HG	1,618602807	3,64716E-06	lincRNA	motif			certain
RP11-379B18.5	1,573924462	0,001311576	lincRNA	motif			certain
U73166.2	1,520500719	1,09237E-05	lincRNA	motif			certain
LINC00883	1,188628453	6,45733E-06	lincRNA	motif			certain
MALAT1	1,127404142	9,94712E-18	lincRNA	motif			certain
LINC00173	1,104753916	0,003408608	processed_transcript	motif			certain
hsa-mir-6080	1,099865175	5,84051E-09	processed_transcript	motif			certain
RPS-1103G7.4	1,064712736	0,000438789	antisense	motif			certain
RP11-102L12.2	2,72951308	1,47019E-07	antisense	motif			PCO
NAV2-AS2	2,33259675	1,00784E-05	antisense	motif			PCO
RP11-442H21.2	2,118315636	9,21341E-54	antisense	motif			PCO
AC135048.13	1,949656499	9,81349E-06	processed_transcript	motif			PCO
AP006621.9	1,6301791	8,10477E-11	antisense	motif			PCO
KB-1471A8.1	1,560251819	4,05361E-05	lincRNA	motif			PCO
RP11-235E17.6	1,419463892	0,002288283	antisense	motif			PCO
RP11-10A14.3	1,35023168	2,08051E-06	antisense	motif			PCO
RP11-429J17.5	1,295518163	0,003849422	antisense	motif			PCO
NICN1-AS1	1,267943097	0,001467487	antisense	motif			PCO
RP11-318M2.2	1,202376519	0,003001572	antisense	motif			PCO
RP11-783K16.5	1,189853425	0,000590127	antisense	motif			PCO
RP11-319G6.1	1,185839597	0,008647871	antisense	motif			PCO
RP11-263K19.4	1,175980339	1,69987E-06	antisense	motif			PCO
RP11-640I15.1	1,169529985	2,17794E-06	antisense	motif			PCO
AP003419.11	1,121418163	2,09628E-06	processed_transcript	motif			PCO
RP11-961A15.1	1,101471903	0,022524199	antisense	motif			PCO

RP11-138P22.1	1,089007386	0,026071757	antisense	motif		PCO
SH3BP5-AS1	1,03745103	3,4258E-15	antisense	motif		PCO
RP11-361L15.4	1,037407704	0,001081907	processed_transcript	motif		PCO
RP11-256I23.2	1,017837575	0,01319531	antisense	motif		PCO
RP11-84A14.5	1,013013428	0,004887223	antisense	motif		PCO
CYP4F8	4,148867962	3,71611E-12	processed_transcript	motif		uncertain
RP11-417J8.6	3,829823665	6,24355E-22	lincRNA	motif		uncertain
MIR205HG	3,149045286	8,11035E-08	lincRNA	motif		uncertain
RP11-696N14.1	2,284312795	3,06228E-07	antisense	motif		uncertain
NR2E3	1,701831772	0,000553872	processed_transcript	motif		uncertain
RP4-773A18.4	1,685273497	5,11231E-05	antisense	motif		uncertain
RP1-8089.2	1,557654225	0,007730789	lincRNA	motif		uncertain
AC092594.1	1,301906536	0,015977914	lincRNA	motif		uncertain
DNAJC27-AS1	1,289725353	0,016550537	antisense	motif		uncertain
AC147651.3	1,192885193	0,027151286	lincRNA	motif		uncertain
AC007563.5	1,1890752	1,21461E-06	antisense	motif		uncertain
RPS-882C2.2	1,148639504	4,10535E-06	antisense	motif		uncertain
AL450992.2	1,129402511	0,006442586	antisense	motif		uncertain
RP11-793H13.11	1,092747657	0,025888811	antisense	motif		uncertain
CTD-2538C1.2	1,091160787	0,022149189	lincRNA	motif		uncertain
HOXA11-AS	1,045798733	0,001670331	antisense	motif		uncertain
RP11-17M16.2	1,00854281	0,023706986	antisense	motif		uncertain
CTC-479C5.17	2,6428206	2,65878E-15	3prime_overlapping_ncrna	motif		uncertain
LINC00663	2,070512989	0,000209633	lincRNA	YES	149,78	uncertain
AD000664.2	1,973258211	0,000984467	sense_intronic	YES	55,21	PCO
CTA-292E10.6	1,174574429	0,014713853	antisense	YES	439,82	certain
RP11-96C23.5	1,38276935	0,000837802	processed_transcript	YES	76,08	uncertain
UCA1	2,589174219	1,37012E-06	lincRNA			certain
RP11-400K9.4	2,42072511	9,53746E-21	lincRNA			certain
PCAT6	1,997444002	2,19173E-08	antisense			certain
AL121578.2	1,658986467	0,000607016	lincRNA			certain
RP11-96C23.14	1,585827358	8,25772E-05	processed_transcript			certain
LINC00865	1,57014214	2,98112E-10	lincRNA			certain
RP11-7K24.3	1,426368263	5,24005E-05	lincRNA			certain
CTD-2162K18.5	1,388527527	4,3392E-05	antisense			certain
RP11-167H9.6	1,369457552	8,26495E-05	lincRNA			certain
RP11-428L9.2	1,302149031	7,31988E-05	lincRNA			certain
LINC00243	1,294009894	7,0462E-05	processed_transcript			certain
PSMG3-AS1	1,255904256	4,70343E-15	lincRNA			certain
AP004372.1	1,250037336	0,003348024	antisense			certain
AC144831.1	1,200365001	0,01470506	lincRNA			certain
GS1-124K5.4	1,192791795	0,007451442	lincRNA			certain
RP1-151F17.2	1,170799536	0,006442586	lincRNA			certain
KB-1732A1.1	1,16034429	0,001182107	lincRNA			certain
RP11-663N22.1	1,158939776	0,001081907	lincRNA			certain
LINC00885	1,157371095	0,006010843	lincRNA			certain
LINC00664	3,480492865	9,76441E-08	lincRNA			PCO
RP11-611O2.5	3,076725649	4,48683E-12	processed_transcript			PCO
GSN-AS1	2,946714495	1,46368E-44	antisense			PCO
RP11-567L7.5	2,749443007	4,9793E-07	antisense			PCO
RP11-787I22.3	2,5592587	2,03782E-22	antisense			PCO
RP11-363E7.4	2,421309923	4,3049E-10	sense_overlapping			PCO
NAV2-AS1	2,165227398	6,07889E-10	antisense			PCO
RP11-290L1.3	1,959313011	2,37064E-06	antisense			PCO
RP11-151B14.6	1,857190459	4,02252E-07	antisense			PCO
RP11-126K1.8	1,631879964	9,94712E-18	antisense			PCO

Nature Medicine: doi:10.1038/nm.4135

RP11-57C13.3	1,524208052	0,004214707	antisense	PCO
RP11-203J24.8	1,480799447	2,61891E-22	antisense	PCO
STEAP3-AS1	1,4805621	1,98559E-09	antisense	PCO
RP4-694B14.5	1,387273192	1,73518E-20	antisense	PCO
RP11-129B22.1	1,304659839	1,92174E-19	antisense	PCO
AC084219.4	1,298515851	4,02467E-06	antisense	PCO
RN7SL1	1,28568911	0,02192029	antisense	PCO
CTD-3060P21.1	1,253041438	0,000463798	antisense	PCO
RP11-98D18.9	1,248502522	4,48975E-06	antisense	PCO
RP11-998D10.7	1,237701156	2,62207E-09	lincRNA	PCO
RP11-465N4.4	1,233885943	0,001928929	antisense	PCO
CTD-2653D5.1	1,223496314	6,0636E-19	antisense	PCO
RP11-15A1.7	1,149992026	0,004214707	antisense	PCO
RP11-512N21.3	1,12720262	0,014090891	antisense	PCO
CTC-458I2.2	1,11836682	2,02587E-05	antisense	PCO
RP11-326F20.5	1,096586488	3,51181E-05	antisense	PCO
AC007283.5	1,082574323	0,008957764	lincRNA	PCO
RP11-666A8.9	1,044426105	0,000172389	antisense	PCO
RP11-407N17.5	1,023870424	0,016550537	antisense	PCO
RP4-724E16.2	1,023716258	0,003749813	antisense	PCO
XXbac-				
BPG246D15.8	1,021446087	1,58355E-07	lincRNA	PCO
RP11-510N19.5	1,018906187	1,33892E-12	sense_intronic	PCO
AC104135.4	3,800338107	1,82015E-87	lincRNA	uncertain
AC104135.3	3,555469081	4,40981E-50	lincRNA	uncertain
RP11-78ZC8.2	3,505962526	2,75003E-18	lincRNA	uncertain
LINC00494	2,15752638	6,9125E-12	processed_transcript	uncertain
RP11-141M1.3	1,966912013	8,4597E-05	lincRNA	uncertain
RP11-774O3.3	1,915271814	0,001568239	lincRNA	uncertain
RP11-261P9.4	1,887903585	0,000122174	sense_intronic	uncertain
U73167.7	1,857785957	0,000585212	antisense	uncertain
PDCD4-AS1	1,476524427	2,01201E-06	antisense	uncertain
RP11-845M18.6	1,45701723	3,64716E-06	antisense	uncertain
RP3-410B11.1	1,40677663	0,001199369	lincRNA	uncertain
HEXA-AS1	1,406326868	0,000374214	antisense	uncertain
C9orf106	1,3960655	0,001853692	processed_transcript	uncertain
RP3-337H4.8	1,288958901	0,025888811	antisense	uncertain
CTD-2410N18.3	1,250889458	0,016550537	lincRNA	uncertain
CTD-2583A14.8	1,156727568	0,007730789	sense_intronic	uncertain
SLC2A1-AS1	1,145850156	0,047913297	lincRNA	uncertain
RP11-1017G21.5	1,103186701	0,005045383	lincRNA	uncertain
AC144450.1	1,037560813	0,03305511	antisense	uncertain

Supplementary Table S1. List of putative p53-target lncRNA genes.

Candidate lncRNA genes that are identified as putative p53 targets according to the protocol described in the M&M section. Dark blue: high confidence p53 target; Lighter and light blue: medium confidence p53 target; white: low confidence p53 target. PCO: overlapping with a protein-coding gene. Annotation reliability is depicted in the last column. Motif in the “iRegulon” column indicates whether a p53 motif is found by the iRegulon algorithm within 10 kb of transcription start site.

Chapter II

The lncRNA *NEAT1_1* is dispensable for normal tissue homeostasis and cancer cell growth

Carmen Adriaens^{1,2}, Florian Rambow^{1,2}, Greet Bervoets^{1,2}, Toomas Silla³, Mari Mito⁴, Tomoki Chiba⁵, Asahara Hiroshi⁵, Tetsuro Hirose⁶, Shinichi Nakagawa⁷, Torben H Jensen³ & Jean-Christophe Marine^{1,2, #}

¹Laboratory for Molecular Cancer Biology, Center for Cancer Biology, VIB, Leuven, Belgium;

²Laboratory for Molecular Cancer Biology, Oncology department, KU Leuven, Leuven, Belgium;

³Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark,

⁴RNA Systems Biochemistry Laboratory, RIKEN Cluster for Pioneering Research,

Saitama, Japan; ⁵Department of Systems BioMedicine, Tokyo Medical and Dental University,

Tokyo, Japan; ⁶Institute for Genetic Medicine, Hokkaido University, Sapporo, Japan; ⁷Faculty

of Pharmaceutical Sciences, Hokkaido University, Sapporo, Japan. [#]correspondence to

jeanchristophe.marine@kuleuven.vib.be

Key words

LncRNA, *NEAT1* isoforms, mouse genetics, cell cycle, RNA exosome, paraspeckles

biorXiv doi:

<https://doi.org/10.1101/574582>

Abstract

The *NEAT1* locus produces two long non-coding RNA isoforms of which the short isoform, *NEAT1_1*, is highly expressed in most tissues. The long isoform, *NEAT1_2*, is the central architectural component of paraspeckles (PS), which are nuclear bodies that are assembled in only selected tissues and cells exposed to various forms of stress. Silencing *NEAT1* causes defects in mammary gland development and corpus luteum formation and protects mice from skin cancer development. Knocking-down *NEAT1* and *NEAT1_2* in established cancer cell lines reduced their growth and sensitize them to DNA damaging agents. Because the *NEAT1_1* isoform completely overlaps with the 5' end of *NEAT1_2*, the contribution of the short isoform to these phenotypes has remained unclear. Using dual RNA-FISH to detect both the *NEAT1_1* and *NEAT1_2* transcripts simultaneously, we observed that *NEAT1_1* levels are dynamically regulated during the cell cycle. We show that *NEAT1_1* is degraded by the nuclear RNA exosome factor in a RRP40-dependent manner. Unexpectedly, however, cancer cells engineered to lack *NEAT1_1*, but not *NEAT1_2*, did not exhibit cell cycle defects. Moreover, *Neat1_1* isoform specific knockout mice did not exhibit any of the phenotypes observed in *Neat1*-deficient mice. We propose that the *NEAT1* functions are mainly, if not exclusively, attributable to *NEAT1_2* and, by extension to PS. The pathophysiological function of *NEAT1_1*, if any, remains to be elucidated.

Introduction

Long non-coding RNAs (lncRNAs) are RNA molecules exceeding 200 nucleotides in length that lack protein coding potential. In the past decade, lncRNAs have become prominent players in a wide range of cellular processes, including in the domains of gene regulation, spatial organization of the genome, and cell plasticity²⁹.

The lncRNA *NEAT1* is required for the assembly of paraspeckle (PS) nuclear bodies^{12,155,156}. The *NEAT1* locus produces two different isoforms: a long (~22.7 kb) isoform, *NEAT1_2*, and a short isoform (~3.7 kb), *NEAT1_1*¹⁵⁵. *NEAT1_1* is a highly conserved and abundant polyadenylated transcript, which is detected in virtually all tissues¹⁶⁵. *NEAT1_2* is produced in the absence of cleavage and 3'-end processing of the *NEAT1_1* transcript. Little is known about the mechanisms underlying this read through event, other than that it is dependent on the activity of a ubiquitous nucleic acid binding protein HNRNP K and a cleavage factor Im (CFIm) complex consisting of CPSF6 and NUDT21¹⁴(Naganuma et al., 2012). Critically, this switch from transcriptional termination to read through is required for *NEAT1_2* expression, and thereby for assembly of paraspeckles^{14,161}. *Neat1_2* foci are only detected in cells exposed to various forms of stress and in specific physiological conditions such as lactation-induced murine mammary gland proliferating tissue¹⁶⁹ or during the formation of the corpus luteum¹⁷⁰. Non-differentiated cells preferentially produce *NEAT1_1*, but lack the expression of *NEAT1_2* and thus PS^{156,177,251}. In human tissue, PS appear in over 65% of epithelial cancers¹⁴¹, where they predict poor prognosis¹⁹⁴. In contrast, they are either completely absent or only sporadically detected in normal adjacent tissues¹⁴¹.

The generation of *Neat1*-deficient mice in which expression of both isoforms was silenced, has been instrumental to assess the pathophysiological relevance of *Neat1*. These mice have highlighted a critical role for *Neat1* in the formation of a functional lactating mammary gland and corpus luteum and as a modulator of oncogenic stress^{141,169,170}. *Neat1* protects pre-neoplastic cells from accumulating excessive DNA damage and, thereby, is required for tumor initiation¹⁴¹. Because PS were detected in the cellular compartments affected by these phenotypes it is tempting to speculate that these phenotypes arose as a consequence of loss of *NEAT1_2* and PS. However, these mice were also deficient for *Neat1_1* and its contribution to these phenotypes therefore remains unclear.

Mechanistically, PS have been implicated in the regulation of gene expression and maintenance of DNA integrity in response to endogenous and exogenous forms of stress^{139,141,177,182,189,190,196,252}, potentially through their interaction with the RNA interference machinery and micro-RNAs^{193–195}, or via the modulation of transcriptional and post-transcriptional regulators^{174,179,180,183,190,191,253}. However, interpretation of these mechanistic

studies are again complicated by the complete overlap of *NEAT1_1* with the 5' end of *NEAT1_2*. Whether these two isoforms functionally interact¹⁰ and/or exert distinct biological functions remains unclear.

To address this, we employed dual RNA-FISH, isoform specific gene editing and knockdown strategies to study the individual contributions of *NEAT1* isoforms both in *in vitro* cultured cells and in mice (*in vivo*). We show that the two isoforms are differentially expressed at various phases of the cell cycle and that *NEAT1_1* is an unstable transcript that is efficiently targeted by the nuclear exosome machinery for degradation. Intriguingly, however, despite being highly conserved, ubiquitously expressed at very high levels and cell-cycle regulated *NEAT1_1* deficient cells and mice did not exhibit any of the phenotypes observed in cells lacking either both isoforms or *NEAT1_2* only. Moreover, the phenotypes observed upon silencing *NEAT1_2* alone were recapitulated upon silencing *NEAT1_2* in *NEAT1_1*-deficient cells. We conclude that the previously identified *NEAT1*'s biological functions are solely attributable to the *NEAT1_2* isoform, and by extension PS formation. The pathophysiological function of *NEAT1_1*, if any, remains to be elucidated. This study pleads for a more careful dissection of individual non-coding RNA isoforms and indicates that high abundance and conservation is not necessarily predictive of functionality.

Results

Differential regulation of the NEAT1 isoforms in response to various stimuli

To dissect the putative differential behavior of the two *NEAT1* isoforms in cultured cancer cells, we performed RNA-FISH with two distinct probes that either target both transcripts (red) and *NEAT1_2* specifically (blue) (Figure 1A). A pink signal (red+blue) therefore marks the presence of both transcripts, whereas red signals indicate the presence of *NEAT1_1* alone. Further molecular characteristics of the locus, including high levels of H3K27Ac and H3K4me3 indicating high transcriptional activity in cell lines derived from epithelial origin, can be found in figure 1A. In untreated, proliferating U2OS cells we observed a large pool of *NEAT1_1* in the nucleoplasm, outside of PS [37.7 ± 15.8 % of the cells] (figure 1B, C, left panel and box plot). Note that these cells are triploid for *NEAT1* and, consistently, three pink dots were detectable indicating the presence of both isoforms at PS. We and others showed that induction of p53 stimulates transcription of *NEAT1* and PS formation^{139,141,210,254}. Accordingly, treatment of the cells with the p53 inducer, Nutlin-3a, increased the size of PS. This was accompanied by a dramatic increase in the proportion of cells displaying nucleoplasmic *NEAT1_1*-specific signal (79.0 ± 8.3 % of the cells; figure 1B, C center panel and box plot). In contrast, exposure to the DNA damaging agent Hydroxyurea (HU), decreased the *NEAT1_1* only red signal (with only 5.2 ± 3.7 % of the cells being positive; figure 1B, C, right panel and box plot).

Real time quantitative PCR (RT-qPCR) with primers detecting both isoforms and *NEAT1_2* only (figure 1A) established that *NEAT1_2* was specifically upregulated in cells exposed to HU (Figure 1D, E). As expected, an increase in the levels of the p53-target *CDKN1A* was also observed, indicating its transcriptional activation (figure 1F). Although we noted that the size of *NEAT1_2*-containing bodies slightly decreased in these cells, we confirmed that these bodies are genuine PS by co-staining with NONO, a canonical PS protein^{13,150} (figure S1). These data indicated that the ratio and localization of *NEAT1_1* and *NEAT1_2* vary dramatically depending on the type of stress inflicted to the cells. These experiments also highlighted the presence of a large pool of *NEAT1_1* that does not overlap with *NEAT1_2*-containing PS^{165,167}.

Figure 1

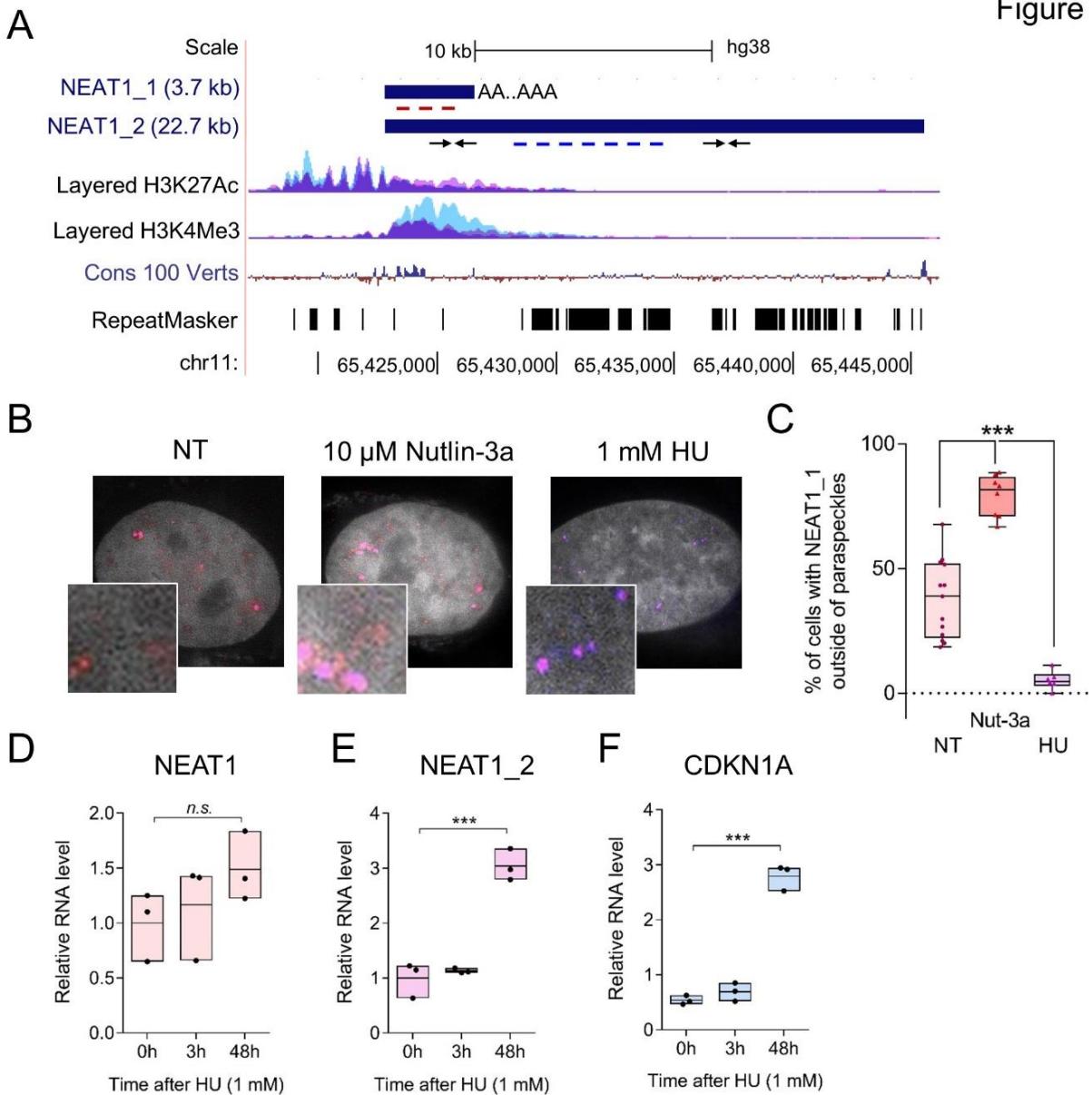


Figure 1: NEAT1 isoforms are induced differentially by different p53 activating stressors. A. Refseq representation of human NEAT1 isoforms in the UCSC genome browser (hg38) with layered H3K27 acetylation and H3K4me3 from three different epithelial cell types; nucleotide-level conservation among 100 vertebrates and the location of repeats in the genome sequence. Red and blue dotted lines represent RNA-FISH probes targeting both and the long NEAT1_2 isoform specifically, respectively. Note that when targeting the long isoform, blue and red probes will overlap and thus show paraspeckles in pink. Small arrows represent approximate locations of RT-qPCR primers used in this study. B. Representative superresolution images of RNA-FISH targeting NEAT1 isoforms in U2OS cells in non-treated (NT), 10 μ M Nutlin-3a (24h) and 1mM hydroxyurea (HU, 48h) conditions. C. Quantification of the percentage of cells in which the short isoform can be observed outside of paraspeckles. Each dot represents an independent experiment ($N = 13, 8, \text{ and } 6$, respectively). D, E and F. Relative levels of NEAT1 (both isoforms, D), NEAT1_2 (E) and the canonical p53 target CDKN1A (F) after 0, 3 and 48h of HU treatment (1mM).

Figure S1

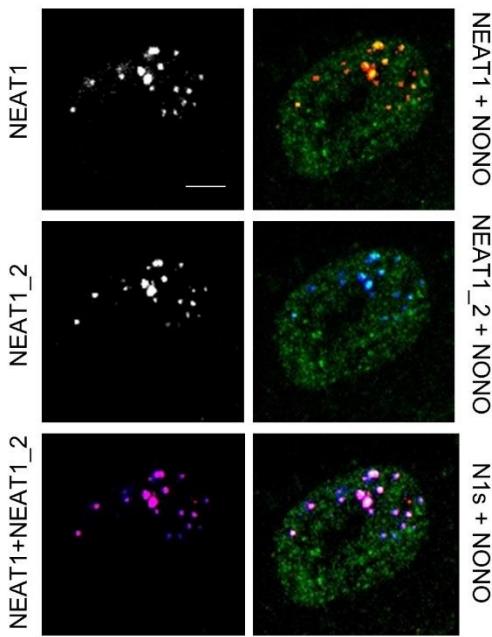


Figure S1: NEAT1_2 accumulations in HU conditions are genuine paraspeckles. NEAT1/NEAT1_2 RNA FISH of a HU treated cell (48h, 1 mM) combined with regular immunofluorescence against NONO, a canonical paraspeckle protein, showing paraspeckle integrity is preserved in HU conditions. Scale bar, 10 μ m.

NEAT1_1 levels are dynamically regulated during the cell cycle

In U2OS, Nutlin-3a stalls cells primarily in the G1 phase of the cell cycle through activation of CDKN1A²⁵⁵. In contrast, HU arrests cycling cells in S phase through inhibition of the deoxynucleotide (dNTP) producing enzyme ribonucleotide reductase (RNR), thereby depleting the dNTP pool during replication²⁵⁶. Because Nutlin-3a and HU induce different effects on the cell cycle, we investigated whether NEAT1_1 and NEAT1_2 levels were differentially regulated at different phases of cell division.

We deprived cells from serum for three consecutive days, halting them in a resting, G0-like state (G0). We then released these cells in 20% serum in the presence of the DNA polymerase inhibitor Aphidicolin (5 μ g/ml) for 24h to halt them at the G1/S phase boundary (Figure S3A). To check proper synchronization, we subjected the cells to DNA content analysis using flow cytometry (figure 2A). Using the above described dual RNA-FISH strategy, we observed that $87 \pm 15\%$ of G0 halted cells expressed the short NEAT1_1 isoform (figure 2B, C; figure S2D). In contrast, only $4.7 \pm 7.8\%$ of the cells arrested at the G1/S phase displayed the NEAT1_1-specific signals that do not overlap with NEAT1_2 (figure 2B, C; figure S2D). RT-qPCR analysis indicated that the total levels of NEAT1, but not of NEAT1_2, were markedly increased in G0 cells (figure S3C). Contrastingly, in G1S cells the levels of both NEAT1 and NEAT1_2

were comparable and slightly lower as compared to non-synchronized cells (figure S3D). Quantification of the number of spots for both *NEAT1* and *NEAT1_2* RNA-FISH probes revealed that G0 cells display on average three *NEAT1_2*-containing PS and on average 24 ± 23 *NEAT1_1* spots (figure S2A, C). In contrast, G1S cells displayed on average 4.5 PS per cell (figure S2B). Together, these data indicated that, in non-proliferating cells, *NEAT1_1* is the predominant isoform.

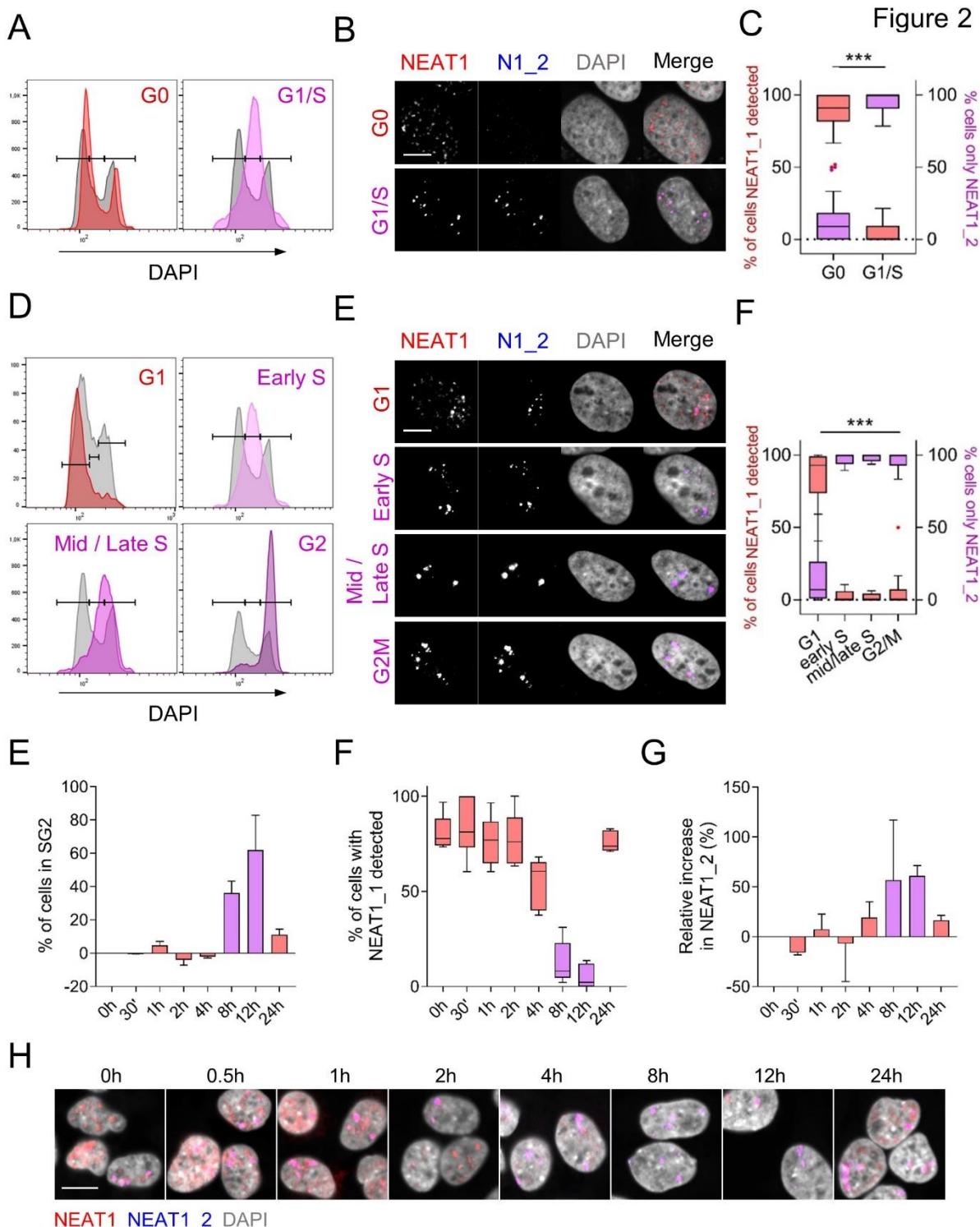


Figure 2: NEAT1 isoforms are alternatively regulated during cell cycle. A. Representative DNA content (DAPI) distribution of U2OS cells in G0 (3 days starvation) and G1/S (G0 cells released in 20% serum + 5 µg/ml Aphidicolin for 24h). Grey background plots are control non-synchronized (NS) cells in culture. B. Representative images of NEAT1/NEAT1_2 RNA FISH of the cells in A. Scale bar, 10 µm. C. Quantification of the % of cells in which NEAT1_1 was detected independently of NEAT1_2 paraspeckles, red boxplots, left y-axis. Percentage of cells in which only NEAT1_2/paraspeckles were detected, right y-axis, purple boxplots. Tukey plots of individual data points (1 point per picture). Significance was calculated using an unpaired, two-sided t-test on independent biological replicates ($N = \text{at least three}$). ***, $p < 0.001$. D. Like in A, but from U2OS cells synchronized by double thymidine block and released for 18 (G1), 2 (early S), 4 (mid/late S) and 8 hours. E. Representative images of NEAT1/NEAT1_2 RNA FISH of the cells in D. F. Same as C for the cells in D. Significance was calculated using a 2-way ANOVA with Dunnett's correction for multiple comparisons on independent replicates ($N = \text{at least three}$). ***, $p < 0.001$ for G1 vs early S, mid/late S, and G2. N1_2 = NEAT1_2. E. % of cells in S and G2 phases at different time points relative to time = 0h in HeLa cells upon release into 20% serum after 3 days starvation. F. Box plots (Min to Max) of individual data points ($N = 5$ and 3 pictures per replicate, respectively) quantifying the % of cells in E in which NEAT1_1 was detected using RNA-FISH against both and the long isoform specifically. G. Relative increase of NEAT1_2 by RT-qPCR of the cells in E. Error bars are standard deviation of $N = 2$ experiments in E and G. H. Representative pictures of cells in E. Scale bar, 20 µm.

To substantiate these data, we used a conventional double thymidine block-and-release protocol to synchronize the cells and released them for 2 (early S), 4-6 (mid/late S), 8 (G2M) and 17-20 hours (G1) (figure 2D, figure S3B). The majority of the cells in G1 ($85.3 \pm 16.7\%$) displayed observable NEAT1_1 signal outside of PS, whereas in early S, mid/late S and G2M phases NEAT1_1 signal was found in a very small fraction of the cells ($2.4 \pm 3.87\%$; $1.63 \pm 2.6\%$ and $5.8 \pm 1.3\%$, respectively; figure 2E-F and figure S2D). Only $2.2 \pm 1.4\%$ of the cells did not display any detectable NEAT1 staining (data not shown). We characterized the numbers of NEAT1 and NEAT1_2 spots per cell in G1, S, and G2 phases. In G1 cells the number of NEAT1_1 foci per cell varies greatly, with an average of 24.7 ± 22 . In contrast, the number of NEAT1_2 spots/PS fluctuates between 6 and 8.5 (figure S2A-C).

RT-qPCR analysis showed elevated levels of total NEAT1 in G1 compared to the other phases of the cell cycle. Since NEAT1_2 levels remained relatively constant throughout the cell cycle, elevated levels of NEAT1 in G1 cells is a consequence of higher NEAT1_1 levels in this particular phase (figure S3E-G). Accordingly, the ratio of the levels of NEAT1_2 over NEAT1 (NEAT1_1 + NEAT1_2) (figure S3H) in early S, mid/late S and G2 phases revolved around 1 (mean = ~1.4 in early S and ~0.98 in both mid/late S and G2). In contrast, this ratio was consistently smaller than 1 (mean = ~0.18) in G1 cells, indicating that NEAT1_1 contributes to the total levels of NEAT1 in these cells. Moreover, in G1/G0 cells, no linear relationship could be established between NEAT1_2 and NEAT1 ($R^2 = 0.1141$, p-value 0.259), whereas in S and G2 samples we observed a strong positive correlation ($R^2 = 0.6488$, $p < 0.001$), and b0 and b1 values of nearly 0 and 1, respectively (b0 = 0.09, b1 = 1.3). These results strongly indicated that NEAT1_1 levels drop as cells engage a new round of cell division. Similar results were

made in another cancer cell line (HeLa cells; figure 2E-H). We concluded that *NEAT1_1* levels fluctuate during the cell cycle. In contrast, *NEAT1_2* levels remain relatively constant during the cell cycle. Because *NEAT1_2* is the product of a read-through event, the downregulation of *NEAT1_1* as cells engage into DNA replication cannot be due to a decrease in transcription, but must instead occur through active degradation of the transcript. Note that this observation contrasts with the previously proposed model in which *NEAT1_1* would be recruited to *NEAT1_2*-containing PS¹⁵.

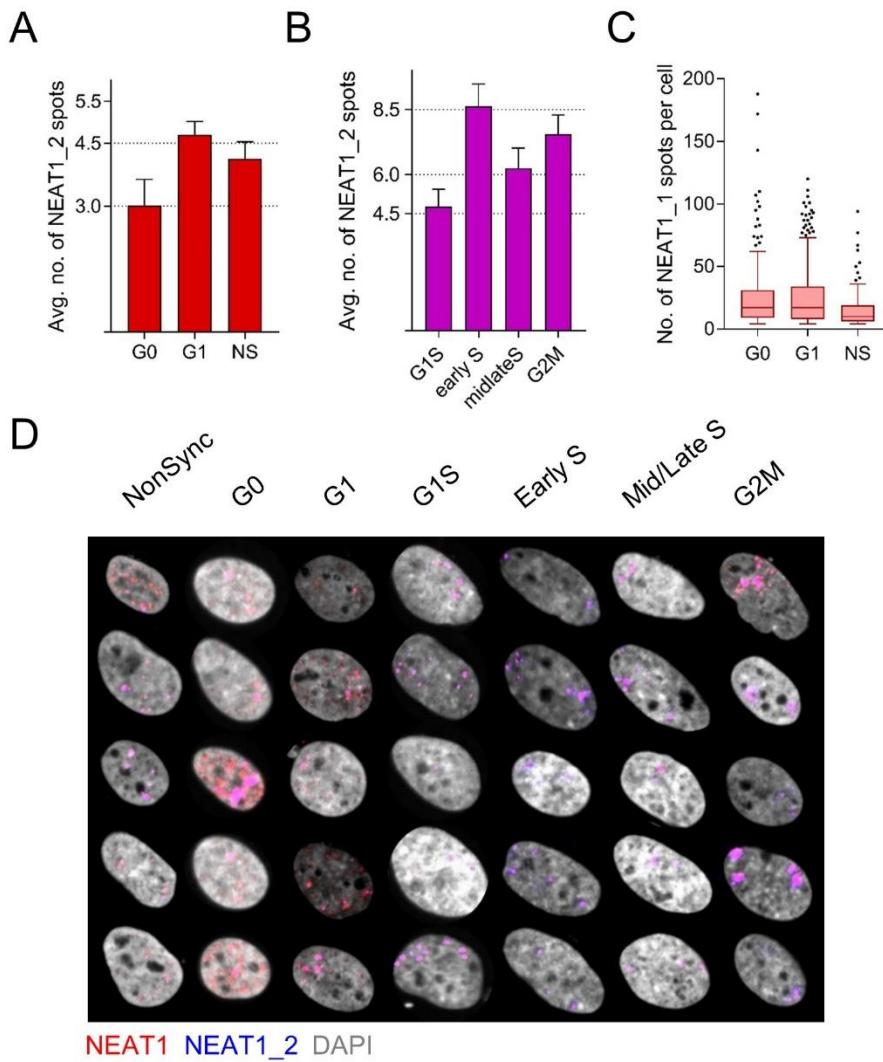


Figure S2

Figure S2: Paraspeckle and *NEAT1_1* characteristics in different cell cycle phases. A-B. Quantification of the average number of *NEAT1_2* spots in cells in which *NEAT1_1* was detected (A) or not (B) in the majority of the cells. Bars show mean + standard error of individually quantified cells. C. Number of microspeckles per cell in NS, G0 and G1 cells, Tukey boxplot representing one data point per cell. D. Additional representative images showing the variation in RNA-FISH patterns of cells in the different cell cycle stages described in figure 2. For visualization purposes, cell sizes in this picture are not to scale.

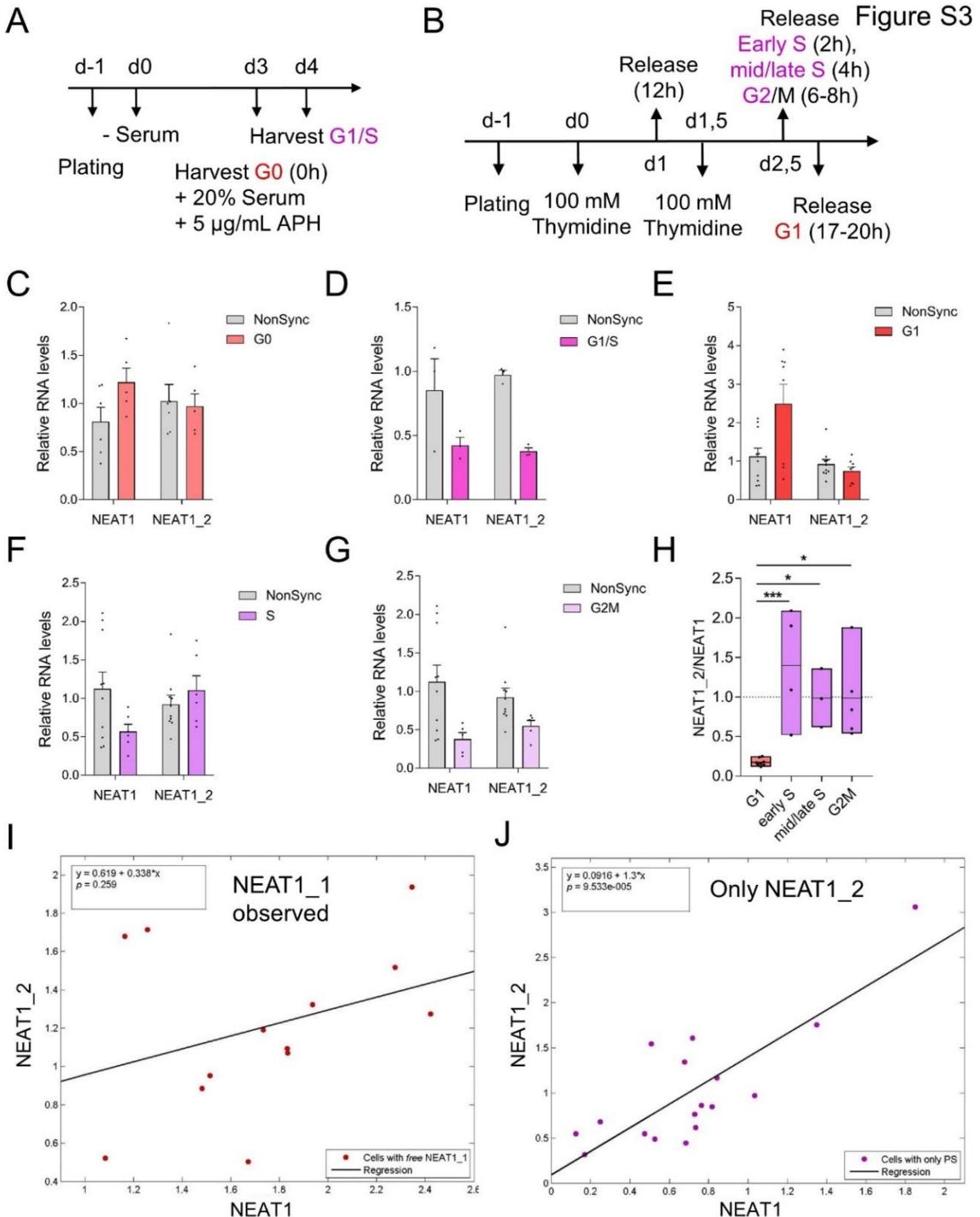


Figure S3: RT-qPCR analysis of NEAT1 isoforms in different cell cycle stages. A. Experiment layout for Figure 2 A-C. B. Experiment layout for Figure 2 D-E. C-G. Relative levels of NEAT1 (both) and NEAT1_2 isoforms by RT-qPCR analysis in G0 (C), G1/S (D), G1 (E), S (F) and G2M (G) phases of the cell cycle as synchronized in A & B of this figure. Each data point is an independent experiment. Bars and error depict mean + sem. H. NEAT1_2 long isoform levels relative to the total of NEAT1 detected in the cells in the experiments in figure 2 D-F. Bars are min to max with each data point an independent experiment. Lines depict the mean. Significance was tested by 1-way ANOVA with Sidak correction for multiple comparisons. *, p<0.05, ***, p<0.001. I-J, Pearson's correlation analysis of relative levels of NEAT1_2 over total levels of NEAT1 in RT-qPCR corrected for primer efficiencies in cells with primarily NEAT1_1 detected outside of paraspeckles (I, $r^2 = 0,1141$) and cells with only NEAT1_2 ("only PS") detected (J, $r^2 = 0,6488$) in the RNA-FISH analysis in figure 2 D-F.

***NEAT1_1* is degraded by the RNA exosome**

In order to identify factors that contribute to the degradation of *NEAT1_1*, we mined publically available datasets and observed that *NEAT1_1* levels were upregulated, whereas *NEAT1_2* levels remained largely unchanged upon silencing of the exosome component RRP40 (figure S4). We confirmed these results by RNA-FISH (figure 3A-B). In addition, RT-qPCR analysis highlighted an increase in *NEAT1*, but not *NEAT1_2*, levels in RRP40 KD cells (figure 3C). KD of RRP40 results in the stabilization of a series of nuclear polyadenylated RNAs, which accumulate in typical foci⁶⁸. Combining RNA-FISH probes targeting *NEAT1* and poly A+ RNA foci, we detected an accumulation of *NEAT1* in the poly A+ foci (figure 3D-E). We also noted that a pool of *NEAT1* did not overlap with poly A+ RNA, which likely represents *NEAT1_2* as this transcript is not polyadenylated and rather stable throughout the cell cycle. To confirm this, we performed *NEAT1* and polyA RNA-coFISH and immunofluorescence to detect the canonical paraspeckle protein NONO. Consistent with our prediction, paraspeckles and *NEAT1*/polyA+ foci do indeed not overlap (figure 3F). Note that RRP40 depletion, and by extension *NEAT1_1* accumulation in polyA+ foci, did not impact on the cell cycle progression (figure 3G-H).

Figure S4

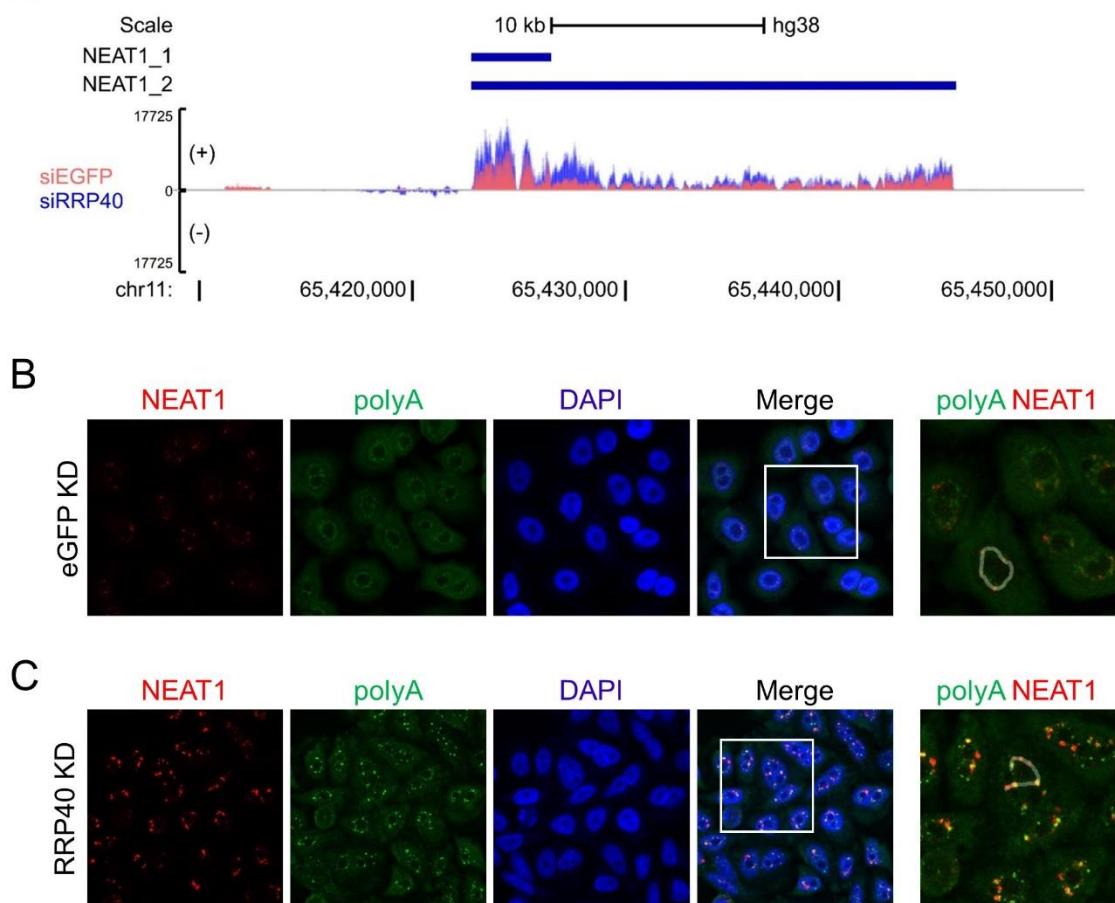
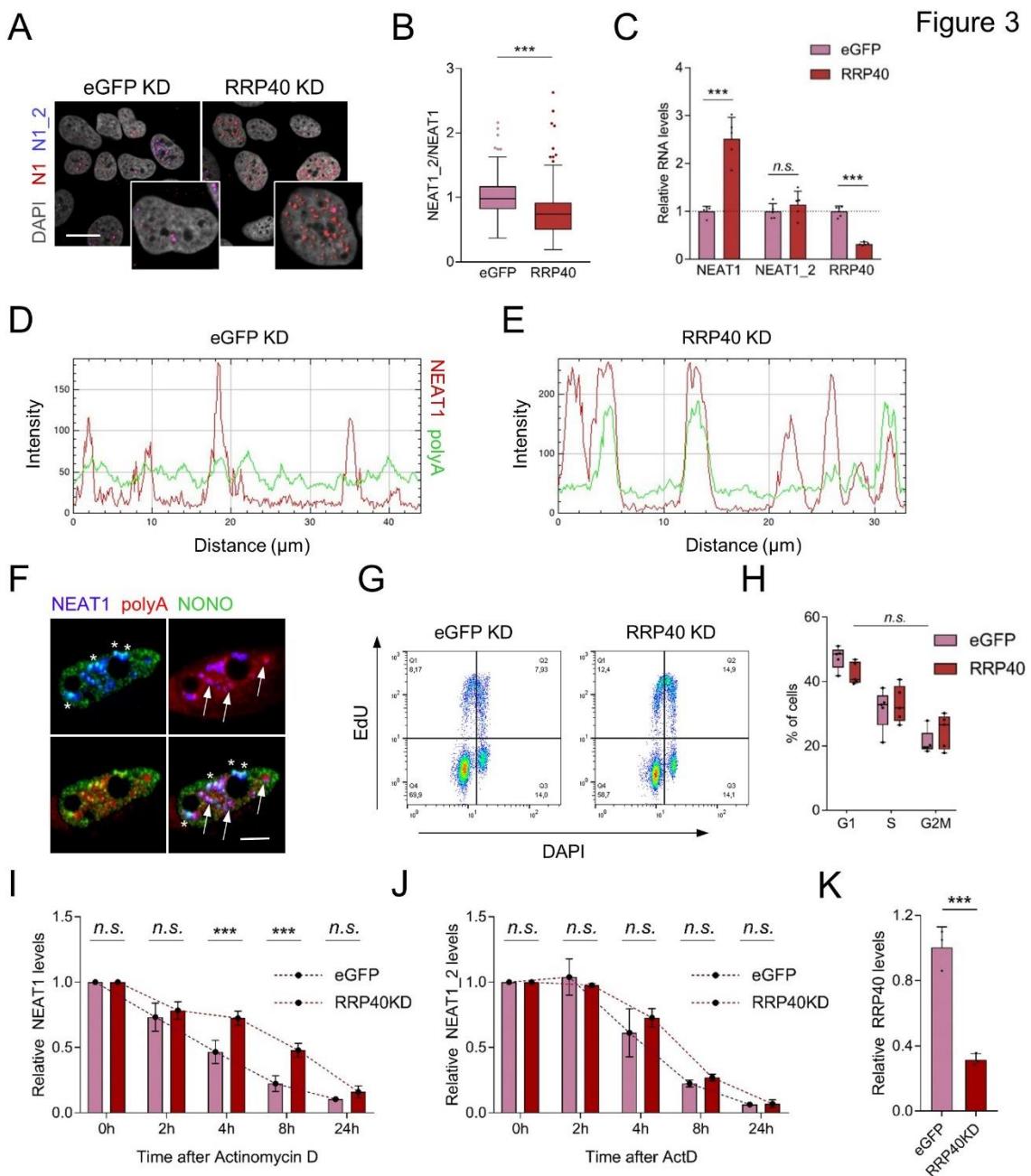


Figure S4: NEAT1_1 is rapidly degraded by the RNA exosome. A. RNA-seq tracks of the NEAT1 locus in which RRP40 is knocked down. B. Image from which the line plot in figure 3D is derived. C. Image from which the line plot in figure 3E is derived. White lines represent lines along which the intensities of both channels are measured.

To further confirm that the observed upregulation of *NEAT1_1* was due to a decrease in RNA degradation rather than a transcriptional (or other processing) effect, we measured *NEAT1* and *NEAT1_2* levels by RT-qPCR in RRP40 knockdown cells at different time points following exposure to the transcriptional inhibitor Actinomycin D. The pace of *NEAT1_1*, *but not NEAT1_2*, decay was significantly delayed in RRP40 KD cells as compared to control cells (figure 3I-K). Collectively, these results indicated that *NEAT1_1* is specifically degraded by the RNA exosome.

Figure 3: NEAT1_1 is rapidly and specifically degraded by the RNA exosome. A. Representative NEAT1/NEAT1_2 RNA-FISH images of U2OS cells in which RRP40, a catalytic subunit of the RNA exosome complex, was knocked down. B. Tukey plots of quantified RNA-FISH signal shown as NEAT1_2 nuclear intensity over total NEAT1 nuclear intensity in $N = 5$ independent experiments. Significance was calculated by an unpaired Mann-Whitney U non-parametric test with each cell as a data point combining data from 5 independent experiments. ***, two-tailed $p < 0.001$. C. Mean and standard deviation of relative RNA levels by RT-qPCR of the same experiments as in B ($N = 5$) in eGFP and RRP40 KD cells suggesting specific upregulation of the *NEAT1_1* isoform. Dots are individual data points. D-E. Line plots of the intensity in eGFP KD (D) and RRP40KD (E) conditions showing NEAT1 localization in polyA+ accumulations upon exosome inhibition. F. NEAT1 (blue) / polyA (red) RNA-FISH and IF in HeLa cells showing NEAT1/polyA+ accumulations in RRP40 KD conditions (arrows) are distinct from paraspeckles (NEAT1 + NONO, asterix). G. Representative flow cytometry graphs of EdU/DAPI staining in eGFP and RRP40 KD conditions. H. % of cells in G1, S and G2M phases of the cell cycle upon eGFP and RRP40 KD as in A-C. $N = 5$ independent experiments. Non-significance was determined using a 2-way ANOVA comparing eGFP and RRP40 conditions in each of the phases with Dunnett's correction for multiple testing. I-J. Relative RNA-levels of both NEAT1 isoforms (I) and the long NEAT1_2 isoform specifically (J) upon addition of 2 mg/ml of the transcription inhibitor Actinomycin D 48h after eGFP or RRP40 KD in HeLa cells. K. Bar graph showing relative RNA levels of RRP40 for the experiment in I-J at time = 0h. Statistical significance was tested using a 2-way ANOVA with Dunnett's correction for multiple testing for $N = 3$ independent experiments. Bars and error are mean and standard deviation. Individual data points are independent experiments.

Figure 3



NEAT1_1 does not contribute to cell growth

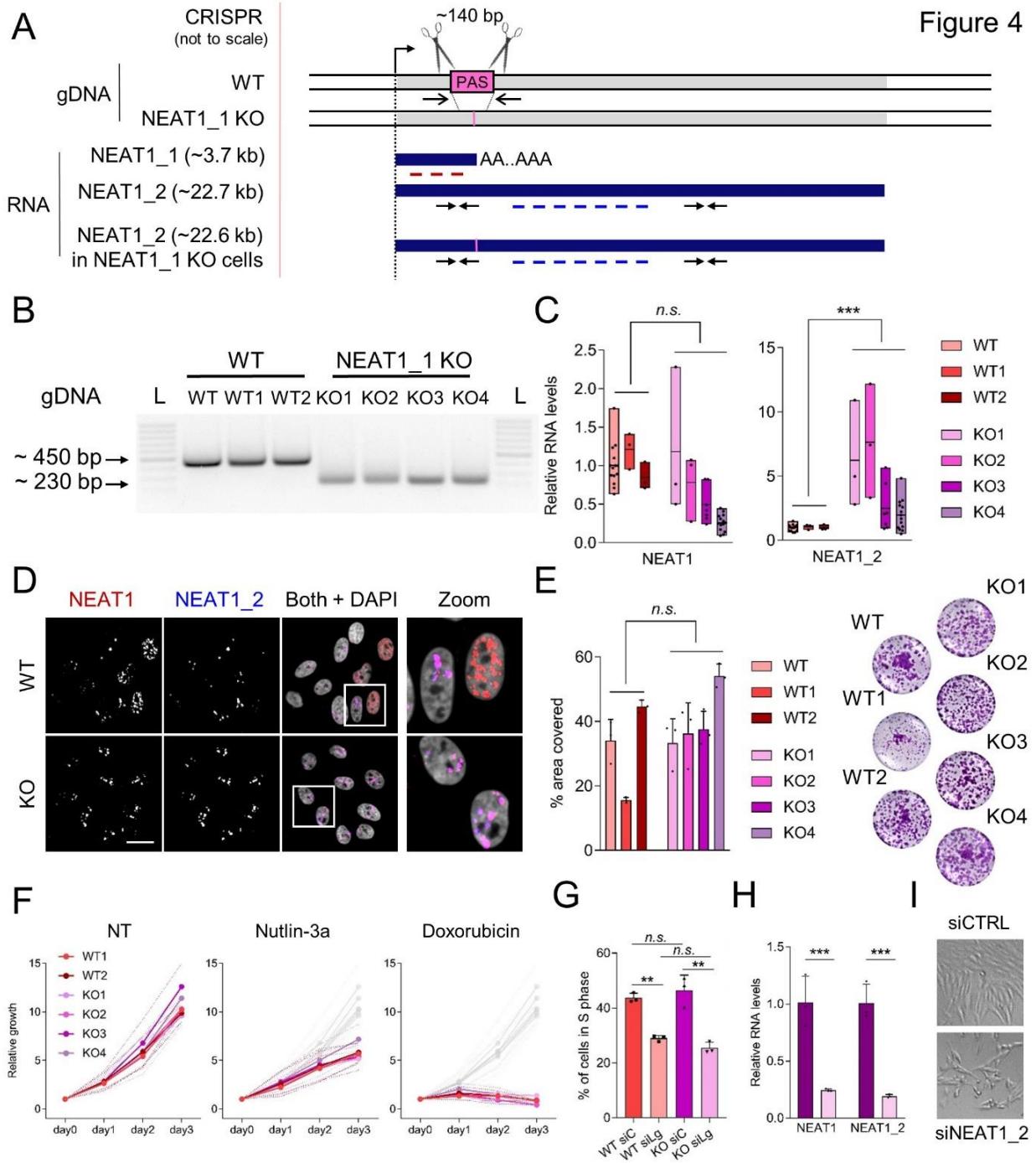
To unravel a putative role for *NEAT1_1* as a regulator of cell cycle progression and/or survival of G1 cells, we used CRISPR editing to delete a small regulatory region (~140 bp) at the 3' end of *NEAT1_1* spanning the CFIIm and hnRNP K binding sites, as well as the polyadenylation signal (PAS). This approach is expected to KO specifically *NEAT1_1* by allowing constitutive read-through, and a switch to *NEAT1_2* expression. This deletion was introduced in U2OS and two other cancer cell lines, HCT116 p53 WT and its isogenic p53 KO line (figure 4A, S5A-B) and single-cell clones were isolated. PCR-based genotyping confirmed successful knockout of the *NEAT1_1* regulatory region in both alleles. 2 wild type (WT) and 4 *NEAT1_1* knock out

(KO) clones for U2OS (figure 4B), as well as 4 WT and 4 KO clones for each of the HCT116 cell lines were generated (figure S5A-B).

To establish that these cells effectively did not express *NEAT1_1*, we quantified total *NEAT1* and *NEAT1_2* levels. Whereas total *NEAT1* levels did not change in the PAS KO clones as compared to WT cells, the levels of *NEAT1_2* were increased, consistent with the prediction that all initiated transcripts contribute to expression of the long isoform (figure 4C, S5E-F). In contrast to the WT controls, RNA-FISH did not detect *NEAT1_1* in the nucleoplasm of the KO clones (figure 4D, figure S5C-D), confirming selective depletion of *NEAT1_1* in the PAS KO cells. Importantly, PS integrity was preserved in *NEAT1_1* KO cells, as evidenced by co-staining these cells with the PS marker NONO and a *NEAT1* RNA FISH probe set (figure S6).

Figure 4: Loss of NEAT1_1 in proliferating cancer cells does not impact on cell growth. A. Scheme of the CRISPR strategy used to knock out *NEAT1_1* by deletion of the regulatory sequences and polyadenylation signal at the 3' end of the *NEAT1_1* short isoform genomic sequence and the resulting RNAs in wild type (WT) and *NEAT1_1* knock out (KO) cells. Closed arrows and dotted lines represent RT-qPCR primers and RNA-FISH probes respectively. Open arrows on the DNA sequence represent approximate locations of genotyping primers. B. Representative inverted gel image of the PCR product from U2OS gDNA used for genotyping individually isolated single cell clones after CRISPR with the primers depicted in A. L = DNA ladder. C. Relative RNA levels of *NEAT1* (total) and *NEAT1_2* specifically showing upregulation of *NEAT1_2* whereas the total levels of *NEAT1* remain the same in WT (shades of red) vs KO (shades of pink and purple) clones. WT without a number is the mother population from which the WT and KO clones were derived. Significance was calculated using a 2-way ANOVA comparing RNA levels in WT and KO clones with Dunnett's correction for multiple testing. ***, p<0.001 for N = at least 3 independent experiments. n.s., not significant. Dots represent data points from each independent experiment. D. Representative *NEAT1/NEAT1_2* RNA-FISH image of WT and *NEAT1_1* KO cells showing the complete loss of *NEAT1_1* upon polyA site knockout. Scale bar, 10 µm. E. Quantification of % area covered (left) and representative images (right) of colony assays 14 days after seeding 2000 cells per well in N = 3 independent replicates, 3 wells per replicate each. Statistical testing was done using a 1-way ANOVA. n.s.: not significant. Dots represent average of 3 wells of the independent experiments. Bars are mean + standard deviation. (Legend continued on the next page)

Figure 4



(continued from previous page) F. Short term growth measured by WST-1 relative to day 0 (1 day after seeding) in non-treated (NT), 10 μ M Nutlin-3a and a low dose of doxorubicin (150 ng/ml) in NEAT1_1 WT and KO clones. All data is the average of $N = 3$ independent experiments. Standard error is depicted as dotted lines above and below the data points. In Nutlin-3a and doxorubicin conditions, values in non-treated conditions are shown as light grey lines in the back of the graph. All data is not significant as tested by 2-way ANOVA with Dunnett's correction for multiple testing in the different time points. G. Quantification of % of EdU positive cells (S phase) in flow cytometry upon CTRL (siC) and NEAT1_2 (siLg) knockdown in WT and KO clones. $N = 3$ independent experiments. Significance was determined using 2-way ANOVA with Dunnett's correction for multiple comparisons. **, $p < 0.01$. n.s.: not significant. I. representative image of siCTRL and siNEAT1_2 KO cells showing decreased cell density 48h after transfection.

We next assessed whether *NEAT1_1* deletion affected the growth and proliferation of the mutant cells. Long-term growth assay indicated that *NEAT1_1* KO cells proliferated at a similar rate as the wild type controls (figure 4E, figure S5J-M). We confirmed these results in a short-term growth assay (WST-1) and following exposure to Nutlin-3a or a low dose of the DNA damaging agent doxorubicin. In contrast to transient depletion of *NEAT1_2*, which sensitized cells to these agents, *NEAT1_1* KO cells were unaffected (figure 4F, S5G-I). These results indicated that *NEAT1_1* is not required for the 2D growth and proliferation of cancer cell lines.

To determine whether *NEAT1_2* functions independently of *NEAT1_1*, we knocked down *NEAT1_2* in the PAS KO cells and analyzed their cell cycle distribution and growth properties. *NEAT1_2* knockdown in these cells induced a similar decrease in EdU-positive cells as it did in WT cells (figure 4G-I). Cell density was also markedly decreased upon *NEAT1_2* KD both in control and PAS KO cells. These data indicated that *NEAT1_1* does not contribute to the ability of *NEAT1_2* to promote the growth of cancer cell lines.

Figure S5: NEAT1_1 KO in HCT116 does not induce growth phenotypes. A-B. Gel images of genotyping PCR using the primers in figure 4 for single-cell derived clones from HCT116 p53 wild type (A) and the isogenic p53 KO colon carcinoma cell line using 2 different guide RNA sets. C-D. Representative NEAT1/NEAT1_2 RNA-FISH images for the HCT116 clones in A (C) and B (D). E-F. RT-qPCR analysis of NEAT1 total and NEAT1_2 long isoforms. Dots represent individual data points of independent replicates. Statistical significance was calculated with 1-way ANOVA and Tukey's correction for multiple testing with ***, p<0.001 and n.s., not significant. G-I: WST-1 analysis of short term growth in HCT116 WT (G), p53 KO gRNA set 1 (H) and set 2 (I) derived clones. J-M: long term growth analysis (2 weeks) of clones in A & B (J, L respectively) and their representative well images after staining (K, M, respectively). Cells seeded in J-K: 6000. Cells seeded in L-M: 2000. Statistical significance was calculated with 1-way ANOVA and Tukey's correction for multiple testing with n.s., not significant.

Figure S5

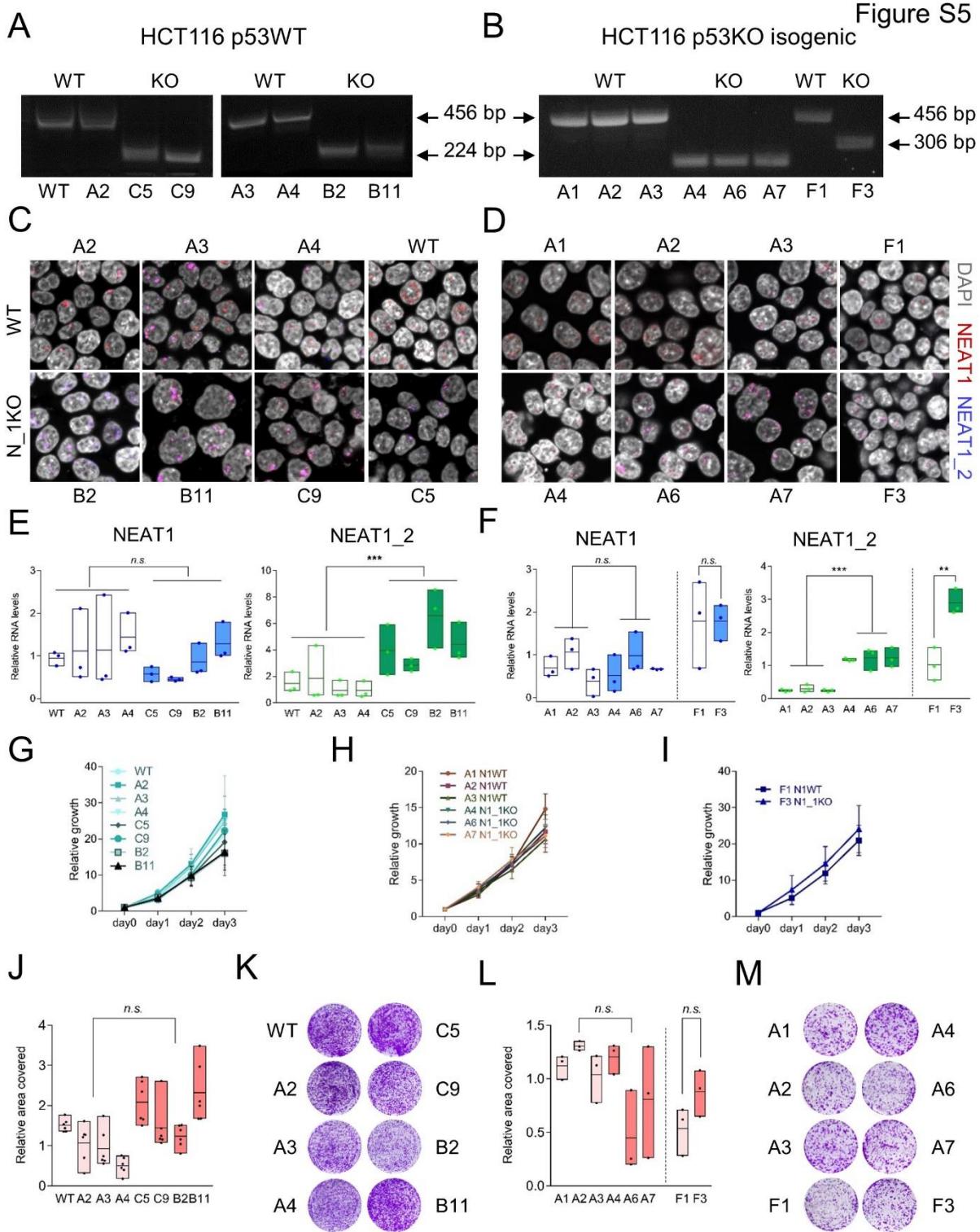


Figure S6

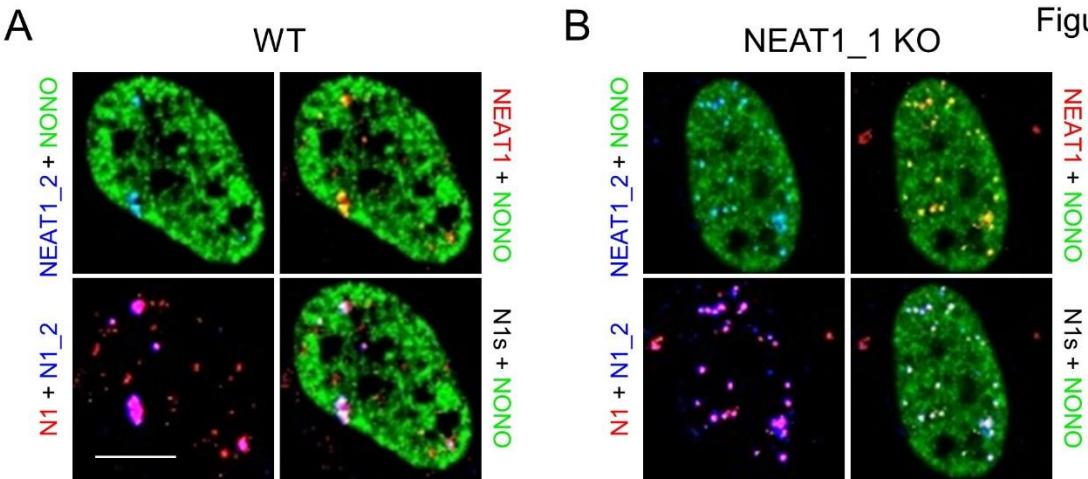


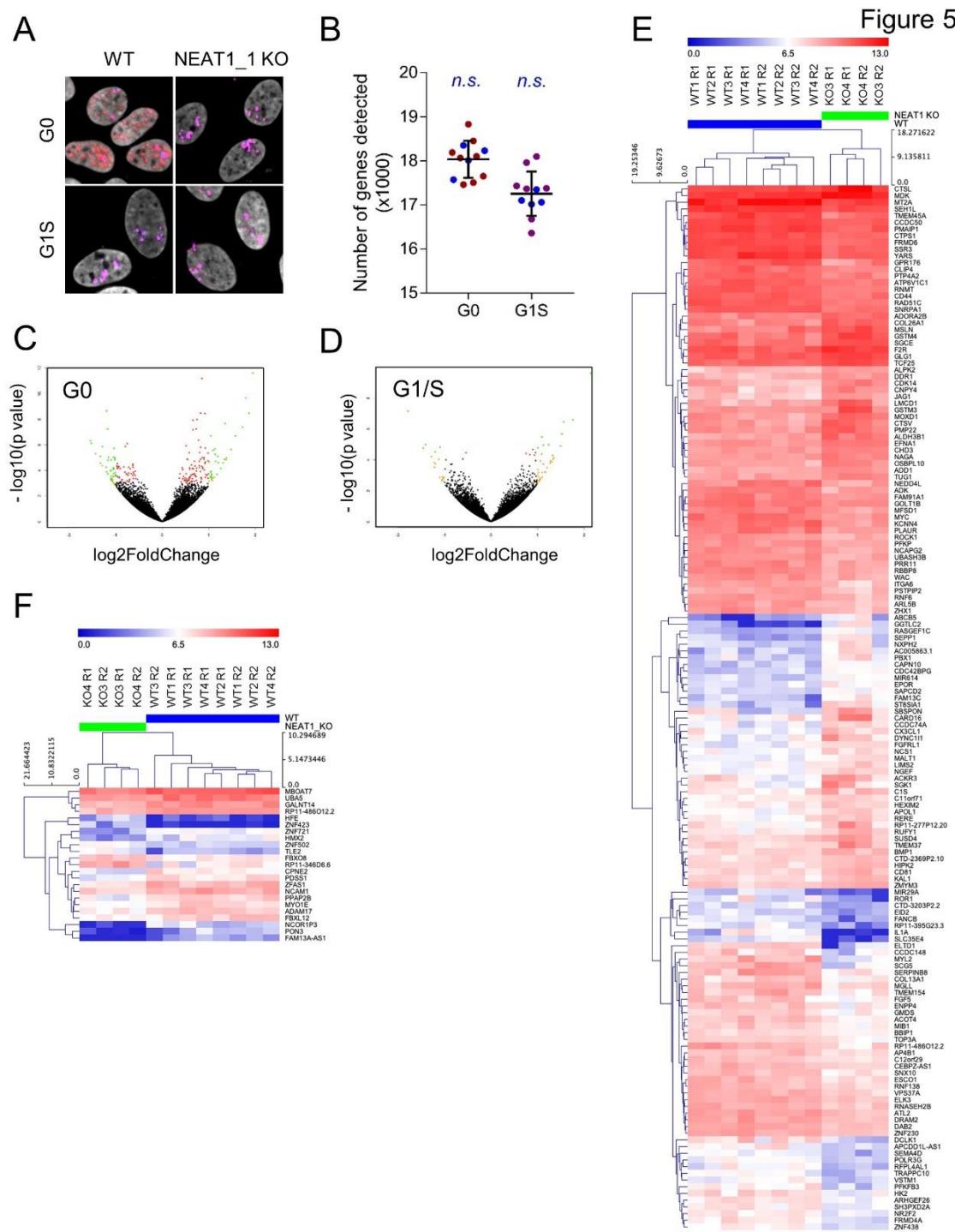
Figure S6: NEAT1_1 KO cells make genuine paraspeckles. A. NEAT1/NEAT1_2 RNA FISH in wild type (WT) U2OS cells in G0 conditions in which the short isoform becomes prominently observed combined with IF for the paraspeckle protein NONO showing NEAT1_2 but not NEAT1_1 colocalization with the paraspeckles. B. Same as in A, but in NEAT1_1 KO conditions lacking the short isoform. N1 = NEAT1; N1_2 = NEAT1_2. Scale bar, 10 μ m.

NEAT1_1 depletion does not overtly impact on the cellular transcriptome

It has been proposed that NEAT1 regulates gene expression by localizing to transcription start sites of actively transcribed genes¹⁶⁸. In order to test whether NEAT1_1, which predominantly localizes to the nucleoplasm of G0/G1 cells, modulates gene transcription we profiled the transcriptome of G0/G1 NEAT1_1 KO and WT cells by RNAseq (figure 5A). We detected on average 18.030 (G0) and 17.250 (G1S) expressed genes, respectively (figure S5B), of which only 156 (~0.86%) and 23 (~0.13%) genes were significantly differentially expressed (DE) in the PAS KO cells compared to WT. Gene ontology analysis did not identify particular pathways or biological processes affected collectively by the depletion of NEAT1_1. Thus, although NEAT1_1 is highly expressed in G0/G1 cells, its loss does not seem to impact significantly on the overall gene expression profiles of resting cancer cells.

Figure 5: NEAT1_1 KO only causes limited changes in gene expression. A. Representative NEAT1/NEAT1_2 FISH in G0 and G1S conditions as in Figure 2 A-C in WT and NEAT1_1 KO cells used for the Smartseq2 RNA sequencing experiment. B. Number of genes detected in G0 and G1S conditions. Red/pink dots are WT conditions, blue dots represent KO conditions. Significance was tested using a two-sided unpaired t-test comparing the number of genes detected in WT and KO conditions. C-D. Volcano plot of gene expression changes (-log2) in G0 (C) and G1S (D) plotted against their p-value (-log10). Dots are color coded red if the adjusted p<0.05, orange if log2 fold change >1 and green if both. E. Hierarchical clustering of G0 samples based on 156 unique differentially expressed genes (FC>1.5, p-adj<0.05). F. Hierarchical clustering of G1/S samples based on 23 unique differentially expressed genes (FC>1.5, p-adj.<0.05).

Figure 5



Neat1_1 KO mice do not exhibit lactation and fertility defects

To further assess the physiological function of *Neat1_1* in normal cells and in the relevant *in vivo* context, we generated a *Neat1_1*-specific KO mouse strain using a strategy similar to the one used to knock out *NEAT1_1* in the cancer cell lines. In brief, 39 basepairs surrounding the polyadenylation signal (PAS) of *Neat1* were excised using CRISPR/Cas9 in mouse Embryonic Stem Cells (mESCs), which were injected into blastocysts and implanted in pseudo-pregnant females. The resulting chimeric mice were back-crossed to generate the final mouse strain deficient for *Neat1_1* (Courtesy of S. Nakagawa and T. Hirose, manuscript in preparation).

Figure S7

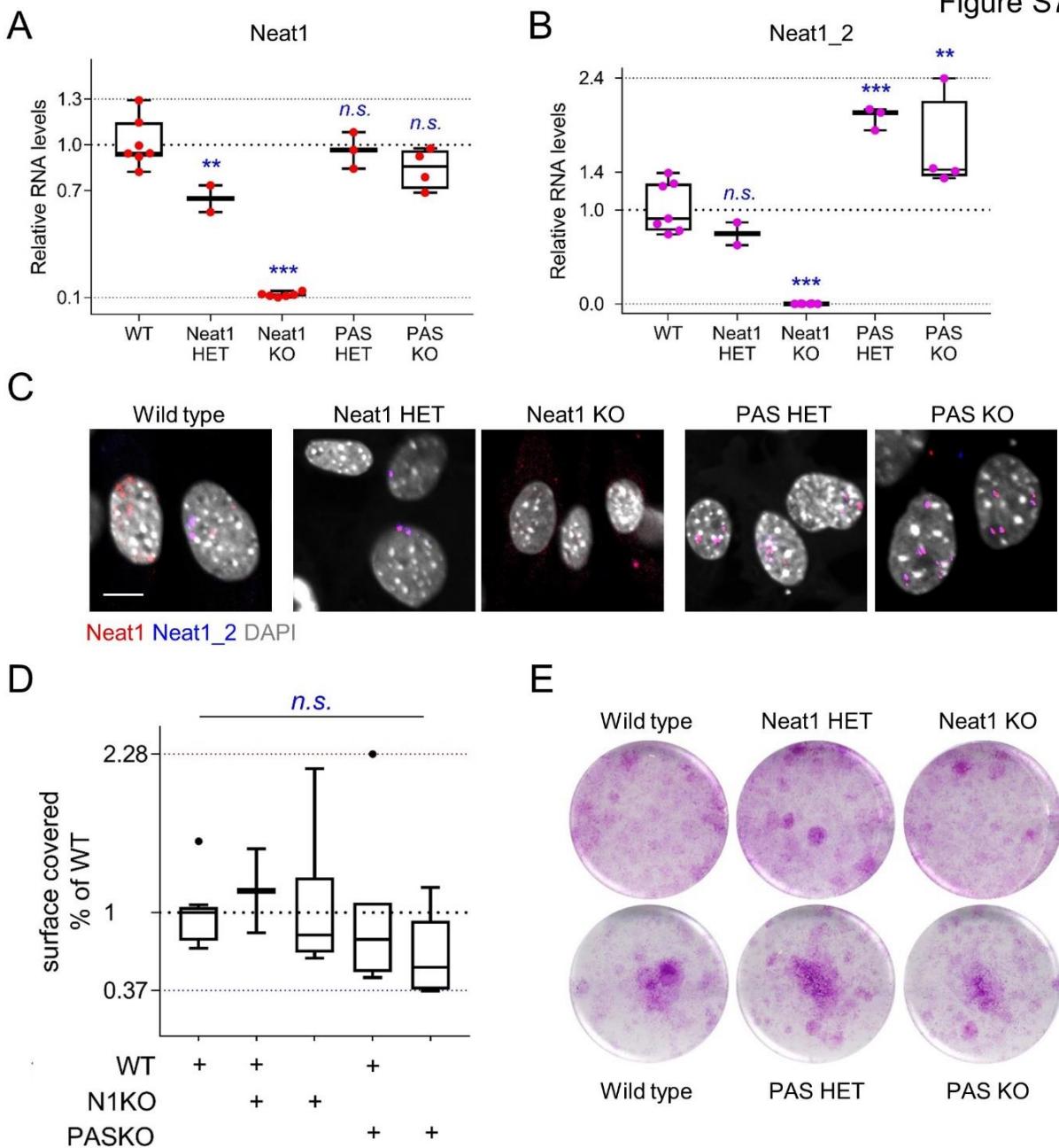


Figure S7: Loss of Neat1_1 or both isoforms in untransformed passage 3 mouse embryonic fibroblasts (MEFs) does not affect cell growth. A-B. Min to max box plots of relative RNA levels by RT-qPCR for both Neat1 isoforms (A), and the long Neat1_2 isoform (B) in MEFs with the respective genotypes. C. Representative images from Neat1/Neat1_2 RNA-FISH in these MEFs. Scale bar, 10 μ m. D-E. Tukey plots (D) and representative wells (E) of long term growth assay with 5000 cells seeded and stained after 2 weeks with $N =$ at least 3 independently derived MEFs per genotype. Each quantification is the average of 3 wells per MEF clone. Statistical testing was done on biological replicates using a 1-way ANOVA with Sidak's correction for multiple testing. n.s., not significant.

PAS KO mice were born at the expected Mendelian ratios (figure 6D) and did not exhibit the previously reported lactation defect observed in *Neat1* full KO¹⁶⁹. Indeed, pups born in nests from wild type, *Neat1* full KO, *Neat1* heterozygous, PAS heterozygous and PAS KO mothers at three and six weeks of age. We confirmed that full KO females were unable to successfully nurture their pups. In contrast, PAS KO females gave birth to normally sized nests and all offspring developed and gained weight normally (figure 5E-F).

To further study the impact of the mutation on the growth of normal cells *in vitro*, we produced PAS KO MEFs. Despite an increase in *Neat1_2* levels, passage 3 MEFs derived from the PAS KO mice grew similarly to WT MEFs. Full *Neat1* KO also grew similarly to WT fibroblasts (figure S7)^{141,165}.

Mouse Neat1_1 does not contribute to DNA damage induction and reduced growth during skin carcinogenesis

The skin of *Neat1* KO exhibits an exacerbated sensitivity to DNA damage and, thereby, an increased resistance to DMBA/TPA-induced skin hyperplasia and tumorigenesis¹⁴¹. To test whether *Neat1_1* mice exhibit a similar phenotype, we subjected these mice to the DMBA-TPA protocol and assessed PS formation, measured hyperplasia and accumulation of DNA damage in their treated back skin. We found that after 11 days of treatment, both PAS KO and WT cells displayed abundant PS (figure 5I), and displayed moderate to severe hyperplasia (figure 5H). In contrast, full *Neat1* KO back skin did not display PS nor hyperplasia (figure 5H-I). In contrast, *Neat1* KO back skin showed a significantly increase in persisting DNA damage in the treated regions as compared to the skin of WT and PAS KO animals (figure 5J-K). We concluded that the phenotypes observed in *Neat1* KO mice are attributable to the loss of the long *Neat1_2* isoform and, thereby likely to be a consequence of loss of PS nuclear bodies.

Figure 6

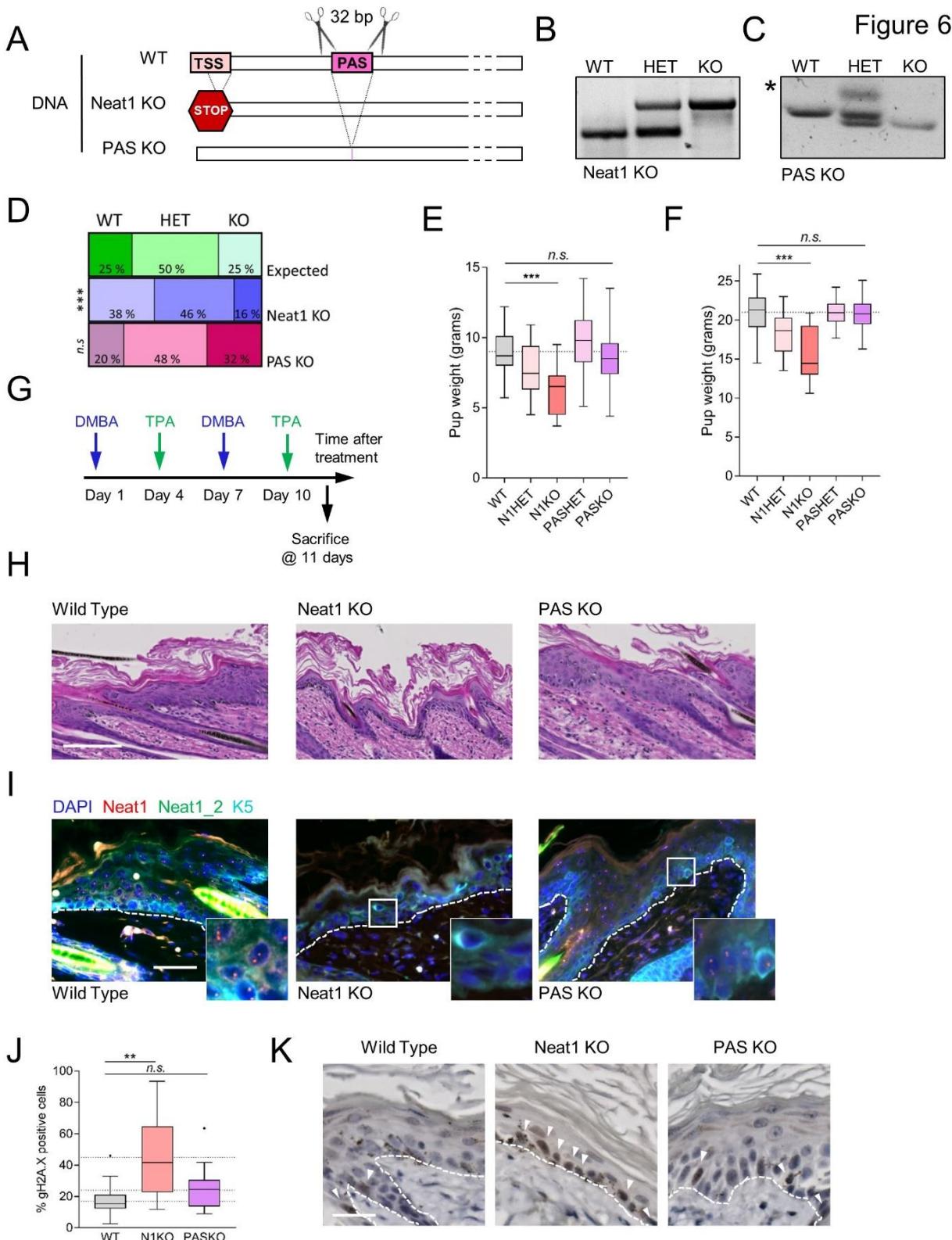


Figure 6: Neat1 full KO phenotypes are due to lack of Neat1_2, not Neat1_1. A. CRISPR strategy to knock out Neat1_1 in mouse embryonic stem cells resulting in a 39 base pair deletion spanning the Neat1_1 polyadenylation signal (PAS) and strategy to knock out both isoforms as described in Nakagawa et al., 2011. B. Genotyping of Neat1 full KO mice. C. Genotyping of Neat1_1 (PAS) KO mice. The asterix indicates an unspecific band in the heterozygous sample. (Legend continued on the next page).

(Continued from the previous page) D. Genotype distribution of pups born from heterozygous parents in full Neat1KO (middle bar) and PAS KO nests (lower bar) as compared to the expected mendelian ratios (upper bar). N = 53 and 15 litters for Neat1 and PAS KO, respectively. P-values were calculated using the chi-squared test. ***, p<0.001. n.s., not significant E-F. Pup weight of offspring from females with the respective genotypes at 3 (E) and 6 (F) weeks of age. Tukey plots of pups from N = between 4 and 19 females per genotype. Statistical significance was calculated using 2-way ANOVA with Dunnett's correction for multiple testing. ***, p<0.001. n.s., not significant. G. Strategy and time line for short-term DMBA/TPA carcinogenesis protocol. H. Representative H&E staining of back skin sections from mice treated as in G. Scale bar, 50 μ m. I. NEAT1 RNA-FISH and Keratin 5 immunofluorescence on back skin section of mice treated with DMBA and TPA as in F. J-K. Quantification (J) and representative images (K) of sections from DMBA/TPA treated back skin stained with the DNA damage marker γ -H2A.X of N = at least 3 mice per genotype according to the method described in Adriaens et al., 2016. The Tukey plot in J graphs individually quantified pictures. Statistical testing was done on biological replicates (averages for individual mice) using a 1-way ANOVA with Sidak's correction for multiple testing. **, p<0.01. n.s., not significant. White arrows in K indicate γ -H2A.X positive cells, white dotted lines separates the dermis from the epidermis. Scale bar, 20 μ m.

Discussion

In this work, we showed that the expression levels of the two *NEAT1* isoforms are dynamically regulated during the cell cycle. We observed that the short isoform, *NEAT1_1* is highly expressed in G0/G1-phase of the cell cycle and that, in line with previous findings¹⁶⁷, it localizes prominently outside of PS. We also observed that *NEAT1_1* levels drop abruptly as cells transit from the G1 to the S phase. This is consistent with the observation that *NEAT1_1* is detected at high levels in terminally differentiated cells in most tissues¹⁶⁵. In contrast, the levels of *NEAT1_2* remained relatively constant throughout the cell cycle and consequently *NEAT1_2* was the only detected *NEAT1* isoform in S-phase onwards. This is in keeping with previous data reporting the presence of PS in amitotic (interphase) cells as evidenced by the typical clustering of canonical PS protein p54nrb/NONO in their nuclei¹⁵⁰.

It had been proposed that the short *NEAT1* isoform is recruited into PS to support their stability and/or functions^{13,15,163}. In contrast, we showed that when both isoforms are co-expressed, in G0/G1, a large fraction of the *NEAT1_1* transcripts localizes outside of PS. In cells residing in other phases of the cell cycle, *NEAT1_2* is the only isoform present and therefore PS are, by and large, *NEAT1_1*-free in these cells. Our data is in line with previous quantifications of *NEAT1_1* molecules indicating that, on the basis of stoichiometry, *NEAT1_1* does not likely locate to PS or at least not in significant amounts^{167,171}. Our data also indicate that PS can be assembled in absence of *NEAT1_1* (in mitotic cells). This observation resonates with previous work showing that *NEAT1_2* is the *NEAT1* isoform that is required for PS assembly¹⁵⁶ and that *NEAT1_1* expression alone is not sufficient to rescue PS formation upon *NEAT1_2* knockdown¹⁵⁵ or in *NEAT1_2* deficient cells¹⁵.

Our data indicated that the *NEAT1_1* transcript is actively degraded as cells commit to divide. We provide evidence that this process is mediated by the main RNA degradation machinery in the nucleus, namely the RNA exosome²⁵⁷. Knockdown of one of its core components, RRP40, led to the specific accumulation of the *NEAT1_1* short isoform within nuclear bodies containing persistent polyA+ RNA transcripts⁶⁸. Importantly, we provided evidence that the RNA exosome only targets *NEAT1_1* (and not *NEAT1_2*). *NEAT1_1*/polyA+-containing bodies were clearly distinct from PS, an observation that is consistent with *NEAT1_1* being an entity that is spatially distinct from PS.

The observations that the evolutionarily conserved *NEAT1_1* isoform is (i) produced at very high levels in most -if not all- resting cells (and cancer cells in G1), where it localizes primarily in the nucleoplasm, outside of PS, and (ii) gets actively degraded as cells commit to divide indicated a putative role for *NEAT1_1* in the regulation of the cell cycle.

However, silencing of RRP40, which led to accumulation of *NEAT1_1*, did not overtly perturb the progression of the cell cycle. Consistently, previous reports showed that neither modulation of RRP40 and its targets^{56,86} nor *NEAT1_1* overexpression affect cell cycle progression (reviewed in Yu et al. 2017). More strikingly, *NEAT1_1* isoform specific KO cells, generated using a similar CRISPR genome editing strategy as described in^{161,167}, grew and responded to stress comparably to wild-type control cells. PS formed normally in these cells. Moreover, the phenotypes observed upon silencing of *NEAT1_2* in various cancer cell lines -such as cell cycle arrest- was also observed in these *NEAT1_1* KO cells. Together, these data indicated that *NEAT1_2* functions independently of *NEAT1_1* and that *NEAT1_2* can promote PS formation in absence of the short *NEAT1* isoform,

Our data showing that *NEAT1_1* localizes outside of PS is consistent with another report showing that there is a fraction of *NEAT1* that localizes diffusely throughout the nucleoplasm¹⁶⁷ and the possibility that *NEAT1* binds active transcription start sites in euchromatin¹⁶⁸. However, only minor changes in the gene expression patterns were observed in *NEAT1_1* KO cells indicating that *NEAT1_1* does not directly impact on gene expression. Instead, we hypothesize that its enrichment to euchromatin might have spatial, physical grounds (i.e. diffusion is easier in euchromatin due to a decrease in molecular crowding) rather than a specific functional role in modulating gene expression. Moreover, in contrast to *Neat1* KO¹⁷⁰, *Neat1_1* KO mice were born at the normal Mendelian ratio. Pups born from *Neat1_1* KO females were not significantly smaller than those from wild type and heterozygous mothers. This is in contrast to our previously published observation that *Neat1* KO females cannot nurture their pups properly owing to mammary gland formation and lactation defects¹⁶⁹. Similarly, *NEAT1_1* KO mice did not exhibit any of the phenotypes observed in *Neat1* KO mice exposed to a 2-step skin carcinogenesis protocol¹⁴¹. Taken together, our data show that *NEAT1_1* is a non-functional, non-essential isoform in both resting and proliferative cells, at least in the experimental conditions used in these studies. This raises the possibility that the highly abundant and conserved short isoform may just be a non-functional byproduct of the *NEAT1* locus and that the only functional unit produced by this locus is the long PS-associated *NEAT1_2* isoform. One interesting speculative hypothesis is that active transcription of the *NEAT1* locus may ensure that cells can rapidly produce *NEAT1_2* in response to stress (by inducing read-through), and thereby the formation of PS. Constant synthesis of a nonfunctional *NEAT1_1* transcript would be the price that cells have to pay to be able to quickly engage a PS-dependent survival pathway when exposed to deleterious stimuli. However, It cannot be fully excluded that *NEAT1_1* does exert a function in very specific stress and/or pathophysiological conditions, such as for instance, during viral infections^{184,189,191,259–261}. The enclosed described *NEAT1_1* mouse model will be a valuable tool to test this possibility.

Materials and methods

Cell lines, culture methods and cell synchronization

U2OS, HeLa and HCT116 WT and p53KO isogenic cell lines were obtained from the LGC ATCC and maintained in DMEM (ThermoFisher Scientific Cat. No. 21885025) plus 10% fetal bovine serum (FischerScientific Cat. No. 10270106). None of the cell lines used were reported in the ICLAC database of commonly misidentified cell lines. All cell lines were tested monthly for mycoplasma contamination and found negative. After their initial purchase, cell lines were not further authenticated. For synchronization in G0, the cells were washed with PBS 24 hours after plating and media were replaced with media containing no serum for three days. For G1/S synchronization, G0 cells were released in 20% serum plus 5 µg/ml aphidicolin (Sigma, A0781) for 24h. For G1, S and G2 synchronization, media was replaced with media containing 2 mM thymidine (Sigma, T1895) for 12h, released in normal growth media for 12h and then grown again in thymidine-containing media for 12h prior to release in normal media and harvesting at the indicated time points. For knockdown experiments with siPOOLs (siTOOLS Biotech), 25 nM siRNA was transfected using Lipofectamine RNAiMAX (ThermoFisher Scientific, Cat. No. 13778075) as previously described¹⁴¹.

Generation and culture of Mouse Embryonic Fibroblasts

Mouse Embryonic Fibroblasts were generated from plug-checked pregnant females at E12.5. The embryos were removed from the uterus and internal organs were discarded with sterile forceps. Heads were used for genotyping. The remainder tissue was pipetted up and down several times in sterile PBS to obtain single cell suspensions before transfer to tissue culture dishes with DMEM containing 10% serum, 1% penicillin/streptomycin (Invitrogen, Cat. No. 15140122) and 50 µM β-Mercaptoethanol (ThermoFisher Scientific, Cat. No. 31350010). The cells were passaged twice before all experiments were performed at passage 3.

RNA-FISH

Dual RNA-FISH was performed according to the Stellaris RNA-FISH protocol essentially as described in ¹⁴¹ but with probesets against both (Cat. No. VSMF-2246-5 and VSMF-2247-5) or NEAT1_2 specifically (Cat. No. VSMF-2251-5) in human cells and against Neat1 both (Cat. No. VSMF-3030-5) or Neat1_2 specifically (Cat. No. VSMF-3035-5) in mouse cells and mouse tissues. Cells were generally plated on 11 mm coverslips in 6-well plates to allow for concomitant RT-qPCR analysis, cell cycle analysis and RNA-FISH. Costaining with the paraspeckle marker p54nrb/NONO ¹³ was done after the ethanol permeabilization step of the RNA-FISH protocol. Briefly, the cells were washed once with PBS followed by a 5' permeabilization step with 0.5% Triton-X100 in PBS. Then, the cells were incubated with a 1/1000 dilution of the antibody in DAKO antibody dilution reagent for 1h at room temperature

followed by three washes in 0.05% Tween-20 in PBS and staining with the secondary antibody (Life Technologies, anti-mouse A488) in DAKO for 1 hour at RT. After two short washes in 0.05% Tween-20 in PBS, the cells were incubated with Wash Buffer (2x SSC, 10% v/v Formamide (Sigma Aldrich, Cat. No. F9037) and the RNA-FISH protocol was continued as described. Hybridization buffer was made using the same formula as wash buffer, adding 10% w/v Dextran (Sigma Aldrich, Cat. No. D8606) and probes at a final concentration of 25 nM. Images were acquired with a Nikon A1 confocal microscope acquired through a Hercules grant type 1 AKUL/09/037 and processed for overlay and brightness and contrast adjustments using ImageJ. RNA-FISH images from mouse back skin tissue were acquired with a ZEISS Axio Scan Z1 microscope using 20x and 40x objectives followed by stitching of the continuous fields using ZEN2 software.

Image analysis

Confocal images were quantified using FIJI software (ImageJ 1.51p. Java version 1.8.0_66, 64-bit, National Institutes of Health, USA). To determine the number of cells that display NEAT1_1 outside of paraspeckle nuclear bodies, we processed the raw images with the Speckle Inspector plugin on each channel after thresholding, with a minimal speckle size 2 pixels, within the nuclear region delineated by the DAPI channel. The number of spots in the NEAT1_2 channel was subtracted from the number of spots per cell in the NEAT1 channel. If the outcome of the subtraction was larger than 3 (arbitrary error margin: 0 (expected) +3 to account for accidental background spots in the Q570 channel), we considered that the cells contained detectable NEAT1_1 outside of the paraspeckles. The total numbers of cells were determined using the Cell Counter plugin. Nuclear RNA-FISH intensity (Figure 3G, 4B) was calculated by thresholding, filling holes and watershed of the DAPI channel, and determination of the nuclear intensity in the NEAT1 and NEAT1_2 channels per cell via the “send to” functionality in Set Measurements before Counting Particles. The percentage of cells containing detectable NEAT1_1 is represented on the left y-axis, whereas the percentage of cells that only displayed paraspeckles was represented on the right y-axis.

To find the ratios per cell of NEAT1_2 signal over total NEAT1 signal, we used the FIJI function “Set Measurements” as above to redirect DAPI-thresholded images to the respective NEAT1 (Quasar 570, measured in the red channel and represented in red) and NEAT1_2 (Quasar 670, measured in the far red channel and represented in blue) channels to obtain their relative intensities, which were then plotted per cell as NEAT1_2/NEAT1 relative integrated density per cell.

RNA isolation and RT-qPCR

RNA isolation, generation of cDNA and RT-qPCR were performed essentially as described in ¹⁴¹. Briefly, after lysis, the cell-lysis buffer mixture was heated for 10 minutes at 55°C according to the protocol described in Chujo et al., 2017¹⁷¹. Then, total RNA was isolated using the NucleoSpin RNA kit (Macherey Nagel, Cat. No. 740955), including rDNase treatment for 15 minutes according to the manufacturer's instructions. The RNA was reverse transcribed using the ThermoFisher Scientific High Capacity cDNA Reverse Transcription Kit (Cat. No. 4368813). RT-qPCR was performed with GC Biotech SensiFast SYBR No-Rox (Cat. No. BIO-98020) and run on a Roche LightCycler-480-384. For normalization, the geometric mean of the two most stable reference genes out of at least three was calculated using geNorm in qBase+ Software (Biogazelle, Zwijnaarde, Belgium - www.qbaseplus.com). RT-qPCR primer sequences were as follows: *NEAT1* fw: 5'-GGAGAGGGTTGGTAGAGAT-3'; *NEAT1* rev: 5'-CCTTCAACCTGCATTCCTA-3'; *NEAT1_2* fw: 5'-GGCCAGAGCTTGCTGCTTC-3'; *NEAT1_2* rev: 5'-GGTGCAGGGCACTTACTTACT-3'; *CDKN1A* fw: 5'-AGCAGAGGAAGACCATGTGGA-3'; *CDKN1A* rev: 5'-AATCTGTCATGCTGGTCTGCC-3'; UBC fw: 5'-ATTGGGTCGCGGTTCTG-3'; UBC rev: 5'-TGCCTTGACATTCTCGATGGT-3'; TBP fw: 5'-CGGCTGTTAACCGCCTTC-3'; TBP rev: 5'-CACACGCCAAGAACAGTGA-3'; B2M fw: 5'-TGCTGTCTCCATGTTGATGTATCT-3'; B2M rev: 5'-TCTCTGCTCCCCACCTCTAAGT-3'; HPRT1 fw: 5'-TGACACTGGCAAAACAATGCA-3'; HPRT1 rev: 5'-GGTCCTTTCACCAAGCT-3'; m*Neat1_2* fw 5'-GCTCTGGACCTTCGTGACTCT-3'; m*Neat1_2* rev 5'-CTGCCTGGCTTGGAAATGTAA-3'; m*Neat1* fw 5'-TTGGGACAGTGGACGTGTGG-3'; m*Neat1* rev 5'-TCAAGTGCCAGCAGACAGCA-3'; mHmbs fw 5'-GCAGTCATGTCCGGTAA-3'; mHmbs rev 5'-GTGGTGGACATAGCAATGATT-3'; mGapdh fw 5'-AGGTTGTCTCCTGCGACTTCA-3'; mGapdh rev 5'-GGTGGTCCAGGGTTCTTACTC-3'. For correlation analysis, primer efficiencies were calculated in qBase+ by combining cDNA of each of the tested samples and producing a serial dilution (1, 0.5, 0.25, 0.125, 0.0625 and 0.03125) to be run simultaneously with the individual samples.

Cell cycle analysis

Cell cycle analysis was performed by pulsing the cells with 10 µM of EdU for 30 minutes before harvesting and trypsinization, or via DNA profiling alone against a non-synchronized control to identify 2N and 4N populations. For EdU staining, the cells were then washed in cold PBS + 10% serum to inactivate the trypsin, collected by centrifugation and fixed for 15 minutes with 4% PFA in PBS. After two washes in PBS, the cells were stored overnight in 15 mL 0.01% Triton-X100 at 4°C. To detect cells in S phase, the cells were subjected to a modified Click-IT reaction protocol (Click-iT® EdU Alexa Fluor® 488, Cat. No. C10420). Briefly, the cells were

collected by centrifugation, the supernatant was discarded, and they were incubated in 50 µL Click-IT reaction cocktail (43.75 µL PBS + 1 µL CuSO₄ 100 mM + 5 µL 100 mM Ascorbic Acid + 0.25 µL A488 Azide dye) for 50 minutes. The cells were then washed once in PBS and resuspended in 300 µL DAPI staining buffer (5 µg/mL DAPI in PBS with 0.1% w/v BSA) before analysis on a MACSQuant® VYB flow cytometer (Miltenyi Biotec, Germany).

CRISPR/Cas9 plasmid construction

Guide RNAs were designed for the 3' regulatory region of *NEAT1_1* using <http://crispr.mit.edu>. Five µg of plasmid pSpCas9(BB)-2A-GFP (pX458) (Addgene, Cat. No. 48138) was digested with BbsI/BpuI (ThermoFisher Scientific) and purified using a NucleoSpin® Gel and PCR Clean-up Kit (Machery Nagel, Cat. No. 740609) followed by In-fusion cloning of annealed gRNA oligos with 20 nucleotide overhangs on both sides (IDT) with sequences according to the manufacturer's instructions (Takara Bio Cat. No. 121416). The gRNA sequences used to generate 4 different Cas9 targeting plasmids were: upstream Guide #1 5'-GTGTATTAGTCACGCATGTATGG-3' quality score 89; upstream Guide #7 5'-GTACTGGTATGTTGCTCTGTATGG-3', quality score 70; downstream Guide #1 5'-GTACATCCAAAGTCGTTATGAAGG-3', quality score 90; downstream Guide #4 5'-GCGTTATGAAGGCAATGTGATAGG-3', quality score 70. Following in-fusion cloning, the plasmids were transformed into competent bacteria (DH5α) grown on ampicillin plates. A colony PCR was performed to check for the correct insertion of the gRNA sequence using GoTaq Green Mastermix (Promega, Cat. No. M712) and primers 5'-GAGGGCCTATTCCCATGATT-3' (fw) and 5'-AAAAAAAGCACCGACTCGGTGCCA-3' (rev). Positive clones were further expanded and their inserted sequences were verified with Sanger sequencing at the VIB Genomic Service Facility, Belgium using the same primers.

Generation of NEAT1_1 KO cells

Once we obtained the desired Cas9/gRNA constructs, we transfected cells plated in 10 cm dishes with 10 µg of downstream and 10 µg of upstream plasmid (Combination dG1/uG1 for U2OS and HCT116 and dG4/uG7 for HCT116) using a standard transient overexpression protocol with Lipofectamine 2000 reagent according to the manufacturer's instructions (ThermoFisher Scientific, Cat. No. 11668019). 48h after transfection, we sorted the cells for GFP expression using a S3™ Sorter (Bio-Rad Laboratories, USA) and diluted the cells at 0.5 cells/100 µL into 96 well plates. After 2 weeks of culture, we visually inspected the wells and selected those containing a single clone. These were collected and replated in duplicate. The cells in one of the two wells were then lysed and subjected to PCR analysis to determine their *NEAT1_1* genotype with primers 5'-CGTTGGATCTTCTGTCT-3' (fw) and 5'-GCTCTCCTACATGGCCTTAAT-3' (rev). These primers were also used for Sanger sequencing to characterize the repair on each allele in homozygous *NEAT1_1* KO clones.

Several homozygous WT and homozygous KO clones were then selected and expanded into new cell lines from the remaining wells.

Cell growth assays

To determine long term cell growth, cells were plated at the indicated densities in three wells per cell line per experiment and grown for 10 or 14 days. They were washed twice in cold PBS, followed by staining for 15 minutes with 0.5% Crystal Violet (Sigma Aldrich, Cat. No. C6158) in 20% Methanol/80%H₂O. The plates were washed and rinsed in tap water and the % area covered of the wells was quantified using FIJI. For short term growth assays, 1500 cells were plated followed by incubation with WST-1 reagent (Roche, Cat. No. 05 015 944 001) and measurement of the luminescence with a VICTOR X3 Multilabel Plate Reader (PerkinElmer) at the indicated time points. Cells were treated with 10 µM Nutlin-3a (Sigma Aldrich, Cat. No. SML05080) or 150 ng/ml doxorubicin (Sigma Aldrich, Cat. No. D1515).

RNA sequencing

Total RNA was extracted as described above using the NucleoSpin RNA kit (Machery Nagel, Cat. No. 740955). The RNA integrity was monitored using Bioanalyzer analysis (Agilent, RIN: 9.7-10). About 500 pg of RNA per sample was reverse-transcribed and amplified using a modified SMARTseq2 protocol (Rambow et al. 2018). Prior to generating sequencing libraries using the NexteraXT kit (Illumina, Cat. No. FC-131-10), cDNA profiles were monitored using the Bioanalyzer. Sequencing was performed on a Nextseq500 platform (Illumina, SE75bp). Reads were then mapped to the human genome (hg19) using STAR (2.4.1b) and quantified with Subread (1.4.6-p2). Differential analyses between *NEAT1_1* KO and WT samples (during G0 and G1/S) were executed using the DeSeq2 pipeline. Samples were grouped using hierarchical clustering (Euclidean distance) based on differentially expressed genes (MeV4_8_1).

KO mice

Neat1 KO, *Neat1_1* KO and WT mice were maintained on a pure C57BL/6J background in a certified animal facility at KU Leuven Campus Gasthuisberg, Leuven, Belgium. They were maintained on a 12/12h light/dark cycle and had access to food and water ad libitum. All animal experiments were carried out in accordance with the guidelines of the Ethical Committee University of Leuven Animal Care and Use under project license 089/2013. Full *Neat1* KO mice were described previously¹⁶⁵ and genotyped with primers 5'-GGTGACGCGACACAAGAGTA-3' (fw), 5'-AAATGTGAGCGAGTAACAAACCC-3' (rev WT) and 5'-CTGTGAAACTTGTGCCCTCC-3' (rev KO) giving rise to PCR products of 612 base pairs (*Neat1* KO) and 336 base pairs (WT). *Neat1_1* KO mice were generated by S. Nakagawa and T. Hirose using a similar CRISPR-Cas9 strategy as described for the cancer cells above

generating a 39 base pair deletion of the polyadenylation signal (5'-ACAGCAAAATAAAGGTTGAGATTGAAGCTTCTTAGAAT-3') and genotyped with primers 5'-GCAAAGTGACAGAGGTCGAGA-3' (fw) and 5'-AGGCAAAGTGACAGAGGTCG-3' (rev) (WT allele: 145 base pairs; KO allele, 106 base pairs) (Unpublished, manuscript in preparation). In order to test for lactation defects, mice with mothers from the indicated genotypes were weighed at 3 and 6 weeks of age¹⁶⁹. Ratios of animals born at indicated genotypes to test against expected Mendelian genotype ratios were calculated from heterozygous x heterozygous parents in both colonies.

DMBA/TPA protocol

The DMBA/TPA protocol was performed as described in¹⁴¹.

H&E and immunohistochemistry

Immunohistochemistry and quantification of images was performed as described in¹⁴¹ using antibodies against γ-H2A.X (Cell Signaling, 2577; 1/1,400) and Keratin 5 (rabbit polyclonal anti-keratin 5; Covance, PRB-160P-0100; 1/1,000). For immunofluorescence the secondary antibody was anti-Rabbit-A488 (Life Technologies). Images were acquired with a ZEISS Axio Scan Z1 microscope using 20x and 40x objectives followed by stitching of the continuous fields using ZEN2 software.

Author contributions

CA designed the study, performed all experiments and analyzed all data apart from the RNA sequencing analysis. FR performed RNA sequencing analysis. GB performed experiments. TS provided images from RRP40 KD conditions and found the link with *NEAT1_1* specific degradation. TH, MM, TC, AH and SN provided reagents and made the mouse models. JCM and TJH designed and supervised the study. CA and JCM wrote the manuscript with input from all authors.

Competing interests

JCM owns a patent to target *NEAT1_2* in cancer (US9783803B2) and is a cofounder of NewCo, a company aiming to develop oligo-based therapeutics to target cancer. The other authors declare no competing interests.

Acknowledgements

We thank Odessa Van Goethem and Veronique Benne for excellent technical support and maintenance of the mouse colonies. We thank Tom Misteli (NCI/NIH, USA), Michael Dewaele (VIB, Belgium), Paulo P. Amaral (Cambridge University, UK) and Andrew Blackford (Oxford University, UK) for helpful discussions throughout the project. CA was supported by an IWT SB fellowship (no. 141372) from the Belgian Research Foundation - Flanders (FWO) and a Gustave Boël - Sofina travel grant (V433017N) from the Koning Boudewijn Stichting, Belgium. The work in the J-C.M. lab was supported by KUL GOA (#14/012), VLK (Vlaamse Liga tegen Kanker), Interreg (SKiN-HUID), and Fonds voor Wetenschappelijk Onderzoek Vlaanderen (#G.0929.16N). This work was supported by JSPS KAKENHI Grant Number 17H03604 and MEXT KAKENHI Grant Number 26113005 granted to S.N. We thank the members of the Research Resource Center at the RIKEN Brain Science Institute for the generation of the PAS KO mice. Work in the T.H.J. laboratory was supported by the ERC (grant 339953). We thank Manfred Schmid for computational support.

References

For readability of this thesis, the references in this chapter are numbered in sequence with the other references used throughout the manuscript, and are integrated in the References section at the end of this book.

Part IV: General Discussion

Chapter I: On NEAT1 and paraspeckles in cancer

In the next paragraphs, I will discuss some of the work described in this thesis in the broader context of the scientific literature. I will also highlight some aspects of *NEAT1* biology which, partly due to space limitations, may not have received due attention in the individual sections of the results.

Altering NEAT1 levels in cancer

NEAT1 levels are significantly altered in several cancers compared to their normal counterpart tissues¹⁹⁴; however, the available data is two-sided. In liquid tumors such as acute lymphocytic leukemia, *NEAT1* levels tend to go down; an effect that may be mediated by increased expression or activity of the transcriptional regulator c-Myc²⁶², who binds directly to the *NEAT1* promoter in several cell lines and promoter locations (Retrieved from UCSC Genome Browser, Hg19. Track: ENCODE with Factorbook Motifs (Txn Factor ChIP-seq)). Data not included in this manuscript, generated by Dr. Laura Standaert and myself confirm a transcriptional downregulation of *NEAT1* upon c-Myc induction in the hyperplastic lymph nodes and B-cell lymphomas in the Eμ-Myc mouse model, suggesting that in tumors of blood and lymphoid origin, the *NEAT1* locus may indeed be predominantly regulated by transcriptional repressor activities of Myc (Myc levels consistently show an inverse correlation with *Neat1* levels). In fact, our data suggests that the loss of *Neat1* in a crossing of the Eμ-Myc model with *Neat1* KO, HET or WT mice does not change the outcome of the tumor progression or the mechanism of transformation; i.e. p53 is still inactivated through various means, and Kaplan-Meier curves generated from the *Neat1* KO, HET and WT animals show that survival outcomes are similar (data not shown).

On the other hand, *NEAT1* is upregulated in many solid tumors and even proves to be of prognostic value for tumor aggression and patient survival¹⁹⁴. *NEAT1* expression is regulated by transcription factors with established roles in cancer formation such as p53^{139,141,210}, ATF2¹⁷⁹, HIF2a¹⁸², ER²²², c-Myc²⁶², or HSP1²⁵². Although some studies have suggested that increased *NEAT1* levels in solid tumors may protect the cells from a toxic accumulation of DNA breaks during the cell cycle^{141,198,263}, no true mechanism of action for *NEAT1* has so far convincingly explained these observations, and a cellular rationale for *NEAT1* up- or downregulation in tumor cells is still lacking (see below).

NEAT1 downstream of p53: not a tumor suppressor?

Although we were the first to show that the lncRNA *NEAT1* is downstream of p53 *in vivo*, previous data had suggested this in an *in vitro* setting in a chronic lymphocytic leukemia cell line²¹⁰. Later on, these findings were confirmed both *in vitro* and *in vivo* in pancreatic cancer¹³⁹,

endothelial cells²⁶⁴ and lung cancer cells²⁵⁴. Despite that fact that p53 has a clear tumor suppressor function, not all of its targets do, in part because of the many feedback loops that affect p53 biology and the requirement for dual outcomes upon its activation, i.e. cell survival and repair of the DNA or cell death and apoptosis¹¹³. Thus, as knockdown of *NEAT1* led to a decrease in cell viability and growth, we hypothesized that *NEAT1* functioned in the cell as an oncogene rather than a tumor suppressor.

In this context, two open questions remain: does *NEAT1* effectively contribute to a negative feedback loop with p53? And, is *NEAT1* really an oncogene, or merely a modulator of cellular responses in times of stress? In the lab, we have generated some preliminary data to answer the first question (e.g. see the doctoral thesis of Dr. Laura Standaert²⁰⁰). Here, RNA-sequencing after *NEAT1* KD in non-treated and Nutlin-3a conditions suggested that depletion of *NEAT1* led to a moderate upregulation of other p53 targets, implying that *NEAT1* suppresses the canonical p53 response. However, it remains unknown whether this observation can be explained by direct interactions of *NEAT1* with p53. Some follow-up experiments in fact suggested that the *NEAT1*-knockdown induced p53 upregulation was very context dependent (e.g. depending on the cell type), indicating that here, just like for many long non-coding RNAs, physiological status and conditions as well as the cell type may play an important role in defining their functional specificity¹¹³.

Second, although *NEAT1* has characteristics of an oncogene (e.g. its loss causes a decrease in viability and cell growth, it is upregulated in tumors and confers poor survival, etc.), an oncogene is usually only defined as a gene of which the product, when overexpressed in the cell, is *sufficient* to induce tumor formation. However, partially because of technical difficulties, we have never formally tested this. Indeed, as *NEAT1* and especially the long isoform *NEAT1_2* is very large (>22.7 kb), overexpression of this RNA remains a challenge. With the advent of novel technologies such as overexpression from the endogenous locus with dCas9 coupled to a transcriptional activator (CRISPRa²⁶⁵), this hypothesis might be more easy to tackle experimentally in the future (see e.g. Yamazaki et al., 2018b²⁶⁶).

Finally, the completely different regulation of the *NEAT1* locus in different settings both from data in resting cells and during the cell cycle prompts caution when interpreting *NEAT1* as downstream target of p53. Indeed, as it seems that *NEAT1_1* and *NEAT1_2* might have very differing functions and localizations in the nucleus, it may be important to keep track of which of the two isoforms is affected in future experiments, and how they are regulated on the 3' end should not be overlooked¹⁰.

Replication stress: cause of *NEAT1* KD induced DNA damage

The (spontaneous) accumulation of DNA damage upon knockdown of *NEAT1/NEAT1_2* in the absence of external stress can be explained by an accumulation of replication stress. Broadly, replication stress is defined as the external and internal events that cause DNA damage, usually triggering a response during replication via the kinases ATR and ATM²⁶⁷. Although not tested formally, we can propose several hypotheses on where the replication stress observed in this study originates from. For instance, reactive oxygen species (ROS) are produced by the cellular metabolism and present in culture media through supra-physiological oxygen levels in the environment²¹⁵ and cause inter-and intra-strand crosslinks in the DNA²⁶⁸. Exposed to ROS, the DNA of the cells is under constant attack, which may stall the replication machinery through an ATR-Chk1 mediated mechanism²¹⁵.

In addition, replication stress can occur when the to-be-incorporated nucleotides in the S phase are insufficiently present²¹⁵. This can be due to a decreased metabolism of nucleic acids in the cells, or, indirectly, because of alterations in the function of the proteins involved in dNTP production. One experiment to test this hypothesis could consist of supplementing nucleosides to the cells in *NEAT1* knockdown conditions, and to verify if with this, the DNA damage accumulation and cell cycle arrest are rescued²¹⁵. How this would function mechanistically in *NEAT1* KD, though, remains to be determined. For instance, we could imagine that *NEAT1_2* serves in the cell as a “hub” in which activation of a nucleotide-producing enzyme takes place (e.g. by facilitating spatial proximity of the relevant players). However, this particular model for *NEAT1* KD-induced replication stress remains highly speculative, as no evidence exists to date that implicates the replication machinery or enzymes important for nucleoside production to bind *NEAT1* directly. To support this, it would be interesting to test if *NEAT1* affects the rates and sequence of replication events in the cell; for instance by replication combing assays in control and *NEAT1* KD conditions²⁶⁹.

Instead, a more likely explanation for the deficiency in nucleotides could be found in the narrow sense of the concept of oncogene-induced replication stress: when the cells receive aberrant signals to start a new cell cycle before they have recovered from the previous one, there will be insufficient building blocks to satisfactorily duplicate the DNA in the S-phase, leading to defective replication (stalling of the forks)²⁶⁷. When stalled replication forks are not restarted before the cell decides to undergo mitosis, they will collapse and cause damage²¹⁵. Similarly, when the stimuli for growth push the cell to proliferate exceedingly fast, the mistakes or aberrations inflicted on the DNA upon replication may remain unrepaired²¹⁵.

Finally, a last potential cause for the accumulation of DNA damage during replication are replication-transcription conflicts. When not resolved, these can lead to aberrant R-loop

formation and activation of the DNA damage machinery in the cell²⁷⁰. R-loops are three stranded hybrid structures in which the RNA product of a locus re-hybridizes with the DNA it is transcribed from, leading to a RNA-DNA hybrid and a loose DNA strand²⁷⁰. In fact, some preliminary data not included in this manuscript suggested that when a nuclear form of the R-loop resolving protein RNaseH1 is overexpressed in cells in which *NEAT1_2* is knocked down, the cells accumulate less damage, but not upon overexpression of a catalytically inactive mutant or a control construct (data not shown). This suggests that, either indirectly or directly, *NEAT1_2* knockdown causes accumulation of R-loops; which may contribute to the DNA damage observed upon its loss in proliferating cells²⁷⁰.

When looking at the small scale, e.g. solely to the *NEAT1* locus, the latter hypothesis does not seem unlikely: DNA damage could occur at the *NEAT1* locus because of R-loop formation (see below). Indeed, due to high levels of transcription, the *NEAT1* locus produces abundant R-loops. It is, for instance, used as the html front page example by the R-loop documenting database R-loopDB²⁷¹. Whether this suggests that *NEAT1_2* production is a protection mechanism against *NEAT1_1*-mediated disruption of normal processes in the nucleus, remains unlikely as *NEAT1_1* abundance upon RRP40 knockdown does not impede proper cell cycle progression, and thus suggests that the cell is not sensitive to changes in *NEAT1_1* levels for its proliferation.

On the other hand, although not directly linked to the work presented in this manuscript, linking *NEAT1* and cancer with R-loop biology allows me to make a side note here on *NEAT1* and its mutation status in cancer. Indeed, *NEAT1* has been shown to be highly mutated, but no functional implications of these mutations are found to date. Therefore, I will discuss in the paragraphs below a possible mechanism and explanation for these observations.

NEAT1 mutations: (not) just bystanders

Tumorigenesis and cancer progression are driven by altering the genetic make-up of the cell through changes in gene expression patterns and their mutational load. Recently, it has become increasingly clear that observations pertaining to protein coding genes in cancer can be extended to the non-coding genome²⁷². For instance, a non-coding mutation was shown to create a super-enhancer influencing the downstream expression of an oncogene in T-ALL²⁷³, and frequent mutations of CTCF/cohesin binding sites in the genome may lead to changes in gene expression, higher genomic instability and cancer^{274–276}. For lncRNAs specifically, it has been hypothesized that mutations in genes may predict sites of functionality in the transcript, e.g. sites of interactions with proteins and DNA or nucleotides crucial for RNA secondary structure which may affect their function.

In addition to changes in expression levels, an alternative, but not mutually exclusive mechanism for *NEAT1* in cancer is suggested by the observation that *NEAT1* and its promoter are frequently mutated in a plethora of cancers²⁷⁷ including breast^{278,279}, liver²⁸⁰ and papillary kidney²⁸¹. *NEAT1* does however not show mutational ‘hotspots’²⁷⁷; the mutations rather occur all the way throughout the gene body both in the shape of Single Nucleotide Variants and as larger, two to five base long insertions and deletions²⁸².

A recently emerging hypothesis to explain this is based on the high *NEAT1* expression levels in these proliferating cells: the active *NEAT1* locus may have a higher likelihood of incurring transcription-replication machinery collisions²⁸². This hypothesis is independently supported, as discussed above, by the observation that the *NEAT1* locus extensively accumulates DNA-RNA hybrids (R-loops)^{271,283}, a nucleic acid structure often occurring at highly active loci, and known to cause DNA breaks and thus mutations²⁷⁰. Furthermore, other highly expressed genes such as ALB, SFTPB, and the lncRNA MALAT1 exhibit similar mutational patterns²⁸². Indeed, like *NEAT1*, the genomic locus of the intron-less MALAT1 exhibits high levels of R-loops, and thus a similar mechanism for these observations may be at play^{271,284}. Finally, it is clearly the highly transcribed *NEAT1* gene body itself that is affected, as promoter mutation in RNA *NEAT1* is not common in solid cancers²⁸⁵. Therefore, although it is unknown whether *NEAT1* mutations have functional implications for paraspeckle formation or how they affect the cancer cells through other means, the individual significance of each mutated position may actually be limited. In conclusion, these observations suggest that individual mutations of *NEAT1* might not influence tumorigenesis and progression, but rather that it is the isoform-specific expression level that is of importance.

Chapter II: On NEAT1 and paraspeckle function

Link with DNA repair

In this work, we find that modulation of *NEAT1_2* levels causes loss of genomic integrity. We discuss some potential causes for these observation stress above. However, how this could function mechanistically remains to be fully understood. One potential explanation could be found in the observation that *NEAT1* and paraspeckle proteins interact with several DNA damage repair factors from the Non-Homologous End Joining (NHEJ) pathway in the context of another nuclear particle, the HDP-RNP¹⁹¹. Here, *NEAT1* and the canonical paraspeckle proteins PSPC1, NONO, SFPQ, RBM14 and MATR3 are retained in a complex containing HEXIM1. The latter protein is mostly known for its negative regulatory function tethering away the lncRNA 7SK from the positive transcription elongation factor p-TEFb through sequestration of CDK9 and Cyclin T1 in the 7SK mRNP²⁸⁶. However, here, the HDP-RNP particle seems to be more important for the cellular stress response than for transcriptional regulation. Indeed, the complex was shown to undergo remodeling, releasing the paraspeckle and DNA repair components in order to provide a rapid response upon viral infection. Another study also found the NHEJ components bound to *NEAT1*, which, although not followed up experimentally, could be the same complex¹⁶⁸.

Because in the above studies *NEAT1* and the paraspeckle proteins are used interchangeably to assign paraspeckle/HDP-RNP functions, it is unclear whether the interactions occur only at paraspeckles or whether they represent completely independent *NEAT1* vs. paraspeckle protein functions. For instance, only small fractions of the transcription component HEXIM1 is involved in the HDP-RNP¹⁹¹ and thus, depending on the cellular conditions, HEXIM1 may have different interaction partners, pointing to a dynamic role of paraspeckle components, *NEAT1*, and the transcription regulation factors. What drives the genesis of these “PS-like bodies”, and what determines which interactions prevail over others at a given moment in a given cell remains to be determined.

In addition, several other reports have suggested roles for individual paraspeckle proteins in the DNA repair process (reviewed in ¹⁴⁹). For instance, the paraspeckle protein NONO has previously been shown to interact with the Ku70 and Ku80, conferring sequence specificity at the DNA²⁸⁷. SFPQ/PSF has been shown to facilitate recruitment of the other paraspeckle proteins and ATM target Matrin3 to DNA breaks²³⁵. Similarly, NONO has been proposed to play a role in DNA repair through its PARylation-dependent recruitment to laser-induced DNA lesions²³³. Finally, also PSPC1 is involved in DNA repair: Li et al. showed that the deletion of NONO is compensated by increased PSPC1 expression in MEFs, and that only their double deletion confers hypersensitivity to radiation¹⁵¹. In a supplementary experiment of the first part

of this study, however, we show using controlled low-power laser experiments that the GFP-tagged paraspeckle proteins from these papers are not relocating to the double strand breaks, but may rather be depleted from them. On the other hand, high power induced ssDNA lesions can recruit the PSPs in a potentially non-specific manner (data not shown). The latter suggests that, although the paraspeckle components may contribute to DNA repair and safeguard DNA integrity (their knockdown also causes replication stress); these functions, consistent with a function through the paraspeckle nuclear bodies, are most likely indirect.

Potential functions for *NEAT1_1*

In this study, we could not pinpoint a role for the short *NEAT1_1* isoform in the phenotypes found for *Neat1* full KO or *NEAT1* both isoform modulation, nor could we corroborate that *NEAT1_1*, even in conditions in which it is highly expressed, was implicated in the regulation of gene expression¹⁶⁸. However, previous work has proposed functions for *NEAT1* which, with our set of experiments, we could have missed. In the end, *NEAT1* is a relatively highly conserved RNA among vertebrates (especially the first half of the short isoform and some regulatory elements), and to content ourselves with the idea that its high production levels in the cell and its relatively high conservation have no meaning at all, is difficult. Thus, here, I will try to lay out a potential functional framework in which *NEAT1_1* could play a role, independently of *NEAT1_2* and paraspeckles, mostly based on a series of experiments by different groups that have implicated *NEAT1* in the response to viral infection.

Before *NEAT1* was associated experimentally with paraspeckles, it had been previously identified as the “virus-inducible” non-coding RNA VINC²⁶¹. If we think of *NEAT1* as a virus-inducible RNA, we can envision that its cell cycle/DNA integrity function occurred later on in evolution, potentially by establishing the paraspeckle nuclear body via the transcriptional read-through described above. This may give us a glimpse of, hypothetically, how or why *NEAT1* has evolved, and where it may play a role independently of growth functions.

Several reports provide ideas on roles for *NEAT1* in which 1) *NEAT1* was not assessed in an isoform-specific manner, and any observation could pertain to either the short or the long isoform or both, and 2) no proliferative effects are expected or tested. First of all, *NEAT1* is clearly upregulated during viral infection, including Human Immunodeficiency Virus 1 (HIV)¹⁸⁴, Herpes Simplex Virus 1^{191,260}, Hanta virus²⁵⁹, and Hepatitis Delta Virus (HDV)²⁸⁸. In addition, *NEAT1* can also be upregulated using non-viral immune stimulatory agents such as interferon-stimulatory DNA¹⁹¹, synthetic double stranded RNA²⁸⁹, or poly-I:C¹⁸⁹, suggesting a common mechanism is at play. Second, *NEAT1* knockdown results in severe effects on viral infections. For example, *NEAT1* enhances beta interferon (IFN-β) production and suppressed Hantaviral infection²⁵⁹ and facilitates activation of DNAPKcs and IRF3 as part of the innate immune

response¹⁹¹. In addition, *NEAT1* KD increases HIV viral transcripts¹⁸⁴. In HDV infection, parapeckles disassemble, despite *NEAT1* upregulation, suggesting that *NEAT1* function in viral infection is independent of the nuclear body²⁸⁸. Together, these results suggest a possible role for *NEAT1*, potentially independently of paraspeckle nuclear bodies, as protectors against viral infections. In the future, it may be important to further dissect the upstream signaling pathways leading to *NEAT1* upregulation and its mechanisms of action independently of paraspeckle nuclear bodies.

Chapter III: On NEAT1 and paraspeckle structure

Paraspeckle formation requires the right environment

Our work suggests that *NEAT1* can exist in multiple conformations in the cell, including as small, dispersed *NEAT1_1* “microspeckles”¹⁶⁷ and as larger paraspeckles built on *NEAT1_2*. While the dependence of paraspeckle formation on *NEAT1_2* is well-established, a recent study showed that not the entire *NEAT1_2* RNA is required for paraspeckle biogenesis but that the middle region of the transcript, rich in LINE and SINE repeats, is able to drive their phase separation¹⁶¹. A major paraspeckle protein necessary for its integrity and implicated in this process is FUSED In Sarcoma (FUS)^{14,163,290}. In the absence of the *NEAT1_2* middle region, the transcripts lose their typical open loop-like-structure within the paraspeckles and become, just like in FUS-KO/KD cells, small and dispersed throughout the nucleus^{163,290}.

Here, we noticed that *NEAT1_2* containing paraspeckles in hydroxyurea conditions also significantly changed their shape: instead of large accumulations of many *NEAT1_2* transcripts, they became reminiscent of paraspeckles in the FUS KO cells¹⁶³ or when the middle region of the long transcript is removed¹⁶¹. This indicates that both the RNA (the repeat-rich middle regions), the essential proteins (e.g. FUS) and the right cellular environment (e.g. the right local levels of nucleic acids) are necessary for their formation. Indeed, as the local levels and the nature of nucleic acids in the cell are crucial to define different liquid phases in the intracellular environment²⁹¹, we could imagine that the dispersion of *NEAT1_2* upon HU treatment is due to a disruption of the correct parameters needed for paraspeckle formation. However, experimental verification of this hypothesis is warranted, e.g. are paraspeckles in HU conditions sensitive to the solvent 1,6-hexanediol? And, what about *NEAT1_1* containing microspeckles? Are these phase-separated, too?

Determining specificity for *NEAT1* nucleation

In conjunction with the contribution of *NEAT1* to the phase separation process, it needs to be noted that the majority of the essential paraspeckle proteins undergo phase separation through their protein-protein interactions^{161,290}. Thus, when considering the mechanisms of paraspeckle formation, several observations may also point to a scaffolding role for RNA irrespective of sequence. For instance, phosphorothioate oligonucleotides can form paraspeckle-like foci with the paraspeckle proteins in the absence of *NEAT1* in cultured cells¹⁸⁶, and in marsupials, an orthologous *NEAT1* with very little sequence conservation nucleates paraspeckle-like foci *in cellulo*¹⁸⁸. These observations prompt us to ask: if paraspeckles could form irrespective of *NEAT1_2* sequence, what drives the nuclear body formation with *NEAT1* and the PS proteins

specifically in an *in vivo* setting? Eg. Why don't the PS proteins form extensive, phase separated nuclear bodies with other (repeated) RNAs?

The answer to these questions may lie in that their formation may be largely dependent on other factors such as RNA secondary structure and modification, and/or posttranslational modification (PTM) of *NEAT1*-binding proteins. In support of this idea, PTMs in FUS, including phosphorylation and methylation, modulate its liquid-demixing properties²⁹². Another prominent PS protein, SPFQ/PSF, is also methylated and citrullinated and these modifications enhance and reduce, respectively, its RNA binding capacities and thus potentially influence its role in liquid demixing²⁹³. Furthermore, PTMs may have inhibitory effects on paraspeckles; notably, CARM1-mediated arginine methylation of the essential paraspeckle protein NONO/p54nrb counteracts paraspeckle function and *NEAT1* transcription¹⁸³. To what extent these PTMs have an impact on phase separation remains to be experimentally determined²⁹⁴.

In conclusion, although *NEAT1* has been proposed to be the central nucleating factor for the paraspeckle, the above observations may in fact point to a largely protein-driven mechanism for their formation. They support a conceptual framework that begins to explain how in certain parts of the nucleus *NEAT1_2* RNA and proteins can aggregate, whereas in others *NEAT1_1* and, speculatively, maybe even *NEAT1_2* occurs 'free' from the paraspeckle nuclear body. It further provides an explanation for how the vast majority of paraspeckle proteins remain dispersed in the nucleoplasmic space¹⁶⁷.

Finally, this may eventually warrant the redefinition of the "paraspeckle nuclear body" as not just a "*NEAT1_2*-containing structure", but rather a complex, phase separated nuclear organelle in which all the essential proteinaceous components and the *NEAT1_2* RNA are structurally working together, and for which the molecular grammar still needs to be fully understood. Indeed, in order to definitively find a function for paraspeckles and *NEAT1*, it will be necessary, in future studies, to very precisely define the roles and components of the system under study. Therefore, we conclude here that, even though bits and pieces of the *NEAT1* story are uncovered, much more biology is to be learned to finally draw the larger picture on this elusive nuclear compartment and RNA.

Chapter IV: Global conclusion and outlook

With the work presented in this thesis, I hope to have provided an insight into *NEAT1* and paraspeckle biology. We can summarize a few key points:

- *NEAT1* is a long, non-coding RNA
- *NEAT1* is a p53 target, transcribed into two isoforms, *NEAT1_1* and *NEAT1_2*
- *NEAT1_2* forms paraspeckle nuclear bodies with a number of paraspeckle proteins
- The contribution of *NEAT1_1* to paraspeckles remains unclear
- *NEAT1_2* knockdown/loss of paraspeckles causes replication stress
- *NEAT1_2* knockdown/loss of paraspeckles sensitizes cancer cells to chemotherapy
- *NEAT1_2* could be an interesting cancer therapeutic target
- *NEAT1* and paraspeckles are not interchangeable, and need to be studied as such
- *NEAT1_1* is predominant in G1/G0 cells, whereas the long isoform prevails in SG2
- *NEAT1_1* can be knocked out by deleting its polyadenylation signal
- *Neat1_1* does not contribute to *Neat1* full KO phenotypes in mice
- *NEAT1_1* is specifically degraded by the RNA exosome
- *NEAT1_1* function remains unknown, but could be linked to the innate immune response.

In the future, it may be important to further dissect the regulation of the different isoforms, and how they each contribute to the physiological changes observed upon stressors. It is clear that *NEAT1* and paraspeckles are regulated on multiple levels, in particular the promoter level, the short and long isoform 3' ends, and through the cofactors required for nuclear body formation. This work suggests that both isoforms have significantly different functions in the cell, and is, I believe, only the tip of the iceberg on what can be known and understood.

However, I hope that this work has provided a framework to start studying paraspeckles and *NEAT1*, and potentially the vast amount of other (non-coding) RNAs and proteins in isoform and component-specific manners, but in particular without losing sight of the greater picture. We could imagine that, once *NEAT1* and paraspeckle biology become well understood, a potential therapeutic exploitation of this data in cancer may help to fight this horrendous disease.

Part V: References

1. Francastel, C., Schubeler, D., Martin, D. I. & Groudine, M. Nuclear compartmentalization and gene activity. *Nat. Rev. Mol. Cell Biol.* **1**, 137–143 (2000).
2. Zhu, L. & Brangwynne, C. P. Nuclear bodies: The emerging biophysics of nucleoplasmic phases. *Curr. Opin. Cell Biol.* **34**, 23–30 (2015).
3. Courchaine, E. M., Lu, A. & Neugebauer, K. M. Droplet organelles? *EMBO J.* **35**, 1603–12 (2016).
4. Spector, D. L. Nuclear domains. *J. Cell Sci.* **114**, 2891–3 (2001).
5. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13–23 (2017).
6. Hancock, R. Internal organisation of the nucleus: Assembly of compartments by macromolecular crowding and the nuclear matrix model. *Biol. Cell* **96**, 595–601 (2004).
7. Banani, S. F. *et al.* Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016).
8. Feric, M. *et al.* Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**, 1686–1697 (2016).
9. Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
10. Fox, A. H., Nakagawa, S., Hirose, T. & Bond, C. S. Paraspeckles: Where long noncoding RNA meets phase separation. *Trends Biochem. Sci.* **43**, 124–135 (2018).
11. Hutchinson, J. N. *et al.* A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 39 (2007).
12. Clemson, C. M. *et al.* An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell* **33**, 717–26 (2009).
13. Souquere, S., Beauclair, G., Harper, F. & Fox, A. Highly Ordered Spatial Organization of the Structural Long Noncoding NEAT1 RNAs within Paraspeckle Nuclear Bodies. *Mol. Biol. Cell* **21**, 4020–4027 (2010).
14. Naganuma, T. *et al.* Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J.* **31**, 4020–34 (2012).
15. Mao, Y. S., Sunwoo, H., Zhang, B. & Spector, D. L. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13**, 95–101 (2011).
16. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
17. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
18. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
19. Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**, 199–208 (2015).
20. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–89 (2012).
21. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).
22. Cabili, M. N. *et al.* Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16**, 20 (2015).
23. Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R. & Johnson, R. Towards a

- complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.* **19**, 535–548 (2018).
- 24. Ulitsky, I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* **17**, 601 (2016).
 - 25. Mattick, J. S. Non-coding RNAs: The architects of eukaryotic complexity. *EMBO Rep.* **2**, 986–991 (2001).
 - 26. Taft, R. J., Pheasant, M. & Mattick, J. S. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**, 288–299 (2007).
 - 27. Zhao, X.-Y. & Lin, J. D. Long Noncoding RNAs: A New Regulatory Code in Metabolic Control. *Trends Biochem. Sci.* **40**, 586–596 (2015).
 - 28. Schmitt, A. M. & Chang, H. Y. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* **29**, 452–463 (2016).
 - 29. Quinn, J. J. & Chang, H. Y. Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* **17**, 47–62 (2016).
 - 30. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
 - 31. Melé, M. & Rinn, J. L. “Cat’s Cradling” the 3D Genome by the Act of LncRNA Transcription. *Mol. Cell* **62**, 657–664 (2016).
 - 32. Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
 - 33. Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* **3**, ra8 (2010).
 - 34. Yao, H. *et al.* Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* **24**, 2543–2555 (2010).
 - 35. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013).
 - 36. Lewis, C. J. T., Pan, T. & Kalsotra, A. RNA modifications and structures cooperate to guide RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.* **18**, 202–210 (2017).
 - 37. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169**, 1187–1200 (2017).
 - 38. Terry, L. J., Shows, E. B. & Wente, S. R. Crossing the nuclear envelope: hierarchical regulation of nucleocytoplasmic transport. *Science* **318**, 1412–1416 (2007).
 - 39. Palazzo, A. F. & Lee, E. S. Sequence Determinants for Nuclear Retention and Cytoplasmic Export of mRNAs and lncRNAs. *Front. Genet.* **9**, 440 (2018).
 - 40. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–63 (2005).
 - 41. Kornienko, A. E., Guenzl, P. M., Barlow, D. P. & Paudler, F. M. Gene regulation by the act of long non-coding RNA transcription. *BMC Biol.* **11**, 59 (2013).
 - 42. Bohmdorfer, G. & Wierzbicki, A. T. Control of Chromatin Structure by Long Noncoding RNA. *Trends Cell Biol.* **25**, 623–632 (2015).
 - 43. Kurokawa, R., Rosenfeld, M. G. & Glass, C. K. Transcriptional regulation through noncoding RNAs and epigenetic modifications. *RNA Biol.* **6**, 233–236 (2009).
 - 44. Li, W., Notani, D. & Rosenfeld, M. G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.* **17**, 207–223 (2016).

45. Vihervaara, A., Duarte, F. M. & Lis, J. T. Molecular mechanisms driving transcriptional stress responses. *Nat. Rev. Genet.* **19**, 385–397 (2018).
46. Doma, M. K. & Parker, R. RNA quality control in eukaryotes. *Cell* **131**, 660–668 (2007).
47. Houseley, J., LaCava, J. & Tollervey, D. RNA-quality control by the exosome. *Nat. Rev. Mol. Cell Biol.* **7**, 529–539 (2006).
48. Skourtis-Stathaki, K. & Proudfoot, N. J. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.* **28**, 1384–1396 (2014).
49. Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* **8**, 113–126 (2007).
50. Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763–776 (2009).
51. Castel, S. E. & Martienssen, R. A. RNA interference in the nucleus: Roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.* **14**, 100–12 (2013).
52. Sarshad, A. A. *et al.* Argonaute-miRNA Complexes Silence Target mRNAs in the Nucleus of Mammalian Stem Cells. *Mol. Cell* **71**, 1040–1050.e8 (2018).
53. Spiller, M. P., Reijns, M. A. M. & Beggs, J. D. Requirements for nuclear localization of the Lsm2-8p complex and competition between nuclear and cytoplasmic Lsm complexes. *J. Cell Sci.* **120**, 4310–4320 (2007).
54. Kufel, J., Bousquet-Antonacci, C., Beggs, J. D. & Tollervey, D. Nuclear pre-mRNA decapping and 5' degradation in yeast require the Lsm2-8p complex. *Mol. Cell. Biol.* **24**, 9646–9657 (2004).
55. Xu, Y.-F. *et al.* Nucleotide degradation and ribose salvage in yeast. *Mol. Syst. Biol.* **9**, 665 (2013).
56. Zinder, J. C. & Lima, C. D. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev.* **31**, 88–100 (2017).
57. Kilchert, C., Wittmann, S. & Vasiljeva, L. The regulation and functions of the nuclear RNA exosome complex. *Nat. Rev. Mol. Cell Biol.* **17**, 227–239 (2016).
58. Gudipati, R. K. *et al.* Extensive degradation of RNA precursors by the exosome in wild-type cells. *Mol. Cell* **48**, 409–421 (2012).
59. Mure, F. *et al.* The splicing factor SRSF3 is functionally connected to the nuclear RNA exosome for intronless mRNA decay. *Sci. Rep.* **8**, 12901 (2018).
60. Molleston, J. M. *et al.* A conserved virus-induced cytoplasmic TRAMP-like complex recruits the exosome to target viral RNA for degradation. *Genes Dev.* **30**, 1658–1670 (2016).
61. Lubas, M. *et al.* The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep.* **10**, 178–192 (2015).
62. Ogami, K., Chen, Y. & Manley, J. L. RNA surveillance by the nuclear RNA exosome: mechanisms and significance. *Non-coding RNA* **4**, (2018).
63. Gerlach, P. *et al.* Distinct and evolutionary conserved structural features of the human nuclear exosome complex. *Elife* **7**, (2018).
64. Tomecki, R. *et al.* The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *EMBO J.* **29**, 2342–2357 (2010).
65. Wang, Z. & Kiledjian, M. Functional link between the mammalian exosome and mRNA decapping. *Cell* **107**, 751–762 (2001).
66. Milanowska K, Mikolajczak K, Lukasik A, Skorupski M, B. Z. & Machnicka MA, Nowacka M, Rother KM, B. J. RNApPathwaysDB - a database of RNA maturation and decay. *RNApPathwaysDB – a database RNA Matur. decay pathways* **41**, D268-72 (2013).
67. Lubas, M. *et al.* Interaction profiling identifies the human nuclear exosome targeting complex. *Mol. Cell* **43**, 624–637 (2011).

68. Silla, T., Karadoulama, E., Mąkosa, D., Lubas, M. & Jensen, T. H. The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export Activity to Retain Target Transcripts. *Cell Rep.* **23**, 2199–2210 (2018).
69. Meola, N. *et al.* Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. *Mol. Cell* **64**, 520–533 (2016).
70. Fasken, M. B. *et al.* Air1 zinc knuckles 4 and 5 and a conserved IWRXY motif are critical for the function and integrity of the Trf4/5-Air1/2-Mtr4 polyadenylation (TRAMP) RNA quality control complex. *J. Biol. Chem.* **286**, 37429–37445 (2011).
71. LaCava, J. *et al.* RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**, 713–724 (2005).
72. Vanacova, S. *et al.* A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol.* **3**, e189 (2005).
73. Wyers, F. *et al.* Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**, 725–737 (2005).
74. Preker, P. *et al.* RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851–1854 (2008).
75. Tomecki, R. *et al.* Multiple myeloma-associated hDIS3 mutations cause perturbations in cellular RNA metabolism and suggest hDIS3 PIN domain as a potential drug target. *Nucleic Acids Res.* **42**, 1270–1290 (2014).
76. Mistry, D. S., Chen, Y. & Sen, G. L. Progenitor function in self-renewing human epidermis is maintained by the exosome. *Cell Stem Cell* **11**, 127–135 (2012).
77. McIver, S. C. *et al.* Exosome complex orchestrates developmental signaling to balance proliferation and differentiation during erythropoiesis. *Elife* **5**, (2016).
78. McIver, S. C. *et al.* The exosome complex establishes a barricade to erythroid maturation. *Blood* **124**, 2285–2297 (2014).
79. Fazzio, T. G., Huff, J. T. & Panning, B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* **134**, 162–174 (2008).
80. Onderak, A. M. & Anderson, J. T. Loss of the RNA helicase SKIV2L2 impairs mitotic progression and replication-dependent histone mRNA turnover in murine cell lines. *RNA* **23**, 910–926 (2017).
81. Lloret-Llinares, M. *et al.* The RNA exosome contributes to gene expression regulation during stem cell differentiation. *Nucleic Acids Res.* **46**, 11502–11513 (2018).
82. Blasius, M., Wagner, S. A., Choudhary, C., Bartek, J. & Jackson, S. P. A quantitative 14-3-3 interaction screen connects the nuclear exosome targeting complex to the DNA damage response. *Genes Dev.* **28**, 1977–1982 (2014).
83. Richard, P., Feng, S. & Manley, J. L. A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. *Genes Dev.* **27**, 2227–2232 (2013).
84. Basu, U. *et al.* The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* **144**, 353–363 (2011).
85. Wan, J. *et al.* Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration. *Nat. Genet.* **44**, 704–708 (2012).
86. Graham, A. C., Kiss, D. L. & Andrulis, E. D. Core Exosome-independent Roles for Rrp6 in Cell Cycle Progression. *Mol. Biol. Cell* **20**, 2242–2253 (2009).
87. Fitzpatrick, G. V., Soloway, P. D. & Higgins, M. J. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat. Genet.* **32**, 426–431 (2002).
88. Grote, P. *et al.* The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall

- development in the mouse. *Dev. Cell* **24**, 206–214 (2013).
- 89. Lewandowski, J. P. *et al.* The Tug1 Locus is Essential for Male Fertility. *bioRxiv* 562066 (2019). doi:10.1101/562066
 - 90. Gomez, J. A. *et al.* The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. *Cell* **152**, 743–754 (2013).
 - 91. Sauvageau, M. *et al.* Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* **2**, e01749 (2013).
 - 92. Li, L. & Chang, H. Y. Physiological roles of long noncoding RNAs: insight from knockout mice. *Trends Cell Biol.* **24**, 594–602 (2014).
 - 93. Amaral, P. P., Dinger, M. E. & Mattick, J. S. Non-coding RNAs in homeostasis, disease and stress responses: an evolutionary perspective. *Brief. Funct. Genomics* **12**, 254–278 (2013).
 - 94. Valadkhan, S. & Valencia-Hipolito, A. lncRNAs in Stress Response. *Curr. Top. Microbiol. Immunol.* **394**, 203–236 (2016).
 - 95. Audas, T. E. & Lee, S. Stressing out over long noncoding RNA. *Biochim. Biophys. Acta* **1859**, 184–91 (2015).
 - 96. Gutschner, T. & Diederichs, S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol.* **9**, 703–719 (2012).
 - 97. Espinosa-Diez, C. *et al.* DNA damage dependent hypomethylation regulates the pro-angiogenic LncRNA MEG9. *bioRxiv* 442699 (2018). doi:10.1101/442699
 - 98. Kartha, R. V & Subramanian, S. Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. *Front. Genet.* **5**, 8 (2014).
 - 99. Lee, S. *et al.* Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* **164**, 69–80 (2016).
 - 100. Munschauer, M. *et al.* The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature* **561**, 132–136 (2018).
 - 101. Tichon, A., Perry, R. B. T., Stojic, L. & Ulitsky, I. SAM68 is required for regulation of pumilio by the NORAD long noncoding RNA. *Genes Dev.* **32**, 70–78 (2018).
 - 102. Michelini, F. *et al.* Damage-induced lncRNAs control the DNA damage response through interaction with dRNAs at individual double-strand breaks. *Nat. Cell Biol.* **19**, 1400–1411 (2017).
 - 103. Francia, S., Cabrini, M., Matti, V., Oldani, A. & d'Adda di Fagagna, F. DICER, DROSHA and DNA damage response RNAs are necessary for the secondary recruitment of DNA damage response factors. *J. Cell Sci.* **129**, 1468–1476 (2016).
 - 104. Huarte, M. The emerging role of lncRNAs in cancer. *Nat. Med.* **21**, 1253 (2015).
 - 105. Doose, G. *et al.* MINCR is a MYC-induced lncRNA able to modulate MYC's transcriptional network in Burkitt lymphoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5261-70 (2015).
 - 106. Lu, Y. *et al.* MYC Targeted Long Noncoding RNA DANCR Promotes Cancer in Part by Reducing p21 Levels. *Cancer Res.* **78**, 64–74 (2018).
 - 107. Hung, C.-L. *et al.* A long noncoding RNA connects c-Myc to tumor metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18697–18702 (2014).
 - 108. Grossi, E., Sanchez, Y. & Huarte, M. Expanding the p53 regulatory network: LncRNAs take up the challenge. *Biochim. Biophys. Acta* **1859**, 200–208 (2016).
 - 109. Wu, X., Bayle, J. H., Olson, D. & Levine, A. J. The p53-mdm-2 autoregulatory feedback loop. *Genes Dev.* **7**, 1126–1132 (1993).
 - 110. Levine, A. J. & Oren, M. The first 30 years of p53: growing ever more complex. *Nat. Rev. Cancer*

9, 749–58 (2009).

111. Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805–816 (1993).
112. el-Deiry, W. S. *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817–825 (1993).
113. Kastenhuber, E. R. & Lowe, S. W. Putting p53 in Context. *Cell* **170**, 1062–1078 (2017).
114. Villunger, A. *et al.* p53- and drug-induced apoptotic responses mediated by BH3-only proteins puma and noxa. *Science* **302**, 1036–8 (2003).
115. Miyashita, T. & Reed, J. C. Tumor suppressor p53 is a direct transcriptional activator of the human bax gene. *Cell* **80**, 293–299 (1995).
116. Khan, M. R. *et al.* TP53LNC-DB, the database of lncRNAs in the p53 signalling network. *Database (Oxford)*. **2019**, (2019).
117. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).
118. Li, X. L. *et al.* Long Noncoding RNA PURPL Suppresses Basal p53 Levels and Promotes Tumorigenicity in Colorectal Cancer. *Cell Rep.* **20**, 2408–2423 (2017).
119. Burgess, A. *et al.* Clinical Overview of MDM2/X-Targeted Therapies. *Front. Oncol.* **6**, 7 (2016).
120. Lane, D. P., Cheok, C. F. & Lain, S. p53-based cancer therapy. *Cold Spring Harb. Perspect. Biol.* **2**, a001222 (2010).
121. Khvorova, A. & Watts, J. K. The chemical evolution of oligonucleotide therapies of clinical utility. *Nat. Biotechnol.* **35**, 238–248 (2017).
122. Stein, C. A. & Castanotto, D. FDA-Approved Oligonucleotide Therapies in 2017. *Mol. Ther.* **25**, 1069–1075 (2017).
123. Nero, T. L., Morton, C. J., Holien, J. K., Wielens, J. & Parker, M. W. Oncogenic protein interfaces: small molecules, big challenges. *Nat. Rev. Cancer* **14**, 248–262 (2014).
124. Beck, A., Goetsch, L., Dumontet, C. & Corvaia, N. Strategies and challenges for the next generation of antibody-drug conjugates. *Nat. Rev. Drug Discov.* **16**, 315–337 (2017).
125. Arun, G., Diermeier, S. D. & Spector, D. L. Therapeutic Targeting of Long Non-Coding RNAs in Cancer. *Trends Mol. Med.* **24**, 257–277 (2018).
126. Leucci, E. *et al.* Melanoma addiction to the long non-coding RNA SAMMSON. *Nature* **531**, 518–522 (2016).
127. Mitobe, Y., Takayama, K.-I., Horie-Inoue, K. & Inoue, S. Prostate cancer-associated lncRNAs. *Cancer Lett.* **418**, 159–166 (2018).
128. Dong, R. *et al.* Targeting long non-coding RNA-TUG1 inhibits tumor growth and angiogenesis in hepatoblastoma. *Cell Death Dis.* **7**, e2278 (2016).
129. Scaiewicz, V. *et al.* Use of H19 Gene Regulatory Sequences in DNA-Based Therapy for Pancreatic Cancer. *J. Oncol.* **2010**, 178174 (2010).
130. Kaczmarek, J. C., Kowalski, P. S. & Anderson, D. G. Advances in the delivery of RNA therapeutics: from concept to clinical reality. *Genome Med.* **9**, 60 (2017).
131. Landskron, L. *et al.* The asymmetrically segregating lncRNA cherub is required for transforming stem cells into malignant cells. *Elife* **7**, (2018).
132. Hosono, Y. *et al.* Oncogenic Role of THOR, a Conserved Cancer/Testis Long Non-coding RNA. *Cell* **171**, 1559–1572.e20 (2017).
133. Renninger, S. L., Schonthaler, H. B., Neuhauss, S. C. F. & Dahm, R. Investigating the genetics

- of visual processing, function and behaviour in zebrafish. *Neurogenetics* **12**, 97–116 (2011).
- 134. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611 (2009).
 - 135. Lapouge, G. *et al.* Identifying the cellular origin of squamous skin tumors. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7431–6 (2011).
 - 136. Frese, K. K. & Tuveson, D. A. Maximizing mouse cancer models. *Nat. Rev. Cancer* **7**, 645–658 (2007).
 - 137. Harris, A. W. *et al.* The E mu-myc transgenic mouse. A model for high-incidence spontaneous lymphoma and leukemia of early B cells. *J. Exp. Med.* **167**, 353–371 (1988).
 - 138. Mou, H., Kennedy, Z., Anderson, D. G., Yin, H. & Xue, W. Precision cancer mouse models through genome editing with CRISPR-Cas9. *Genome Med.* **7**, 53 (2015).
 - 139. Mello, S. S. *et al.* Neat1 is a p53-inducible lincRNA essential for transformation suppression. *Genes Dev.* **31**, 1095–1108 (2017).
 - 140. Wu, J. *et al.* Identification of a long non-coding RNA NR_026689 associated with lung carcinogenesis induced by NNK. *Oncotarget* **7**, 14486–14498 (2016).
 - 141. Adriaens, C. *et al.* p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat Med* **22**, 861–868 (2016).
 - 142. Ponzio, G. *et al.* A new long noncoding RNA (lncRNA) is induced in cutaneous squamous cell carcinoma and down-regulates several anticancer and cell differentiation genes in mouse. *J. Biol. Chem.* **292**, 12483–12495 (2017).
 - 143. Arun, G. *et al.* Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev.* **30**, 34–51 (2016).
 - 144. Richards, E. J. *et al.* Long non-coding RNAs (lncRNA) regulated by transforming growth factor (TGF) beta: lncRNA-hit-mediated TGFbeta-induced epithelial to mesenchymal transition in mammary epithelia. *J. Biol. Chem.* **290**, 6857–6867 (2015).
 - 145. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
 - 146. Rupaimoole, R. *et al.* Long Noncoding RNA Ceruloplasmin Promotes Cancer Growth by Altering Glycolysis. *Cell Rep.* **13**, 2395–2402 (2015).
 - 147. Andersen, J. S. *et al.* Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12**, 1–11 (2002).
 - 148. Fox, A. H. *et al.* Paraspeckles: a novel nuclear domain. *Curr. Biol.* **12**, 13–25 (2002).
 - 149. Knott, G. J., Bond, C. S. & Fox, A. H. The DBHS proteins SFPQ, NONO and PSPC1: A multipurpose molecular scaffold. *Nucleic Acids Res.* **44**, 3989–4004 (2016).
 - 150. Fox, A. H., Bond, C. S. & Lamond, A. I. P54nrb Forms a Heterodimer with PSP1 That Localizes to Paraspeckles in an RNA-dependent Manner. *Mol. Biol. Cell* **16**, 5304–5315 (2005).
 - 151. Li, S. *et al.* Double-strand break repair deficiency in NONO knockout murine embryonic fibroblasts and compensation by spontaneous upregulation of the PSPC1 paralog. *Nucleic Acids Res.* **42**, 9771–9780 (2014).
 - 152. Passon, D. M., Lee, M., Fox, A. H. & Bond, C. S. Crystallization of a paraspeckle protein PSPC1-NONO heterodimer. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67**, 1231–1234 (2011).
 - 153. Huang, J. *et al.* Crystal structure of a SFPQ/PSPC1 heterodimer provides insights into preferential heterodimerization of human DBHS family proteins. *J. Biol. Chem.* **293**, 6593–6602 (2018).
 - 154. Bond, C. S. & Fox, A. H. Paraspeckles: nuclear bodies built on long noncoding RNA. *J. Cell Biol.*

186, 637–44 (2009).

155. Sasaki, Y. T. F., Ideue, T., Sano, M., Mituyama, T. & Hirose, T. MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2525–30 (2009).
156. Sunwoo, H. *et al.* MEN e / b nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* **19**, 347–359 (2009).
157. Li, W. *et al.* Distinct regulation of alternative polyadenylation and gene expression by nuclear poly(A) polymerases. *Nucleic Acids Res.* **45**, 8930–8942 (2017).
158. Yamazaki, T. & Hirose, T. The building process of the functional paraspeckle with long non-coding RNAs. *Front. Biosci. (Elite Ed.)* **7**, 1–41 (2015).
159. Wilusz, J. E. *et al.* A triple helix stabilizes the 3??? ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev.* **26**, 2392–2407 (2012).
160. Brown, J. A., Valenstein, M. L., Yario, T. A., Tycowski, K. T. & Steitz, J. A. Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MEN noncoding RNAs. *Proc. Natl. Acad. Sci.* **109**, 19202–19207 (2012).
161. Yamazaki, T. *et al.* Functional domains of NEAT1 architectural lncRNA induce paraspeckle assembly through phase separation. *Mol. Cell* **70**, 1038–1053.e7 (2018).
162. Clark, M. B. *et al.* Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**, 885–98 (2012).
163. West, J. A. *et al.* Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J. Cell Biol.* **214**, 817–30 (2016).
164. Mao, Y. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies. *Trends Genet.* **27**, 295–306 (2011).
165. Nakagawa, S., Naganuma, T., Shioi, G. & Hirose, T. Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *J. Cell Biol.* **193**, 31–39 (2011).
166. Nakagawa, S. & Hirose, T. Paraspeckle nuclear bodies--useful uselessness? *Cell. Mol. Life Sci.* **69**, 3027–36 (2012).
167. Li, R., Harvey, A. R., Hodgetts, S. I. & Fox, A. H. Functional dissection of NEAT1 using genome editing reveals substantial localization of the NEAT1_1 isoform outside paraspeckles. *RNA* **23**, 872–881 (2017).
168. West, J. A. A. *et al.* The Long Noncoding RNAs NEAT1 and MALAT1 Bind Active Chromatin Sites. *Mol. Cell* **55**, 791–802 (2014).
169. Standaert, L. *et al.* The long noncoding RNA Neat1 is required for mammary gland development and lactation. *RNA* **20**, 1844–9 (2014).
170. Nakagawa, S. *et al.* The lncRNA Neat1 is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. *Development* **141**, 4618–27 (2014).
171. Chujo, T. *et al.* Unusual semi-extractability as a hallmark of nuclear body-associated architectural noncoding RNAs. *EMBO J.* **36**, 1447–1462 (2017).
172. Aguzzi, A. & Altmeyer, M. Phase separation: Linking cellular compartmentalization to disease. *Trends Cell Biol.* **26**, 547–558 (2016).
173. Nott, T. J., Craggs, T. D. & Baldwin, A. J. Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters. *Nat. Chem.* **8**, 569–575 (2016).
174. Kawaguchi, T. *et al.* SWI/SNF chromatin-remodeling complexes function in noncoding RNA-dependent assembly of nuclear bodies. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 4304–4309 (2015).
175. Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**, 1267–1279 (2016).

176. Lin, Y., Schmidt, B. F., Bruchez, M. P. & McManus, C. J. Structural analyses of NEAT1 lncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture. *Nucleic Acids Res.* **46**, 3742–3752 (2018).
177. Chen, L.-L. & Carmichael, G. G. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol. Cell* **35**, 467–78 (2009).
178. Prasanth, K. V *et al.* Regulating gene expression through RNA nuclear retention. *Cell* **123**, 249–63 (2005).
179. Wang, Y. *et al.* Genome-wide screening of NEAT1 regulators reveals cross-regulation between paraspeckles and mitochondria. *Nat. Cell Biol.* **20**, 1145–1158 (2018).
180. Torres, M. *et al.* Circadian RNA expression elicited by 3'-UTR IRAU-paraspeckle associated elements. *Elife* **5**, e14837 (2016).
181. Wang, J. *et al.* RNA-binding protein PSPC1 promotes the differentiation-dependent nuclear export of adipocyte RNAs. *J. Clin. Invest.* **127**, 987–1004 (2017).
182. Choudhry, H. *et al.* Tumor hypoxia induces nuclear paraspeckle formation through HIF-2[alpha] dependent transcriptional activation of NEAT1 leading to cancer cell survival. *Oncogene* **34**, 4482–90 (2014).
183. Hu, S. Bin *et al.* Protein arginine methyltransferase CARM1 attenuates the paraspeckle mediated nuclear retention of mRNAs containing IRAUs. *Genes Dev.* **29**, 630–645 (2015).
184. Zhang, Q. *et al.* NEAT1 Long Noncoding RNA and Paraspeckle Bodies Modulate HIV-1 Posttranscriptional Expression. *Am. Soc. Microbiol.* **4**, e00596-12 (2013).
185. Anantharaman, A. *et al.* Paraspeckles modulate the intranuclear distribution of paraspeckle-associated Ctn RNA. *Sci. Rep.* **6**, 34043 (2016).
186. Shen, W., Liang, X. H. & Crooke, S. T. Phosphorothioate oligonucleotides can displace NEAT1 RNA and form nuclear paraspeckle-like structures. *Nucleic Acids Res.* **42**, 8648–62 (2014).
187. Česnik, A. B. *et al.* Nuclear RNA foci from C9ORF72 expansion mutation form paraspeckle-like bodies. *J. Cell Sci. jcs.224303* (2019). doi:10.1242/jcs.224303
188. Cornelis, G., Souquere, S., Vernochet, C., Heidmann, T. & Pierron, G. Functional conservation of the lncRNA NEAT1 in the ancestrally diverged marsupial lineage: Evidence for NEAT1 expression and associated paraspeckle assembly during late gestation in the opossum *Monodelphis domestica*. *RNA Biol.* **13**, 826–36 (2016).
189. Imamura, K. *et al.* Long Noncoding RNA NEAT1-Dependent SFPQ Relocation from Promoter Region to Paraspeckle Mediates IL8 Expression upon Immune Stimuli. *Mol. Cell* **53**, 393–406 (2014).
190. Hirose, T. *et al.* NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol. Biol. Cell* **25**, 169–83 (2014).
191. Morchikh, M. *et al.* HEXIM1 and NEAT1 Long Non-coding RNA Form a Multi-subunit Complex that Regulates DNA-Mediated Innate Immune Response. *Mol. Cell* **67**, 387–399 (2017).
192. Lahaye, X. *et al.* NONO Detects the Nuclear HIV Capsid to Promote cGAS-Mediated Innate Immune Activation. *Cell* **175**, 488–501 (2018).
193. Jiang, L. *et al.* NEAT1 scaffolds RNA-binding proteins and the Microprocessor to globally enhance pri-miRNA processing. *Nat. Struct. Mol. Biol.* **24**, 816–824 (2017).
194. Li, S., Li, J., Chen, C., Zhang, R. & Wang, K. Pan-cancer analysis of long non-coding RNA NEAT1 in various cancers. *Genes Dis.* **5**, 27–35 (2018).
195. Shuaib, M. *et al.* AGO1 in association with NEAT1 lncRNA contributes to nuclear and 3D chromatin architecture in human cells. *bioRxiv* 525527 (2019). doi:10.1101/525527

196. Ahmed, A. S. I. *et al.* Long noncoding RNA NEAT1 (nuclear paraspeckle assembly transcript 1) is critical for phenotypic switching of vascular smooth muscle cells. *Proc. Natl. Acad. Sci.* **115**, E8660–E8667 (2018).
197. Chen, Q. *et al.* Long noncoding RNA NEAT1, regulated by the EGFR pathway, contributes to glioblastoma progression through the WNT/b-catenin pathway by scaffolding EZH2. *Clin. Cancer Res.* **24**, 684–695 (2018).
198. Li, X. *et al.* Oncogenic properties of NEAT1 in prostate cancer cells depend on the CDC5L–AGRN transcriptional regulation circuit. *Cancer Res.* **78**, 4138–4149 (2018).
199. Major, A. T. *et al.* Specific interaction with the nuclear transporter importin 2 can modulate paraspeckle protein 1 delivery to nuclear paraspeckles. *Mol. Biol. Cell* **26**, 1543–58 (2015).
200. Standaert, L. The functional relevance of lncRNA downstream of p53 - A NEAT1 story. (KU Leuven, 2015).
201. Vogelstein, B., Lane, D. & Levine, a J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
202. Sharpless, N. E. & DePinho, R. a. How stem cells age and why this makes us grow old. *Nat. Rev. Mol. Cell Biol.* **8**, 703–13 (2007).
203. Vassilev, L. T. *et al.* In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848 (2004).
204. Khoo, K. H., Hoe, K. K., Verma, C. S. & Lane, D. P. Drugging the p53 pathway: understanding the route to clinical efficacy. *Nat. Rev. Drug Discov.* **13**, 217–36 (2014).
205. Carvajal, L. a & Manfredi, J. J. Another fork in the road--life or death decisions by the tumour suppressor p53. *EMBO Rep.* **14**, 414–21 (2013).
206. Huang, B., Deo, D., Xia, M. & Vassilev, L. T. Pharmacologic p53 activation blocks cell cycle progression but fails to induce senescence in epithelial cancer cells. *Mol. Cancer Res.* **7**, 1497–509 (2009).
207. Yildirim, E. *et al.* Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell* **152**, 727–42 (2013).
208. Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* **43**, 621–9 (2011).
209. Marín-Béjar, O. *et al.* Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. *Genome Biol.* **14**, R104 (2013).
210. Blume, C. J. *et al.* P53-dependent non-coding RNA networks in chronic lymphocytic leukemia. *Leukemia* **29**, 2015–2023 (2015).
211. Janky, R. *et al.* iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections. *PLoS Comput Biol* **10**, e1003731 (2014).
212. Herrmann, C., Van De Sande, B., Potier, D. & Aerts, S. i-cisTarget: An integrative genomics method for the prediction of regulatory features and cis-regulatory modules. *Nucleic Acids Res.* **40**, (2012).
213. Reinhardt, H. C. & Schumacher, B. The p53 network: cellular and systemic DNA damage responses in aging and cancer. *Trends Genet.* **28**, 128–36 (2012).
214. Nassar, D., Latil, M., Boeckx, B., Lambrechts, D. & Blanpain, C. Genomic landscape of carcinogen-induced and genetically induced mouse skin squamous cell carcinoma. *Nat. Med.* **21**, 946–54 (2015).
215. Gaillard, H., García-Muse, T. & Aguilera, A. Replication stress and cancer. *Nat. Rev. Cancer* **15**, 276–289 (2015).
216. Llopis, A. *et al.* The stress-activated protein kinases p38??? and JNK1/2 cooperate with Chk1 to inhibit mitotic entry upon DNA replication arrest. *Cell Cycle* **11**, 3627–3637 (2012).

217. Branzei, D. & Foiani, M. Maintaining genome stability at the replication fork. *Nat. Rev. Mol. Cell Biol.* **11**, 208–19 (2010).
218. Dobbelstein, M. & Sørensen, C. S. Exploiting replicative stress to treat cancer. *Nat. Rev. Drug Discov.* **14**, 405–423 (2015).
219. Sullivan, K. D. et al. ATM and MET kinases are synthetic lethal with nongenotoxic activation of p53. *Nature Chemical Biology* **8**, 646–654 (2012).
220. Ferriss, J. S. et al. Multi-gene expression predictors of single drug responses to adjuvant chemotherapy in ovarian carcinoma: Predicting platinum resistance. *PLoS One* **7**, (2012).
221. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
222. Chakravarty, D. et al. The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat. Commun.* **5**, 5383 (2014).
223. Zhen, L., Yun-hui, L., Hong-yu, D., Jun, M. & Yi-long, Y. Long noncoding RNA NEAT1 promotes glioma pathogenesis by regulating miR-449b-5p/c-Met axis. *Tumor Biol.* **37**, 673–83 (2015).
224. Guo, S. et al. Clinical implication of long non-coding RNA NEAT1 expression in hepatocellular carcinoma patients. *Int. J. Clin. Exp. Pathol.* **8**, 5395–402 (2015).
225. Pan, L.-J. et al. Upregulation and clinicopathological significance of long non-coding NEAT1 RNA in NSCLC tissues. *Asian Pac. J. Cancer Prev.* **16**, 2851–5 (2015).
226. Sengupta, S. & Harris, C. C. p53: traffic cop at the crossroads of DNA repair and recombination. *Nat. Rev. Mol. Cell Biol.* **6**, 44–55 (2005).
227. Harris, S. L. & Levine, A. J. The p53 pathway: positive and negative feedback loops. *Oncogene* **24**, 2899–2908 (2005).
228. Cha, H. et al. Wip1 directly dephosphorylates gamma-H2AX and attenuates the DNA damage response. *Cancer Res.* **70**, 4112–22 (2010).
229. Moon, S.-H. et al. Wild-type p53-induced phosphatase 1 dephosphorylates histone variant gamma-H2AX and suppresses DNA double strand break repair. *J. Biol. Chem.* **285**, 12935–47 (2010).
230. Chowdhury, D. et al. gamma-H2AX dephosphorylation by protein phosphatase 2A facilitates DNA double-strand break repair. *Mol. Cell* **20**, 801–809 (2005).
231. Altmeyer, M. et al. Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat. Commun.* **6**, 8088 (2015).
232. Rulten, S. L. et al. PARP-1 dependent recruitment of the amyotrophic lateral sclerosis-associated protein FUS/TLS to sites of oxidative DNA damage. *Nucleic Acids Res.* **42**, 307–314 (2014).
233. Krietsch, J. et al. PARP activation regulates the RNA-binding protein NONO in the DNA damage response to DNA double-strand breaks. *Nucleic Acids Res.* **40**, 10287–301 (2012).
234. Ha, K., Takeda, Y. & Dynan, W. S. Sequences in PSF/SFPQ mediate radioresistance and recruitment of PSF/SFPQ-containing complexes to DNA damage sites in human cells. *DNA Repair (Amst.)* **10**, 252–259 (2011).
235. Salton, M., Lerenthal, Y., Wang, S. Y., Chen, D. J. & Shiloh, Y. Involvement of Matrin 3 and SFPQ/NONO in the DNA damage response. *Cell Cycle* **9**, 1568–1576 (2010).
236. Mastrocola, A. S., Kim, S. H., Trinh, A. T., Rodenkirch, L. A. & Tibbetts, R. S. The RNA-binding protein fused in sarcoma (FUS) functions downstream of poly(ADP-ribose) polymerase (PARP) in response to DNA damage. *J. Biol. Chem.* **288**, 24731–24741 (2013).
237. Wickramasinghe, V. O. & Venkitaraman, A. R. RNA Processing and Genome Stability: Cause and Consequence. *Mol. Cell* **61**, 496–505 (2016).
238. Britton, S. et al. DNA damage triggers SAF-A and RNA biogenesis factors exclusion from chromatin coupled to R-loops removal. *Nucleic Acids Res.* **42**, 9047–9062 (2014).

239. Shanbhag, N. M., Rafalska-Metcalf, I. U., Balane-Bolivar, C., Janicki, S. M. & Greenberg, R. A. ATM-Dependent chromatin changes silence transcription in cis to dna double-strand breaks. *Cell* **141**, 970–981 (2010).
240. Specks J, Lecona E, Lopez-Contreras AJ, F.-C. O. A Single Conserved Residue Mediates Binding of the Ribonucleotide Reductase Catalytic Subunit RRM1 to RRM2 and Is Essential for Mouse Development. *Mol. Cell. Biol.* **35**, 2910–7 (2015).
241. Nakagawa, S., Yamazaki, T. & Hirose, T. Molecular dissection of nuclear paraspeckles: towards understanding the emerging world of the RNP milieu. *Open Biol.* **8**, (2018).
242. Jacks, T. et al. Tumor spectrum analysis in p53-mutant mice. *Curr. Biol.* **4**, 1–7 (1994).
243. Vasioukhin, V., Degenstein, L., Wise, B. & Fuchs, E. The magical touch: genome targeting in epidermal stem cells induced by tamoxifen application to mouse skin. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8551–8556 (1999).
244. Abel, E. L., Angel, J. M., Kiguchi, K. & DiGiovanni, J. Multi-stage chemical carcinogenesis in mouse skin: fundamentals and applications. *Nat. Protoc.* **4**, 1350–1362 (2009).
245. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
246. Anders, S., Pyl, P. T. & Huber, W. HTSeq--A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–9 (2015).
247. Harrow, J. et al. GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
248. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
249. Zhang, Y. et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
250. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011).
251. Modic, M. et al. Cross-regulation between TDP-43 and paraspeckles promotes pluripotency-differentiation transition. *Cell Press Sneak Peek* (2018). doi:10.2139/ssrn.3155931
252. Lellahi, S. M. et al. The long non-coding RNA NEAT1 and nuclear paraspeckles are upregulated by the transcription factor HSF1 in the heat shock response. *J. Biol. Chem.* **293**, 18965–18976 (2018).
253. Hupalowska, A. et al. CARM1 and Paraspeckles Regulate Pre-implantation Mouse Embryo Development. *Cell* **175**, 1902–1916.e13 (2018).
254. Idogawa, M., Ohashi, T., Sasaki, Y., Nakase, H. & Tokino, T. Long non-coding RNA NEAT1 is a transcriptional target of p53 and modulates p53-induced transactivation and tumor-suppressor function. *Int. J. Cancer* **140**, 2785–2791 (2017).
255. Shen, H., Moran, D. M. & Maki, C. G. Transient Nutlin-3a treatment promotes endoreduplication and the generation of therapy-resistant tetraploid cells. *Cancer Res.* **68**, 8260–8268 (2008).
256. Singh, A. & Xu, Y. J. The cell killing mechanisms of hydroxyurea. *Genes (Basel)* **7**, 99 (2016).
257. Vanacova, S. & Stef, R. The exosome and RNA quality control in the nucleus. *EMBO Rep.* **8**, 651–657 (2007).
258. Yu, X., Li, Z., Zheng, H., Chan, M. T. V & Wu, W. K. K. NEAT1: A novel cancer-related long non-coding RNA. *Cell Prolif.* **50**, (2017).
259. Ma, H. et al. The Long Noncoding RNA NEAT1 Exerts Antihantaviral Effects by Acting as Positive Feedback for RIG-I Signaling. *J. Virol.* **91**, e02250-16 (2017).
260. Wang, Z. et al. NEAT1 modulates herpes simplex virus-1 replication by regulating viral gene transcription. *Cell. Mol. Life Sci.* **74**, 1117–1131 (2017).

261. Saha, S., Murthy, S. & Rangarajan, P. N. Identification and characterization of a virus-inducible non-coding RNA in mouse brain. *J. Gen. Virol.* **87**, 1991–5 (2006).
262. Zeng, C. *et al.* The c-Myc-regulated lncRNA NEAT1 and paraspeckles modulate imatinib-induced apoptosis in CML cells. *Mol. Cancer* **17**, 130 (2018).
263. Adriaens, C. & Marine, J. C. J.-C. NEAT1-containing paraspeckles: Central hubs in stress response and tumor formation. *Cell Cycle* **16**, (2017).
264. Fuschi, P. *et al.* Central role of the p53 pathway in the noncoding-RNA response to oxidative stress. *Aging (Albany. NY)*. **9**, 2559–2586 (2017).
265. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180–2196 (2013).
266. Yamazaki, T., Fujikawa, C., Kubota, A., Takahashi, A. & Hirose, T. CRISPRa-mediated NEAT1 lncRNA upregulation induces formation of intact paraspeckles. *Biochem. Biophys. Res. Commun.* **504**, 218–224 (2018).
267. Mazouzi, A., Velimezi, G. & Loizou, J. I. DNA replication stress: causes, resolution and disease. *Exp. Cell Res.* **329**, 85–93 (2014).
268. Jena, N. R. DNA damage by reactive species: Mechanisms, mutation and repair. *J. Biosci.* **37**, 503–517 (2012).
269. Tourrière, H., Saksouk, J., Lengronne, A. & Pasero, P. Single-molecule Analysis of DNA Replication Dynamics in Budding Yeast and Human Cells by DNA Combing. *Bio-protocol* **7**, e2305 (2017).
270. Aguilera, A. & García-Muse, T. R Loops: From Transcription Byproducts to Threats to Genome Stability. *Mol. Cell* **46**, 115–24 (2012).
271. Jenjaroenpun, P., Wongsurawat, T., Sutheeworapong, S. & Kuznetsov, V. A. R-loopDB: A database for R-loop forming sequences (RLFS) and R-loops. *Nucleic Acids Res.* **45**, D119–D127 (2017).
272. Piraino, S. W. & Furney, S. J. Beyond the exome: The role of non-coding somatic mutations in cancer. *Ann. Oncol.* **27**, 240–8 (2016).
273. Mansour, M. R. *et al.* An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science (80-.).* **346**, 1373–1377 (2014).
274. Katainen, R. *et al.* CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.* **47**, 818–821 (2015).
275. Canela, A. *et al.* Genome organization drives chromosome fragility. *Cell* **170**, 507–521.e18 (2017).
276. Amaral, P. P. *et al.* Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. *Genome Biol.* **19**, 32 (2018).
277. Lanzos, A. *et al.* Discovery of Cancer Driver Long Noncoding RNAs across 1112 Tumour Genomes: New Candidates and Distinguishing Features. *Sci. Rep.* **7**, 41544 (2017).
278. Rheinbay, E. *et al.* Recurrent and functional regulatory mutations in breast cancer. *Nature* **547**, 55–60 (2017).
279. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 1–20 (2016).
280. Fujimoto, A. *et al.* Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* **48**, 500–9 (2016).
281. Li, S., Shuch, B. M. & Gerstein, M. B. Whole-genome analysis of papillary kidney cancer finds significant noncoding alterations. *PLoS Genet.* **13**, e1006685 (2017).
282. Rheinbay, E. *et al.* Discovery and characterization of coding and non-coding driver mutations in

more than 2,500 whole cancer genomes. *bioRxiv* 237313 (2017). doi:10.1101/237313

283. Dumelie, J. G. & Jaffrey, S. R. Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisD RIP-seq. *Elife* **6**, e28306 (2017).
284. Bonnet, A. et al. Introns Protect Eukaryotic Genomes from Transcription-Associated Genetic Instability. *Mol. Cell* **67**, 608–621 (2017).
285. Kim, M. S., Jeon, E. H., Mo, H. Y., Yoo, N. J. & Lee, S. H. Promoter mutation in long non-coding RNA NEAT1 is not common in common solid cancers. *Pathology, research and practice* **214**, 1912–1913 (2018).
286. Li, Q. et al. HEXIM1 is a promiscuous double-stranded RNA-binding protein and interacts with RNAs in addition to 7SK in cultured cells. *Nucleic Acids Res.* **35**, 2503–12 (2007).
287. Yang, Y. S., Yang, M. C. W., Tucker, P. W. & Donald Capra, J. NonO enhances the association of many DNA-binding proteins to their targets. *Nucleic Acids Res.* **25**, 2284–92 (1997).
288. Beeharry, Y., Goodrum, G., Imperiale, C. J. & Pelchat, M. The Hepatitis Delta Virus accumulation requires paraspeckle components and affects NEAT1 level and PSP1 localization. *Sci. Rep.* **8**, 6031 (2018).
289. Shelkovnikova, T. A. et al. Protective paraspeckle hyper-assembly downstream of TDP-43 loss of function in amyotrophic lateral sclerosis. *Mol. Neurodegener.* **13**, 30 (2018).
290. Hennig, S. et al. Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol.* **210**, 529–539 (2015).
291. Maharana, S. et al. RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science (80-).* **360**, 918–921 (2018).
292. Rhoads, S. N., Monahan, Z. T., Yee, D. S. & Shewmaker, F. P. The Role of Post-Translational Modifications on Prion-Like Aggregation and Liquid-Phase Separation of FUS. *Int. J. Mol. Sci.* **19**, E886 (2018).
293. Snijders, A. P. et al. Arginine methylation and citrullination of splicing factor proline- and glutamine-rich (SFPQ/PSF) regulates its association with mRNA. *RNA* **21**, 347–59 (2015).
294. Bah, A. & Forman-Kay, J. D. Modulation of intrinsically disordered protein function by post-translational modifications. *J. Biol. Chem.* **291**, 6696–705 (2016).

Part VI: Addendum

Curriculum Vitae

CARMEN ADRIAENS

Bankstraat 86

3000 Leuven, BELGIUM

+1 (301) 385 9449

+32 (486) 25 63 58

"YOU CERTAINLY USUALLY FIND SOMETHING, IF YOU LOOK,
BUT IT IS NOT ALWAYS QUITE THE SOMETHING YOU WERE
AFTER."

– J.R.R. TOLKIEN

carmen.adriaens@kuleuven.vib.be

carmen.adriaens@gmail.com

Research interests

Long non-coding RNA, Mouse Models, Nuclear Bodies, Genetics, Chromatin Organization, Cancer

Publications

Adriaens C, Rambow F, Bervoets G, et al. The lncRNA NEAT1_1 is dispensable for normal tissue homeostasis and cancer cell growth. (2019) **bioRxiv**, <https://doi.org/10.1101/574582>

Adriaens C*, Serebryannyy LA*, Feric M* et al., Blank spots on the map: Some Current Questions on Nuclear Organization and Genome Architecture. (2018) *Equal contribution by all authors*, **Histochem. Cell Biol.** 150(6):579-592

Rambow F*, Rogiers A*, Marin-Bejar O, Aibar S, Femel J, Dewaele M, Karras P, Brown D, Hwan Chang Y, Debiec-Rychter M, Adriaens C et al., Towards minimal residual disease-directed therapy in melanoma. (2018) **Cell.** 174, (4), 843-855.E19

Adriaens C & Marine JC. NEAT1-containing Paraspeckles: central hubs in stress response and tumor formation. **Cell Cycle.** 16, 137-138 (2017).

Adriaens C*, Standaert L*, Barra J, Latil M, Verfaillie A, et al. P53 induces formation of NEAT1 lncRNA containing paraspeckles that modulate replication stress response and chemosensitivity. **Nat. Med.** 22, 861-868 (2016).

Standaert L, Adriaens C, Radaelli E, Van Keymeulen A, Blanpain C, Hirose T, Nakagawa S, Marine JC. The long noncoding RNA *Neat1* is required for mammary gland development and lactation. **RNA.** 20, 1844–9 (2014).

Preprint Highlights for preLights at The Company of Biologists, UK - since launch of the website Feb.2018.
<https://prelights.biologists.com/profiles/carmena/>

Grants and Awards

- 2018 FWO Travel Grant for attendance of the ASCB EMBO 2018 meeting in San Diego, December 2018
- 2018 EMBO Travel Grant for attendance of the ASCB EMBO 2018 meeting in San Diego, December 2018
- 2018 RNA Society Travel Grant for attendance of the RNA Society Meeting June 2018.
- 2018 DMM Travel Grant, The company of Biologists.
- 2017 Gustave Boël – Sofina Travel Fellowship, 12 months, Research Stay at the NCI/NIH in the USA.
- 2015 Best research presentation, Oncoforum, Leuven, Belgium.
- 2014 IWT Graduate Student Fellowship, Applied Biomedical Sciences. 4 years. Belgian Research Foundation – Flanders.

Education

2014 - 2019	Ph.D. in Molecular and Cellular Biology, Laboratory for Molecular Cancer Biology (Prof. Chris Marine, IWT SB fellowship), VIB – KU Leuven, Belgium . Cancer Doctoral School.
2012 - 2014	M.Sc. in Biology, major in Molecular and Physiological Biology (Research track), KU Leuven, Belgium . Diploma, <i>Magna cum laude</i> .
2012 - 2013	First year of M.Sc. in Biology, Specialty Cellular biology, Physiology and Pathology / Neurobiology, Université de Bordeaux II, France (ERASMUS).
2009 - 2012	B.Sc. in Biology, KU Leuven, Belgium . Diploma, <i>Magna cum laude</i> .

Research experience

Oct. 2017 - Oct. 2018	Research stay in the Cell Biology of Genomes Group, lab of dr. Tom Misteli, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD, USA
Jan. - Feb. 2016	Research stay in the Steve Jackson Laboratory, Gurdon Institute, Cambridge, UK
Sept. - June 2014	<i>"The role of the long non-coding RNA NEAT1 in p53 Biology and beyond."</i> Master thesis, Laboratory for Molecular Cancer Biology, VIB - KU Leuven (17.6/20)
April - June 2013	<i>"Detection of a Side Population in Chronic Myelogenous Leukemia peripheral blood."</i> Internship, iNSERM 1035, Laboratory of biotherapies for genetic diseases and Cancer, Bordeaux, France (19/20)
Feb. - May 2012	<i>"Effects of allatotropin on protease secretion."</i> Department of Animal Physiology and Neurobiology, KU Leuven (18/20)
Oct. - Dec. 2011	<i>"Optimizing an RNAi screen on eukaryotic cell lines."</i> Department of Animal Physiology and Neurobiology, KU Leuven (18/20)

Posters and Talks

ASCB/EMBO meeting, San Diego, CA, USA. **Poster**, December 2018. Title: *Dissecting the (patho-) physiological functions of NEAT1 isoforms*. Travel supported by EMBO & FWO travel grants.

RNA Society Meeting, UC Berkeley, Berkeley, CA, USA. **Poster**, June 2018. Title: *Dissecting the (patho-) physiological functions of NEAT1 isoforms*. Travel supported by RNA Society and DMM travel grants.

6th course on Non-Coding Genome, Institut Curie, Paris, France. **Poster**, February 2017. Title: *NEAT1 isoform dynamics*.

The Dark Side of the Genome, fTALES conference, Ghent, Belgium. **Talk**, September 2016. Title: *NEAT1/paraspeckles: central hubs in cellular stress response and cancer*.

Humanitas Ph.D. Course on long non-coding RNAs, Milan, Italy. **Talk**, June 2016. Title: *NEAT1/paraspeckles: central hubs in cellular stress response and tumorigenesis*.

Interuniversity Attraction Pole, Brussels, Belgium. **Talk**, June 2016. Title: *p53-induces paraspeckles formation to modulate replication stress response and chemosensitivity*.

Departmental Research seminar, Campus Gasthuisberg, Leuven, Belgium. **Talk**, May 2016. Title: *The long non-coding RNA NEAT1 modulates the replication stress response and sensitivity to chemotherapy*.

Oncoforum, Campus Gasthuisberg, Leuven, Belgium. **Poster**, May 2016. Title: *The long non-coding RNA NEAT1 controls DNA repair, tumorigenesis and response to cancer therapy.*

EMBO Workshop Systems biology of non-coding RNAs, Rehovot, Israel. **Poster**, February 2016. Title: *p53-induced NEAT1 paraspeckles: central hubs in tumorigenesis and modulators of response to cancer therapy.*

Departmental Evaluation Board Symposium, Campus Gasthuisberg, Leuven. **Poster**, November 2015. Title: *The long non-coding RNA NEAT1 controls DNA repair, tumorigenesis and response to cancer therapy.*

Theo Murphy Long non-coding RNA Royal Society Meeting, Milton Keynes, UK. **Poster**, September 2015. Title: *The lncRNA Neat1 enables skin cancer development.*

VIB11/NERF PhD week, Campus Gasthuisberg, Leuven, Belgium. **Talk**, September 2015 (Award for best presentation). Title: *Assigning key (patho)physiological functions to the lncRNA NEAT1.*

Oncoforum, Campus Gasthuisberg, Leuven, Belgium. **Talk**, June 2015. Title: Assigning key (patho)logical functions to the lincRNA *NEAT1*.

Interuniversity Attraction Pole, Brussels, Belgium. **Poster**, June 2014. Title: The role of the long non-coding RNA *NEAT1* in p53 biology and beyond.

Informatics, Skills & Languages

MS Office	qRT-PCR	Dutch (Mother tongue)
Fiji / ImageJ	Western Blot	French (Fluent)
Graphpad Prism	Immunostaining (IF & IHC)	English (Fluent)
qBase Plus	Mouse work, Cell culture	Spanish (Notions)
FlowJo	Flow Cytometry	Portuguese (Notions)
UCSC genome browser	Cloning, CRISPR-Cas9 genome editing	
Benchling	RNA-FISH	
CLC Main Workbench	Confocal & superresolution microscopy	

Additional Ph.D. level courses and activities

Leadership in Academia (EMBO Career Development course)	GraphPad Prism Statistical Analysis (VIB)
Statistical Thinking and Smart Experimental Design (VIB)	Basics Statistics Theory (VIB)
Introduction to Next Generation Sequencing Analysis (VIB)	qPCR analysis using qbase+ (VIB)
Basics of UCSC Genome Browser – hands on (VIB – L. Lipovich)	Scientific Integrity part I & II (KU Leuven)
Writing for Medical Journals (KU Leuven – Stuart Spencer)	The Non-Coding Genome (VIB)
Cancer & Signaling (KU Leuven – Cancer Doctoral School)	Essentials for Image Editing (VIB – Somersault)
Cancer Therapy (KU Leuven – Cancer Doctoral School)	Systems Biology of non-coding RNAs (EMBO)
Career development training (RNA Society meeting)	Basics in R (NCI/NIH Bioinformatics unit)
Essential Tools for R (Leuven Statistics Research Centre)	Light on the dark side of the genome (fTALES)
FELASA B & C Certificates	Introduction to ggplot2 & statistics in R (LSTAT)

preLights

1. The chromatin remodeling factors dMLL3/4 regulate a small gene repertoire important for embryonic fate during fly oocyte development.

The Trithorax group protein dMLL3/4 instructs the assembly of the zygotic genome at fertilization

Pedro Prudêncio, Leonardo G. Guilgur, João Sobral, Jörg D. Becker, Rui Gonçalo Martinho, Paulo Navarro-Costa

Preprint posted on January 04, 2018 <https://www.biorxiv.org/content/early/2018/01/04/242008>

Selected by Carmen Adriaens
Categories: developmental biology

Global context and preprint summary

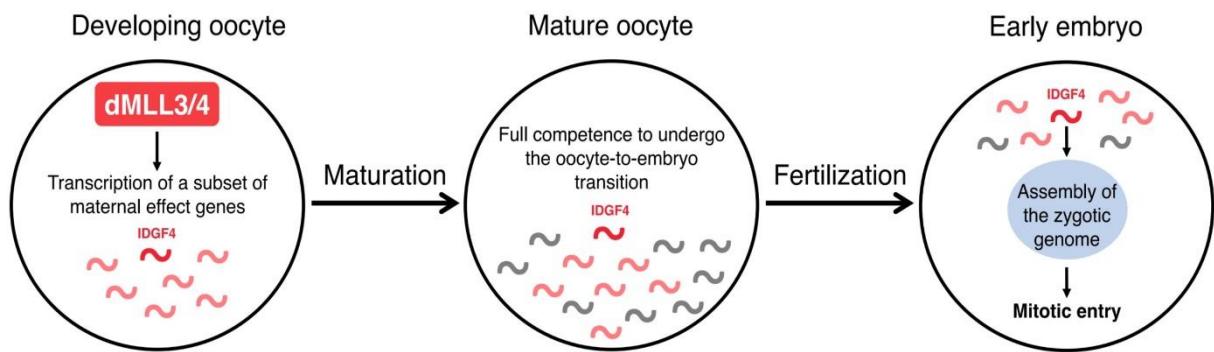
Oocyte-to-embryo (O-to-E) transition requires the assembly of two completely differently configured haploid genomes. For the pronuclei to merge into a single zygotic genome after fertilization, the parental chromatin needs to be extensively remodeled. A previous study has shown that the Trithorax-group chromatin remodeling proteins MLL3 and MLL4 are essential for the zygote to reprogram to pluripotency, but are dispensable for the maintenance of cell identity. In the current work, the authors find that the Drosophila dMLL3/4 proteins play a role in O-to-E transition through their chromatin remodeling and/or gene regulatory capacities. They identify a novel gene, IDGF4, which is required for O-to-E and is under the transcriptional control of dMLL3/4.

My favorite experiment in this study:

Although it is known that MLL3/4 was needed to establish the pluripotent cell state, the timing of this requirement has been unclear. To test this, the authors depleted these proteins specifically in the fly germline, and counted egg-hatching events after fertilization. Eggs from dMLL3/4 depleted females did not hatch due to a failure to enter embryogenesis at the first mitotic division, whereas depletion of these proteins in males did not affect the embryogenesis process. I think this experiment is great because it shows conclusively that maternally-provided dMLL3/4 is necessary before O-to-E can occur upon fertilization.

What I like about this work:

Despite our relatively broad knowledge on embryogenesis, still many aspects of this intriguing process remain elusive. This study from Prudêncio et al. strengthens the emerging concept that chromatin remodeling factors are important to reshape the genomic and transcriptomic landscapes in each of the embryogenesis stages independently. I like this study because it does not only pinpoint the exact timing of a known phenotype, but it also provides the first mechanistic insights into its establishment through the identification of a maternally encoded factor acting on the paternal genome.



The figure shows the model proposed by the authors of how the dMLL3/4-dependent gene expression signature might be necessary for the cell to undergo oocyte-to-embryo transition and accomplish the assembly of the zygotic genome. (From preprint, made available under a CC-BY-NC-ND 4.0 International license).

Tags: dmll3/4, drosophila, maternal effect genes, oocyte-to-embryo transition

Posted on: 16th February 2018

2. RNA in unexpected places: dilncRNAs at resected DSB ends facilitate homologous recombination-mediated repair of broken DNA

A role for RNA and DNA:RNA hybrids in the modulation of DNA repair by homologous recombination

Giuseppina D'Alessandro, Marek Adamowicz, Donna Whelan, Sean Michael Howard, Corey Winston Jones-Weinert, Valentina Matti, Eli Rothenberg, Petr Cejka, Fabrizio d'Adda di Fagagna

Preprint posted on January 29, 2018 <https://www.biorxiv.org/content/early/2018/01/29/255976>

Selected by Carmen Adriaens

Categories: cancer biology, cell biology, molecular biology

The idea:

Upon double stranded DNA break (DSB) formation, non-coding RNAs at the resected ends of the break enhance BRCA1 recruitment and facilitate homologous recombination (HR) in the S and G2 phases of the cell cycle.

What is this preprint about?

Because of its potential mutagenic effect, DNA damage is one of the most dangerous insults for the cell. Although the protein component of the DNA repair machinery has been relatively well documented, recent work has also identified (non-coding) RNA as a major contributor to fine-tune the DNA repair pathways. In 2017, the group of Dr. Fabrizio d'Adda di Fagagna showed that in mammalian cells, a novel class of long non-coding RNAs (dilncRNAs) is transcribed from broken ends upon DSB formation after resection, a process necessary to free up DNA ends before HR can occur. They evidenced that these dilncRNAs contribute to the DNA repair process both by acting as precursors for DNA damage-associated small non-coding RNAs (DDRNA, previously characterized in Francia et al., 2012) and through the recruitment of DDRNAs by RNA-RNA pairing (see Michelini et al., NCB, 2017). In their recent preprint, the same authors show that dilncRNAs enhance HR-mediated DNA repair in the HR-prone S and G2 phases of the cell cycle by recruiting major HR factors. Moreover, they provide evidence that RNase H2 is in a complex with the HR proteins BRCA1, PALB2, BRCA2 and RAD51, and that this R-loop-resolving protein is essential to remove the DNA-damage induced DNA-RNA hybrids so that proper repair can occur.

My opinion:

What I love about this preprint is that this study helps to put forward the notion that RNA molecules play crucial roles in processes commonly considered to be dominated solely by a protein system.

Specifically, in the nucleus, ephemeral RNA molecules act not only as messengers, but also as guides, supports, protectors, inhibitors or physical barriers, recruiters, or flag signs. The dilncRNAs help generic and abundant proteins, such as those of the DNA repair machinery,

to make the right choice in the right place and time. Because these RNAs are so versatile and short-lived, they provide the perfect “LINC” between a robust response to damage and an elegant fine-tuning of the complex nuclear environment.

In the future, it would be interesting to determine how the formation of RNA:DNA hybrids per se contributes to the DNA repair process, especially since they have been reported to be both the cause and consequence of DNA damage. Furthermore, it will be important to study to a further degree how the breaks and their repair are influenced by the RNA species (either bound or not to the broken and resected DNA ends), and how the proteins that regulate the biogenesis and metabolism of these RNA species affect the repair process.

Further reading:

1. Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. Ohle, Tesorero et al., 2016, Cell, Volume 167, Issue 4, 1001 – 1013.e7

This paper used a yeast system (*S. pombe*) to show how resected ends are transcribed bidirectionally and serve as substrates for DNA-RNA hybrid formation. It conceptually led the way for the current study.

2. Damage-induced lncRNAs control the DNA damage response through interaction with DDRNAs at individual double-strand breaks. Michelini et al., 2017, Nature Cell Biology 19, pages 1400–1411.

In this paper, the authors beautifully show a role for RNA polymerase II at DNA breaks and advance the idea that non-coding RNA may play a crucial role in DSB repair.

3. The emerging role of RNAs in DNA damage repair. Hawley et al., 2017. Cell death and differentiation. (2017) 24, 580–587

A great review on how in the last few years we came to realize how RNAs are more important than initially thought for the DNA repair processes in the cell.

4. Site-specific DICER and DROSHA RNA products control the DNA damage response. Francia et al. Nature (2012) 488(7410): 231–235.

The initial paper characterizing the involvement of DDRNA in the DNA repair process, through the DICER and DROSHA proteins, but not other components of the RNAi machinery.

Tags: dilncrna, dna repair, dna:rna hybrid, homologous recombination, Rnase H

Posted on: 19th March 2018, updated on: 20th March 2018

3. It's all about folds and RIPPLes: principles for higher order organization of pre-translational mRNPs

Higher-Order Organization Principles of Pre-translational mRNPs

Mihir Metkar, Hakan Ozadam, Bryan R. Lajoie, Maxim Imakaev, Leonid A. Mirny, Job Dekker, Melissa J. Moore

Preprint posted on March 08, 2018 <https://doi.org/10.1101/278747>

Selected by Carmen Adriaens

Categories: biochemistry, bioinformatics, cell biology, molecular biology

The idea:

A new technique named RIPPLiT combines RNA immunoprecipitation with proximity ligation to determine folding of RNAs within macromolecular messenger ribonucleoprotein (mRNP) particles.

What is this paper about?

In recent years, it has become increasingly clear that the 3D conformation of DNA and the correct folding of chromatin into contact frequency-determined topologically associating domains (TADs) is tremendously important for cellular identity and function. To study mRNA architecture and compaction in conjunction with their protein binding partners, in this work the authors apply the concept of contact frequency (here: chimeric junction frequency) and 3D conformation of macromolecules to develop a technique they name RNA ImmunoPrecipitation and Proximity Ligation in Tandem (RIPPLiT). With this method, they investigate the architecture of RNA within stable mRNPs, allowing to study the biogenesis, stability, folding and compaction of transcripts inside ubiquitous megadalton particles that protect and shuttle the RNA prior to translation.

The authors use this new technique to reveal the secondary structure of pre-translational RNA complexes through pulldown in tandem of two Exon Junction Complex (EJC) components. By means of proof-of-concept, they determine inter- and intramolecular junctions in ribosomal RNA. Because [as the authors put it] “the rules governing RNA polymerase II (Pol II) transcript packaging remain largely undefined”, they analyze the structure of Pol II transcripts in these complexes. Unlike previous cross-linking based studies, they find that Pol II transcripts, in their dataset, do not exhibit significant intermolecular contacts. In addition, they find that, independent of the length of the transcript, messenger RNA folding is relatively non-specifically distributed throughout the mRNP structural scaffold, resulting in densely packed, but flexible rods, rather than highly structured particles.

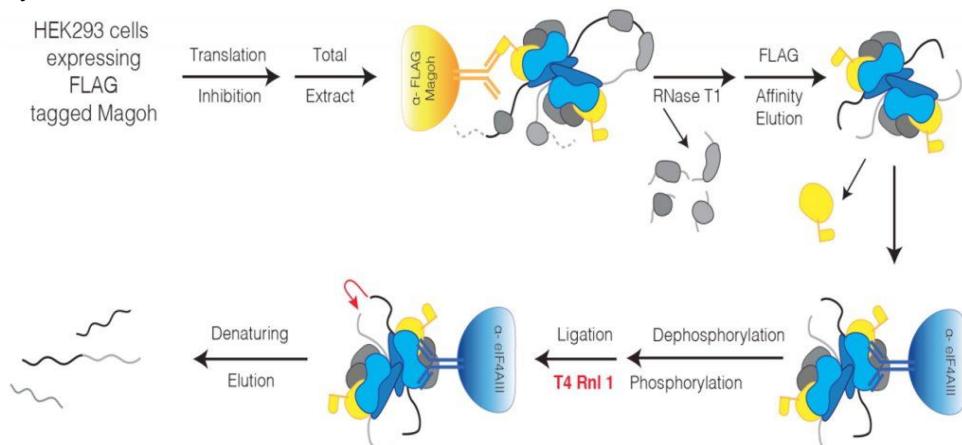
In short, they provide the first insights into how mRNAs are packaged in conjunction with their protein interactors, which are proposed to both protect them from degradation and shuttle them to the correct destination after their biogenesis.

My opinion on this preprint:

I like the fact that the authors apply a well-established concept – that of the importance of 3D architecture of macromolecules and complexes – to the study of (m)RNPs. They can now apply this technique to find the conformation of flexible RNA molecules within the protein particles that dictate their fate (e.g. translation, degradation, localization etc.), rather than to have to depend mainly on computational modeling. Furthermore, a major advantage is that unlike most other techniques for studying conformation of RNA, with RIPPLiT, they overcome the requirement for direct base-pairing of the RNA molecules to uncover spatial proximity.

It may be possible that the authors of this preprint detect very little intermolecular mRNA contacts because of the specific design of the technique (i.e. the use of Harringtonine to halt translation prior to protein pulldown), and the particular focus on the EJC and pre-translational complexes, rather than necessarily finding a global rule for mRNP biology. For instance, in another recent preprint (Morf et al., bioRxiv, 2017, <https://doi.org/10.1101/196147>), abundant spatial proximities are described for different RNAs in distinct nuclear particles, and it would be interesting to integrate data from these two techniques and others to ultimately obtain an integrative view on mRNP/RNA particle constitution in the cell.

Further implementation of RIPPLiT with other proteins in various cellular compartments will help to understand how RNAs interact within themselves and how they behave with their binding partners. Indeed, with RIPPLiT, it will be possible to investigate how these interactions and the global mRNP conformation affect the functions of protein complexes, and vice-versa. RIPPLiT may also provide an opportunity to further uncover fundamental differences and similarities between non-coding and protein-coding RNAs, and may be a useful platform to study, potentially using RNA-guided interference experiments, how (m)RNP particles dynamically behave in different cellular contexts.



This figure shows the schematic of the RIPPLiT workflow applied to the EJC in a mammalian cell line. Here, the authors adapt their previously developed technique (see Singh et al., Methods 2014, doi: 10.1016/j.ymeth.2013.09.013) for the initial steps of the protocol, followed by ligation of proximal RNA ends and sequencing. – This is Figure 1A from the preprint, made available under a CC-BY-NC-ND 4.0 International license.

Tags: 3D conformation, contact frequency, mrnp organization, ripplit, technique

Posted on: 2nd April 2018 , updated on: 4th April 2018

4. An open information resource for patient-derived tumor models: PDX Finder as a searchable portal to improve visibility, transparency and communication in cutting-edge cancer research.

PDX Finder: A Portal for Patient-Derived tumor Xenograft Model Discovery

Nathalie Conte, Jeremy Mason, Csaba Halmagyi, Steven B Neuhauser, Abayomi Mosaku, Dale A Begley, Debra M Krupke, Helen Parkinson, Terrence Meehan, Carol J Bult
Preprint posted on April 05, 2018 <https://www.biorxiv.org/content/early/2018/04/05/291443>

Selected by Carmen Adriaens

Categories: cancer biology, cell biology, clinical trials, genetics, genomics

Background

In the past 15 years, tens of laboratories all over the world have implemented patient-derived tumor xenograft (PDX) models as a powerful tool to study cancer. The principle is simple: upon biopsy in the clinic, a piece or a suspension of the resected tumorous tissue is engrafted either orthotopically or subcutaneously in an immunocompromised mouse. If the graft is successful, the grown tumor can be re-transplanted (expanded) into another set of mice (2nd generation) and so on, until several mice bear the same parent tumor and can be used for experiments. Over generations of host mice, the human tumor stroma is usually replaced by mouse stroma, but the cancerous tissue retains its principal molecular and histological characteristics.

PDX models have several advantages over in vitro- or cell line graft- based techniques. For instance, they retain well the cellular heterogeneity found in the original tumors and they usually respond to a treatment in a manner similar to the donor. They can even be predictive of resistance development or response rate. Moreover, as opposed to genetically engineered mouse models (GEMMs), the cells are of human origin, and data generated for the models is directly relevant to the study of human cancer biology.

PDX models are used for several cancer-related applications. For instance, the engraftment rate and success of a tumor can be examined diagnostically to determine stage and aggressiveness. They are employed to resolve important biological questions such as what the tumor heterogeneity is, whether a cell-of-origin exists, or if certain tumor cells are intrinsically resistant to therapy. Recently, the field has imagined a more clinically relevant application: although expensive and not devoid of risk (and if anything, still in a very controversial light), PDX models could be used as “avatars” for cancer patients. In such a scenario, several drug combinations and treatment courses can be tested and characterized before an actual therapy scheme is established in the clinic.

The problem

“Alone, we can do so little; together, we can do so much.” — Helen Keller

Although many labs have developed PDX models, the techniques to do so are not standardized and the reporting and availability of data, especially, are limited and scattered, or

very heterogeneous among different groups. In a community effort, several world-wide consortia recently established PDX-MI, a Minimal Information set of guidelines that recommends PDX users to report “essential” and “desirable” data modules when new models or research is published (Meehan, Conte et al. *Cancer Research*, 2017). Since the use of PDX is rapidly expanding, there is a pressing need for a centralized database containing the available models and their minimal information. A such, they can be shared globally to prevent redundant studies and the perfect PDX and associated data can be easily found for the specific research question asked.

The preprint

In this preprint, Conte et al. present PDX Finder (www.PDXfinder.org), an online database aiming to solve the decentralization of available models and the information about them. For now, seven PDX centers (one in Italy, six in the US) have joined the database. The authors invite all other centers from around the world to participate.

PDX finder provides a searchable platform for PDX models of different cancer types, and searches can be narrowed to the geographical location of the model, the anatomical system, collection site, molecular characteristics, type of host and patient, mouse generation etc. It will be updated with clinically and biologically relevant information, where available, and aims to improve the visibility of cancer models for future studies.

My opinion

Overall, I think this preprint is a clear and concise description of the website and serves its citation and dissemination purposes well. Importantly, I believe this manuscript is a worthy attempt to address a recurrent issue in modern science: the lack of transparency, standardization, data availability, and mobility of reagents and skills.

This preprint and the PDX Finder website is a small but important step forward by aiming to centralize the information. It may help to increase the visibility of available models and stimulate communication between biologists, clinicians, and informaticians. Hopefully, the platform will encourage exchange among different centers and laboratories, which will be especially valuable when done within the framework of the PDX-MI. Furthermore, the use of animal models for research and clinical purpose is obviously not without controversy. This database may hence contribute to the three R's – Replace, Reduce, Refine. A central platform can help to avoid duplicate models for fundamental research, and in contrast, may reveal gaps in molecular subtypes for individual tumor type PDX models.

In conclusion, I believe that, although as with all animal models their use needs reasonable caution, overall PDX models are an excellent and robust tool to study cancer. In past years they appear to have taken center stage successfully and seem ready to revolutionize cancer research. I like to think we all strive towards the same objective of understanding and trying to cure cancer, and we will need to bundle our efforts to succeed. Ultimately, PDX Finder may help to achieve that goal in an effective and transparent manner.

Tags: database, disease model, issues in science, patient derived tumor xenograft, worldwide consortia

Posted on: 20th April 2018

5. It takes two to (un)tangle: Dual Lac/Tet operator arrays report on chromatin dynamics in live cells

Live-cell imaging of marked chromosome regions reveals dynamics of mitotic chromosome resolution and compaction

John K Eykelenboom, Marek Gierlinski, Zuojun Yue, Nadia Hagarat, Hillary Pollard, Tatsuo Fukagawa, Helfrid Hochegger, Tomoyuki U Tanaka

Preprint posted on April 20, 2018 <https://doi.org/10.1101/305391>

And

Quantitative imaging of chromatin decompaction in living cells

Elisa Dultz, Roberta Mancini, Guido Polles, Pascal Vallotton, Frank Alber, Karsten Weis

Preprint posted on April 22, 2018 <https://doi.org/10.1101/219253>

Selected by Carmen Adriaens and Gautam Dey

Categories: bioengineering, biophysics, cell biology, molecular biology

Interphase cell nuclei are compartmentalized into loose, transcribed euchromatin and compact, largely untranscribed heterochromatin. The consensus is that for a given gene to be transcribed, the active chromatin needs to decompact further through changes in the chromatin landscape. Another form of chromatin dynamics occurs during cell division. Indeed, to divide correctly, the DNA is neatly folded into an extreme state of compaction (the mitotic chromosome) after replication. In two recent independent studies, researchers have used a two-color live cell reporter system to study chromatid separation before and chromatin compaction during mitosis, and the process of decompaction upon transcriptional activation. Since the experimental setup is very similar, we will discuss these papers jointly and highlight the different insights obtained from this live cell imaging system.

The experimental setup

At a specific site in the genome, a lactose operon array (LacO) is integrated in the DNA. The repeats can be recognized by the Lac inhibitor protein (LacI), which is here overexpressed and tagged with a green fluorescent protein (GFP) to be visualized as a single dot per lacO in the cells. At another site in the DNA, a Tet operon, TetO, is integrated and its cognate protein (the Tet repressor, TetR) is tagged and visualized with the red fluorescent reporter mCherry. The two dots (green and red) can be followed using microscopy in live cells, to enable the monitoring and modeling of chromatin dynamics at high resolution in real time in different biological settings.

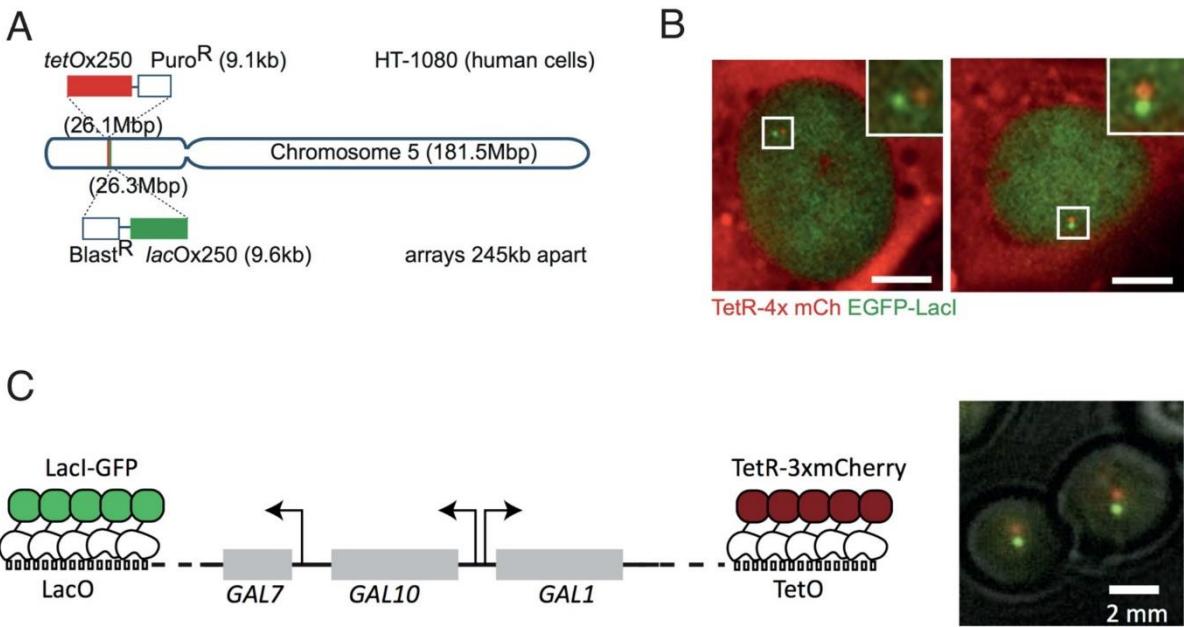


Figure 1: (A) Experimental set-up in human HT-1080 cells used by Eykelenboom et al. with sample images in (B), scale bar = 10 μm. (C) Experimental set-up in budding yeast used by Dultz et al. with sample images on the right. Taken from Figure 1 of Eykelenboom et al. 2018 and Figure 1 of Dultz et al. 2018 under Creative Commons CC-BY-4.0 licenses.

What are the papers about?

Eykelenboom et al. used CRISPR to integrate the Tet and Lac operator arrays separated by a 250 kb interval on one arm of chromosome 5 in HT-1080 human cells. Adding in a 4x-mCherry-tagged Tet repressor and an EGFP-tagged Lac repressor enabled them to follow dynamics of chromatid segregation and compaction throughout the unperturbed cell cycle in single cells. Imaging the 2 green and 2 red spots, representing the chromatids after replication, revealed 4 discernible states: non-resolved, partially resolved, resolved and compacted.

Surprisingly, cells begin to cycle between the partially resolved and unresolved states rather early in G2. This cycle is driven by an apparent antagonism between cohesin holding chromatids together on one hand, and WAPL trying to remove cohesin from the chromosome arms on the other. Exploiting a fortuitous side effect of the integration sites of their operator arrays, the authors showed that the timing of separation depends on the local cohesin concentration. Inhibiting topoisomerase II activity, and thereby preventing the resolution of DNA catenanes, caused partially resolved chromatids to revert to the unresolved state early in mitosis. Finally, the authors showed that condensin II depletion caused defects in sister chromatid separation, while condensin I depletion affected only the final compaction stage.

In a second paper, Dultz and colleagues used very similar experimental principles to study chromatin decompaction upon transcriptional activation. The authors integrated the LacO and TetO repeats on either side of the budding yeast GAL locus and studied chromatin dynamics by measuring distances between red and green dots in different conditions. In the absence of glucose and galactose, the locus is inactive, but derepressed. In the presence of glucose, the locus is actively repressed. In both scenarios, the distance between the dots was small, reflecting an inactive chromatin state. When, on the other hand, the cells were grown in or

induced with galactose, the locus became highly active and the 3 linked genes (GAL7, GAL10 and GAL1) in the locus became transcribed. In this active state, the green and red dots lining the GAL locus on either side reversibly increased in distance, reminiscent of decompaction and transcriptional activation.

Next, the authors asked whether these observations are dependent specifically on either spatial or linear decompaction. Spatial decompaction is primarily mediated by post-translational modifications of the histones and hence manipulation of nucleosome-nucleosome interactions, whereas in linear decompaction the wrapping density changes through eviction and remodeling of nucleosomes at the site of transcription. They found that a change in histone modifications, and, by extension, inter-nucleosomal interactions, does not dramatically alter the decompaction dynamics. Conversely, both manipulation of transcription itself and nucleosomal density do.

These observations led to the hypothesis that decompaction of chromatin is principally mediated by the number of nucleosomes evicted. To test this, the authors set up an exogenous reporter system 10 kb upstream of the original locus and measured changes in distances between dots when no, one, or two open reading frames were present. In this way, they found that the extent of decompaction indeed correlates with the length of the transcribed region. Finally, the authors concluded that although decompaction is transcription dependent, the inverse isn't necessarily true: cells can continue to transcribe the GAL locus even in the absence of significant changes in distances between the two reporter dots.

Takeaways and questions for the future

It's a very exciting time for the field of chromatin reorganization and dynamics! To take just one example, new experimental approaches have validated an old molecular model for condensin function¹ and mapped a pathway to mitotic chromosome organization mediated by condensin I and II². In parallel, a proliferation of live cell reporters have uncovered layers of cell cycle regulation that were hitherto invisible to bulk assays^{3,4}. Taken together, these advances represent the perfect storm to preface an analysis of chromosome dynamics in single cells as they progress through the cell cycle.

The most surprising finding from the first paper (Eykelenboom et al.) is that chromosomes cycle between unresolved and partially resolved states through most of their G2 phase, driven by an apparent concentration-dependent antagonism between cohesin and WAPL.

What then is the purpose of this (futile?) cycle, and how is it broken during prophase? Saturated feedback loops of this type can be used to generate switch-like responses (also termed zero-order ultrasensitivity⁵). In turn, then, one could ask: is it necessary to have a switch-like transition between unresolved and resolved chromatids during prophase?

Moreover, the authors' observations complement other recent work² helping to resolve long-standing questions in the field about the individual and collective roles of the two condensin complexes.

It has long been recognized that actively transcribed chromatin is less compact than inactive chromatin, and much is known about the packaging and histone modifications responsible for these states. However, being able to visualize actual decompaction directly in live cells is impressive: Dultz et al. leverage the ultra-simplicity of the experimental setup to reveal the solution of the chicken-and-egg problem of loose chromatin during transcription.

Yet, there may be some caveats to these experiments as well: because the GAL locus is highly active when growing in galactose and the regulation of the system is rather unique compared to the rest of the yeast genome (ON/OFF), these observations may confer a local rather than a global rule. The physical decompaction (i.e. the increase in physical distance between the two arrays) in other loci may not be as striking as here, and will potentially be more difficult to study.

Furthermore, we note that both studies employ population averaging to draw certain conclusions, faced with high levels of cell-to-cell variability. Other recent work⁶ shows that chromosomal contacts observed by high-throughput sequencing approaches such as Hi-C occur as infrequently as in ~10% of cells only (as observed by microscopy), highlighting the extraordinary differences even within relatively homogenous or synchronized populations of cells. Additional work will be required to understand the sources of this variability and its functional consequences, if any.

References

- Ganji, M. et al. Real-time imaging of DNA loop extrusion by condensin. *Science* 360, 102–105 (2018).
- Gibcus, J. H. et al. A pathway for mitotic chromosome formation. *Science* 359, eaa06135 (2018).
- Chao, H. X. et al. Evidence that the cell cycle is a series of uncoupled, memoryless phases. *bioRxiv* 283614 (2018). doi:10.1101/283614
- Spencer, S. L. et al. The Proliferation-Quiescence Decision Is Controlled by a Bifurcation in CDK2 Activity at Mitotic Exit. *Cell* 155, 369–383 (2013).
- Ferrell, J. E. & Ha, S. H. Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity. *Trends Biochem. Sci.* 39, 496–503 (2014).
- Finn, E. et al. Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *biorXiv* (2017). doi: <https://doi.org/10.1101/171801>

Tags: cell cycle, chromatin dynamics, live cell imaging, teamwork makes the dream work, transcription

Posted on: 15th May 2018

6. A powerful call for reproduction studies: overexpression of a prostate cancer specific non-coding RNA does not function by depletion of SWI/SNF from chromatin to increase invasiveness.

SWI/SNF remains localized to chromatin in the presence of SChLAP1

Jesse R Raab, Keriayn N Smith, Camarie C Spear, Carl J Manner, J. Mauro Calabrese, Terry Magnuson

Preprint posted on May 14, 2018 <https://doi.org/10.1101/322065>

Selected by Carmen Adriaens

Categories: biochemistry, bioinformatics, cancer biology, cell biology, genetics, molecular biology

The preprint and its context

In recent years, the scientific community together with funding agencies and publishing companies more strongly recognizes the necessity for reproduction studies. Fortunately, the perceived value of these studies lies not only in calling out irreproducible work or revealing scientific misconduct, but also in strengthening observations from colleagues in the field and in helping to put forward different interpretations for existing work when more information is obtained and new ways of proving a hypothesis are developed.

In 2013, the group of Prof. Chinnaiyan at the Michigan Center for Translational Pathology published a study showing that high expression of a novel long non-coding RNA (lncRNA), SChLAP1, can be used as a prognostic factor for poor outcomes in prostate cancer patients (Presner et al., Nat. Gen., 2013). Using knockdown and overexpression assays, the authors elegantly demonstrated that SChLAP1 increased the invasiveness of prostate cancer cells both *in vitro* and *in vivo*. Mechanistically, they proposed that this lncRNA is bound by the SWI/SNF chromatin remodeling complex known to both positively and negatively influence gene expression. Through SChLAP1 binding, the complex would be depleted from the chromatin to increase the expression of pro-invasion and pro-metastasis genes, hence causing the phenotype. SChLAP1 functioning was interpreted based on two observations: (1) the lncRNA binds directly to the SNF5 (SMARCB1) subunit of the SWI/SNF complex; (2) SChLAP1 overexpression causes genome-wide loss of SNF5 binding to chromatin as assessed by ChIP-sequencing.

In a recent preprint, the group of Prof. Magnuson at the University of North Carolina partially challenged and partially corroborated the above findings. On the one hand, they provided additional evidence that SChLAP1 has an important role in the proliferation, invasion and metastasis of prostate cancer cells and they validated that the SWI/SNF subunit SNF5 physically interacts with this lncRNA. On the other hand, the authors performed a series of experiments to determine whether SWI/SNF is globally depleted from the chromatin. They found that SChLAP1 overexpression did not affect broad SWI/SNF chromatin localization and binding both with the same experimental approach in the same conditions and with the same and additional reagents as in the original study. Probing changes in chromatin accessibility by ATAC-sequencing, the authors then found that the chromatin state of a set of pro-invasive

genes (GO terms NFKB signaling, epithelial to mesenchymal transitions, and nucleotide metabolism) is more open, although they could not pinpoint a direct role for the SWI/SNF complex at these loci nor find evidence in the microarray data from Presner et al. for changes in their expression. Finally, the authors demonstrated that the observed binding of SChLAP1 to SWI/SNF is due to a rather promiscuous binding ability of SWI/SNF to RNA rather than to a specific effect between this lncRNA and the complex.

My opinion

Why this preprint caught my attention is because it is different from most scientific literature: it is a reproduction study. Although the scientific community agrees that these kinds of studies are necessary, still very few are conducted, and even less are published in high-visibility journals. The reasons for this are manifold. For instance, it may often make the challenger unpopular with the author(s) of the initial work, or, there is no money available to conduct a study much less flashy and novel than when original research is performed. However, the current preprint proves they are a must.

Indeed, this preprint is an excellent example of a rigorous reproduction study: although not all the results were reproduced, part of them are, and the validity of specifically these findings are strengthened independently. In the case of consistent SChLAP1 upregulation in aggressive prostate cancer, and its role in the metastatic propensity of the tumor, this could for instance prompt a company to confidently develop SChLAP1 detection as a biomarker to determine prostate cancer malignancy. Conversely, the authors took several diverse experimental approaches to show that the initially proposed mechanism of action was probably not real. Therefore, as the authors stated in the beginning of the work, an alternative hypothesis needs to be explored and I look forward to discovering what the actual mechanism of action of SChLAP1 is.

Finally, I really like the way the authors approached their biological question. I like to imagine that the concept under study is a physical object (e.g. a cube), and each face provides a different angle to look at it. Here, SChLAP1 tethering away the SWI/SNF complex from the DNA was the cube. Maybe the authors of the original article only looked at one or two faces of it, and through looking at it from multiple angles in this preprint, a different, more accurate and more complete view was obtained, and the cube was not at all what it originally seemed to be. (Figure 1)

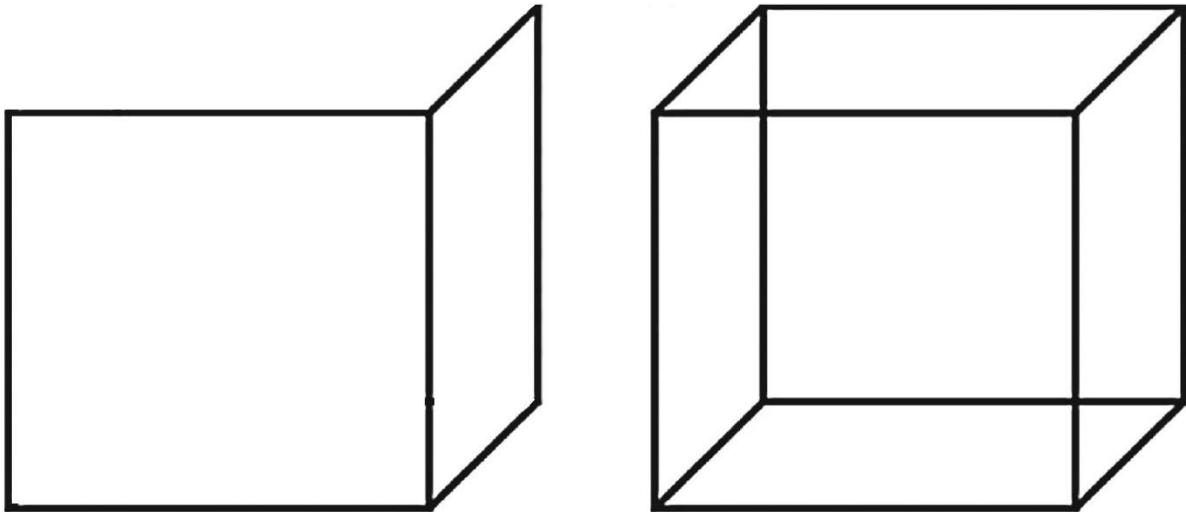


Figure 1: Two (left) vs the many (right) faces of a “conceptual cube” in science give a very different view on what the object looks like.

Questions and concluding remarks

What prompted the authors to conduct this study? Which event or observation or results made them doubt the proposed mechanism of action?

On a more philosophical note, would access to the peer review file have helped to understand the publishing of the results? Would the absolute requirement for mechanism in high-impact factor journals lead to the risk of either over-interpretation of the work, or a potentially less careful evaluation of a striking result? (Admittedly, this is a different debate.)

From a more experimental point of view, to assess whether the dynamic chromatin sites also change in expression, the authors used the microarray data from Presner et al. I wonder if it would not be necessary to simultaneously probe gene expression changes and chromatin dynamics to be able to reach the conclusion that the one doesn't influence the other, especially because the GO terms in the ATAC experiment point to the “opening” of genes that could explain the phenotype.

Finally, the biology of SChLAP1 seems to be different in immortalized and transformed cells. Maybe both papers use the wrong cell type to assess the mechanism, and a system that also displays the invasion phenotype would be more appropriate to find out how this lncRNA works?

Posted on: 15th June 2018

7. Lessons on chromatin architecture from senescent cells: Nuclear pore complex density drives internal heterochromatic focus formation

Nuclear pore density controls heterochromatin reorganization during senescence

Charlene Boumendil, Priya Hari, Karl Olsen, Juan-Carlos Acosta, Wendy Bickmore
Preprint posted on July 21, 2018 <https://www.biorxiv.org/content/early/2018/07/21/374033>

Selected by Carmen Adriaens

Context

In non-transformed cells, excessive oncogene expression has long been known to cause oncogene-induced senescence (OIS). This cellular state is marked by several features, including permanent proliferative arrest, a displacement of the normally peripherally located heterochromatin to several distinct internal areas (Senescence-Associated Heterochromatin Foci, SAHF), and an elevated secretion of immune signaling molecules (SASP, the Senescence Associated Secretory Phenotype).

Recent findings suggest that the peripheral localization of mostly inactive heterochromatin is established by an active tethering mechanism of the chromatin to the nuclear lamina1, thought to protect the nuclei from mechanical assault2. On the contrary, internal heterochromatin is proposed to be the default state of its organization, as proposed from observations of chromatin architecture in nocturnal animal eye rod cells3, 4. Indeed, it is hypothesized that the peripheral heterochromatin of most cells is maintained by a careful balance of repulsing and attracting forces from and to the nuclear envelope. Additionally, to maintain a stable flux of molecules between the nucleus and the cytoplasm, the nuclear envelope contains numerous nuclear pore complexes (NPCs). At these large multiprotein complexes, the heterochromatic fibers are excluded, largely through the repulsing action of the Translocated Promoter Region protein (TPR) component of the NPC5.

Although it has been proposed that the induction of SAHF is in part due to reduced interactions of heterochromatin with the nuclear periphery through, among others, lower levels of Lamin B16, the role of NPCs, and potentially their repelling forces in establishing SAHF is unknown. Moreover, the order of phenotypic manifestations of OIS is not well understood, i.e. it is unclear if the SASP is required for SAHF formation and cell cycle arrest, or if cell cycle arrest is required for SAHF.

The preprint

In this preprint, the authors hypothesized that OIS would lead to an increase in NPCs, thus promoting the establishment of internal heterochromatic foci. During differentiation, normal quiescent cells downregulate their nucleoporin mRNA expression to stabilize the number of NPCs. In contrast, using an inducible constitutively active Ras overexpression system leading to OIS in non-transformed lung fibroblasts, the authors show that the mRNA levels of NPC components are stabilized, resulting in an increased number of NPCs. By further manipulating the levels of either an internally located component (POM121) or the nuclear interior-facing

protein TPR, they find that the increase in NPCs is required for the establishment of the senescence associated heterochromatic foci. Consequently, the authors show that knockdown of the NPC not only prevents SAHF but also abrogates the senescence associated secretory phenotype, arguing that the correct order of events is cell cycle arrest, followed by SAHF and then SASP. The requirement of SAHF for SASP was further strengthened by knocking down a chromatin chaperone preventing SAHF formation (ASF1). Indeed, with this experiment they could effectively halt the secretory phenotype upon SAHF disruption, while the number of NPCs remained the same.

In conclusion, heterochromatic focus formation in oncogene induced senescence requires an increased number of nuclear pore complexes, leading to the sequential induction of two of its main phenotypic manifestations, SAHF and SASP. This process is proposed to function through an increased repulsing force of the chromatin by the higher numbers of NPCs.

My opinion

I like this preprint for several reasons. First and foremost, the study is straightforward, has a clear rationale and hypothesis, and confers simple but on-point experiments to address it. A second reason is that the findings provide a major advancement in the thinking of how, in addition to the lamina, the NPC helps the nucleus organize into mainly peripheral heterochromatin and central euchromatin. It allows us to ask what the evolutionary drives and functional advantages are for this organization. Furthermore, the study proposes an almost effortless mechanism by which this internal SAHF occur in the cell – especially if the hypothesis on repulsing and attracting forces turns out to be correct. Moreover, in line with these observations, recent work has suggested that heterochromatin forms a phase-separated compartment in the nucleus⁷. It would be interesting to further study in this context how the heterochromatin upon establishment of SAHF behaves on the physicochemical level, and to what extent NPCs influence these behaviors.

Considering the recent findings on heterochromatin organization in the rod cells of nocturnal animals, I also wonder what the role of the nuclear envelope and its interactions with the chromatin are in these cell types, and if the NPC may also contribute to the nuclear architecture in this specific context. Additionally, I am curious to learn more about the hypothesized forces that drive this organization both in senescent cells and in normal circumstances. One might ask, in light of the findings of this preprint, how NPCs contribute to general chromatin architecture during differentiation, proliferation, etc.

References

- 1 Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Van Steensel B and Belmont AS, (2017) Cell 169, 780-791.
- 2 The tethering of chromatin to the nuclear envelope supports nuclear mechanics. Schreiner et al. (2015) Nature Communications (6), 7159
- 3 Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. Solovei et al. (2009). Cell 137, 356–368.

4 Heterochromatin drives organization of conventional and inverted nuclei. Falk et al. (2018) bioRxiv, <https://doi.org/10.1101/244038> – preLighted by Boyan Bonev

5 Protein Tpr is required for establishing nuclear pore-associated zones of heterochromatin exclusion. Krull et al. (2010) *Embo J*: 29 (10).

6 Redistribution of the Lamin B1 genomic binding profile affects rearrangement of heterochromatic domains and SAHF formation during senescence. Sadaie et al. *Genes Dev.* (2013) ;27(16):1800-8

7 The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. Larson et al. (2018) *Biochemistry*. 57, 2540–2548

Posted on: 30th July 2018 , updated on: 1st August 2018

8. From 2D to 3D: A chromatin interactome map of human pancreatic islets provides a 3D view into type 2 diabetes and pancreatic islet biology

Human pancreatic islet 3D chromatin architecture provides insights into the genetics of type 2 diabetes

Irene Miguel-Escalada, Silvia Bonàs-Guarch, Inès Cebola, Joan Ponsa-Cobas, Julen Mendieta-Esteban, Delphine M.Y. Rolando, Biola M. Javierre, Goutham Atla, Irene Farabella, Claire C. Morgan, Javier García-Hurtado, Anthony Beucher, Ignasi Morán, Lorenzo Pasquali, Mireia Ramos, Emil V.R. Appel, Allan Linneberg, Anette P. Gjesing, Daniel R. Witte, Oluf Pedersen, Niels Garup, Philippe Ravassard, David Torrents, Josep Maria Mercader, Lorenzo Piemonti, Thierry Berney, Eelco J.P. de Koning, Julie Kerr-Conte, François Pattou, Iryna O. Fedko, Inga Prokopenko, Torben Hansen, Marc A. Martí-Renom, Peter Fraser, Jorge Ferrer

Preprint posted on August 27, 2018 <https://www.biorxiv.org/content/early/2018/08/27/400291>
Selected by Carmen Adriaens

Introduction and context

One of the main two types of diabetes, diabetes mellitus (type 2, T2D) is a polygenic disease that occurs usually in obese subjects and results in systemic peripheral tissue insulin resistance (DeFronzo et al. 2015). A popular approach to study common polygenic diseases is to look at genetic variants that occur in the population and correlate them with the risk for the disease (Torkamani et al. 2018). In the case of T2D, several recent efforts have implicated common Single Nucleotide Polymorphisms (SNPs) in the T2D phenotype and identified a number of susceptibility loci e.g. (K. Sanghera and R. Blackett 2012; Mishra and Hawkins 2017; Mahajan et al. 2018). However, deriving functional and causal information from these variants and their associated genes often proves difficult, especially when they occur in the non-coding portion of the genome, and when information is lacking on what the target genes are and how the variants affect them.

So far, previously identified T2D SNPs and their function in the T2D phenotype have only been studied in the context of a linear DNA map (eg. Which cell-specific transcription factors bind in the enhancer clusters in which the SNP occurs? Or, how do the different variants contribute to gene expression within the Topologically Associated Domain (TAD)?). However, how these SNPs specifically contribute to the disease phenotype on the molecular level in the context of the 3D genome remains to be elucidated.

The preprint

In a recent preprint, members of the group of Prof. Jorge Ferrer and colleagues set out to construct a chromatin interactome map of human pancreatic islets, with the goal to determine beta-cell specific enhancer-promoter interactions and to link variation in the non-coding portion of the genome to the relevant target genes. First, they combined ATAC-sequencing and ChIP-sequencing for Mediator, Cohesin and H3K27Ac with promoter-capture hi-C (pcHi-C). This latter technique increases sensitivity of Hi-C by using fragment baits that cover all known promoter regions (Schoenfelder et al. 2018) to detect lower-frequency, cell type-specific

interactions between promoters and their potential regulatory sequences. These experiments demonstrated that islet-specific interactions are strongly enriched for islet-specific enhancers showing ample binding of Mediator, a protein complex known to regulate gene expression by bridging the signals from transcription factors in regulatory regions to RNA Polymerase II. Because these enhancers also showed the most prominent H3K27 acetylation, the authors named these enhancers Class I enhancers.

To identify the target genes of (Class I) pancreatic islet enhancers from Hi-C, the authors first needed to address an important limitation of the Hi-C technique: its data analysis is highly stringent and shows a strong proximity bias and excessive tissue/physiological condition specificity. Therefore, cell type-specific DNA proximity outcomes can easily be underestimated. To overcome this problem, the authors imputed additional potential interactions based on promoter-associated 3D space (PATs). Here, a PAT is defined as the ‘space’ that is constructed from taking together all pCHi-C interactions of a specific bait. Like this, they found that the PATs that show high confidence interactions with islet specific enhancers have a higher chance of also interacting with other enhancers. Next, the PATs were used to identify the target genes of the T2D-relevant enhancers (i.e. those that showed disease-relevant regulatory variation in their sequences). These functional interactions were then validated with several approaches, a.o. using CRISPR-Cas9 modulation of the PAT enhancers, some computational work, and a dynamic-perturbation strategy in which the activity of the enhancers could be modulated by challenging the cells with high or low glucose media.

Next, the authors assessed the T2D-associated SNPs locating to these 3D regions and asked if they impact on the structure and function of enhancer ‘hubs’ defined by the PATs and their secondary interactors (i.e. their target genes, and other interacting enhancers). Suggestive of a strong tissue-specific role, they observed that the SNPs in the 109 T2D – Fasting Glucose (FG) risk loci were enriched in high-confidence pCHi-C regions with increased Class I (high H3K27Ac – Mediator occupancy) enhancers. Furthermore, they showed also that these SNPs showed higher genetic heritability scores, and thus that they play a prominent role in the heritability of islet function and T2D.

Finally, the authors asked if they could use the islet enhancer hub SNPs to better predict the risk for T2D. They found that, although quantitatively the hub-SNVs were not better at predicting T2D risk than the collection of all common SNPs genome-wide, hub-SNPs could help stratify patients qualitatively. For instance, the Polygenic Risk Scores (PRS) calculated from hub-SNPs in non-obese, younger patients showed higher Odds Ratios than the PRS in obese, older patients, indicating that in these younger individuals the variation occurring in the genomic regions that affect islet transcription impacts more strongly on T2D. Like this, the authors postulated that hub-only SNVs may be useful to further study how islet-specific regulatory variation affects T2D pathogenesis.

What I think about this preprint & open questions

I chose this preprint because it addresses a critical question in cell / systems biology: How does the three-dimensionality of chromatin contribute to cellular identity and disease?

First of all, I like that the 3D interactome map constructed by the authors is browsable with a user-friendly interface; and thus that any individual could go to his/her favorite gene and look

at its 3D environment. Furthermore, the authors push the limits of their data by several means, and, although many of these imputed interactions A) may be dynamic/transient and B) may need to be subjected to experimental validation, I love the idea of using ‘hubs’ to represent cell type specific, 3D spatial interactions. Indeed, with this approach, the authors overcome some of the limitations inherent to genome-wide chromatin-capture techniques and identify potential novel, cell type specific or even disease-specific domains, laying the groundwork for further studies.

The fact that many of the ‘hub’ interactions are imputed may also confer a strong limitation to the work: there is no direct evidence that these hubs or all the interactions within them are actually formed (simultaneously), and interactions may be mutually exclusive, or temporally distinct. Therefore, it would be interesting to study if these interactions occur and what their dynamics are using for instance imaging-based methods (e.g. DNA-FISH, or CRISPR/Cas-guided locus identification in live cells, see eg. (Chen et al. 2013)).

With this work, another question emerges: are these 3D interactions functionally relevant to pancreatic islet biology? For instance, would it be interesting to study a time course and see if the enhancer becomes inactive (loses active chromatin marks) before its target gene expression does? Are the 3D connections lost in these and other (experimental/physiological) circumstances?

I am further intrigued by the Class I enhancers defined in this study by high Mediator occupancy and H3K27Ac. Several recent studies (eg. Cho et al. 2018; Sabari et al. 2018) demonstrated that Mediator and RNA Pol II are involved in a phase-separated complex involving super-enhancers. Thinking of how 3D chromatin architecture helps to establish cell identity mechanistically, the question emerges whether the observations in these recent papers can be extended to the islet-specific Class I enhancers observed in this study. More generally, does the liquid demixing through Mediator and RNA Pol II binding provide a chemically favorable environment to establish relevant DNA-DNA contacts in a cell-type specific manner? Thus, it will be interesting to study further what these Class I enhancers are, how they emerge and are maintained, and if they contribute to phase-separation mediated transcriptional output. Furthermore, it remains to be determined what the driving forces are to establish the liquid-demixed environment in this cell type specific manner.

In conclusion, I like this study because it advances the understanding of 1. islet-specific enhancer-promoter interactions and their potential influence on cell-type specific gene expression in the context of human pancreatic beta cell islets; and 2. the functional and mechanistic role of disease-associated variants in establishing relevant interactions in the 3D genome leading to disease through disruption of the normophysiological 3D architecture. Finally, this study also demonstrates the importance for continuous validation studies of predicted interactions in a cell type or physiological condition specific manner, and ultimately may be an important resource to further our understanding of cell identity and how it is influenced by local chromatin architecture.

References/Further reading

- Chen B, Gilbert LA, Cimini BA, et al (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155:1479–1491. doi: 10.1016/j.cell.2013.12.001
- Cho WK, Spille JH, Hecht M, et al (2018) Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361:412–415. doi: 10.1126/science.aar4199
- DeFronzo RA, Ferrannini E, Groop L, et al (2015) Type 2 diabetes mellitus. *Nat Rev Dis Prim* 1:15019
- Sanghera D, R. Blackett P (2012) Type 2 Diabetes Genetics: Beyond GWAS. *J Diabetes Metab.* doi: 10.4172/2155-6156.1000198
- Mahajan A, Taliun D, Thurner M, et al (2018) Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *bioRxiv*. doi: 10.1101/245506
- Mishra A, Hawkins RD (2017) Three-dimensional genome architecture and emerging technologies: Looping in disease. *Genome Med* 9:87. doi: 10.1186/s13073-017-0477-2
- Sabari BR, Dall'agnese A, Boija A, et al (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361:6400. doi: 10.1126/science.aar3958
- Schoenfelder S, Javierre B-M, Furlan-Magaril M, et al (2018) Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions. *J Vis Exp* 136:. doi: 10.3791/57320
- Torkamani A, Wineinger NE, Topol EJ (2018) The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* 19:581–590. doi: 10.1038/s41576-018-0018-x

Posted on: 20th September 2018, updated on: 6th October 2018

9. An (un)expected surprise? NORAD lncRNA regulates tissue homeostasis, genome integrity and mitochondrial function in vivo.

PUMILIO hyperactivity drives premature aging of Norad-deficient mice

Florian Kopp, Mehmet Yalvac, Beibei Chen, He Zhang, Sungyul Lee, Frank Gillett, Mahmoud Elguindy, Sushama Sivakumar, Hongtao Yu, Yang Xie, Prashant Mishra, Zarife Sahenk, Joshua T Mendell

Preprint posted on October 01, 2018 <https://www.biorxiv.org/content/early/2018/10/01/432112>

Selected by Carmen Adriaens

Categories: biochemistry, bioinformatics, cell biology, developmental biology, evolutionary biology, genetics, genomics, molecular biology

Introduction/background

After the realization that a substantial amount of the transcribed genome does not code for proteins, long non-coding RNAs have been an intense subject of study in the past couple of decades. However, still few have been shown to be essential in vivo, and even less are really functionally and mechanistically understood. One of the more recently studied lncRNAs is Noncoding RNA activated by DNA damage, NORAD, named after its induction upon doxorubicin treatment. NORAD has been initially shown to interact with Pumilio proteins involved in post-transcriptional regulation of RNA targets in the cytoplasm. With its 15+ conserved Pumilio Response Elements (PREs), NORAD was shown to be the preferential binding partner of these proteins, thus serving as a decoy (much like lncRNA can serve as a decoy for microRNAs), so that other PUMILIO targets are not downregulated when they are needed in the cell^{1,2}. An important subset of these targets are mRNAs involved in processes such as chromosomal segregation in cell division and DNA replication^{1,2}. When NORAD is lost, Pumilio proteins become hyperactive and excessively inhibit these targets thus causing aberrant mitosis and aneuploidy. Another recent study has identified NORAD as a central scaffold of a novel nuclear complex, NARC1 (NORAD Activated Ribonucleoprotein Complex 1), which is built on the RNA binding protein RBMX3. Upon DNA damage, NARC1 is formed by shuttling NORAD from the cytoplasm back to the nucleus, where it tethers Topoisomerase I and other factors important for genomic stability together with RBMX.

Although a variety of functions and binding partners of NORAD have been investigated by different groups using cultured mammalian cells^{1–5}, its functions have so far not been explored in vivo. In a recent preprint, the group of Prof. J.T. Mendell tackled the question of whether Norad was also, and similarly, functional in living tissues by making a mouse model constitutively deleted for the mouse Norad ortholog, and comparing the phenotypes in these mice with those overexpressing PUM2.

What are the main findings?

First, the authors found that although Norad KO mice were viable and developed initially in a similar fashion as their WT littermates, soon after they reached adulthood they showed signs of premature aging. For instance, they observed increased baldness, greying hairs, spine deformations, weight loss and muscle and neuronal malfunction in Norad KO mice, and noted they had shorter lifespans as compared to wild types. When looking at the molecular cause for these phenotypes, they initially focused their attention to the well-studied interaction of Norad with Pumilio proteins, which had previously been implicated in aging. The authors could indeed identify a conserved interaction between Norad and Pum2, and found that the published canonical Pumilio targets were largely the same *in vivo* as they were in cultured cells. Furthermore, Pumilio occupancy on its targets was significantly increased in Norad KO mice, which led to a decrease in their expression in this model.

Next, the authors assessed if the aging phenotypes were due to aberrant mitosis previously reported *in vitro*. For this, they assessed the highly proliferative blood lineage and could indeed establish that both lymphocytes and splenocytes displayed increased aneuploidy.

Because muscle cells were heavily affected by Norad loss and the observed phenotypes pointed to aging-related diseases, the authors assessed if mitochondrial function was affected (for instance, it is known that mitochondrial defects can lead to premature aging⁶). They noted that Norad KO mice displayed severe mitochondrial abnormalities, including in their morphology and function, which led to a pathological accumulation of reactive oxygen species and oxidative damage, as well as faulty cellular metabolism and decreased mitochondrial respiration rates. Upon further investigation, they found that Norad-depleted tissues and cell lines showed significantly decreased levels of mitochondria-related genes, a subset of which were shown in PUM2 CLIP experiments to be direct targets of this protein.

As Norad was known to tether away Pum2 to prevent unwanted downregulation of its mRNA targets, the authors hypothesized that Norad KO phenotypes could be phenocopied upon Pum2 overexpression if it functioned in the same way in tissues as it does *in vitro*. Thus, to definitively establish the link between the Norad KO observed phenotypes and Pum2 hyperactivity, they created a mouse model in which doxycycline (dox) administration ubiquitously induced Flag-PUM2 expression. In this model, dox induction did not cause a significant increase in the levels of PUM2 protein, but rather seemed to replace endogenous protein with the tagged version, an observation potentially explained by the known tight regulation of PUM proteins via negative and positive feedback loops. When PUM2 was forcibly expressed, the mice neatly phenocopied, but with an even faster onset, the absence of Norad, both on the gross physiological level and on the cell biological level including the mitochondrial dysfunction.

Overall, the data in this preprint make a compelling case for Norad lncRNA function in adult tissue physiology through interactions with a tightly regulated RNA binding protein.

What do I think about the work?

In short: I really, really like it. It's a well-controlled study with a beautiful design and compelling results in living animals. I think it is eye-opening that this abundant lncRNA can have such late-

onset phenotypes, and I believe it urges the research community to look further than just the binary “viable or not” question. As this study shows, very highly conserved proteins can undergo lineage-specific fine tuning by less conserved RNAs (eg. here, in the mammalian lineage). Furthermore, I am always captivated by the idea that the biology of adult tissues can differ so significantly from these same tissues during development (why do the phenotypes only occur upon reaching adulthood? What changes are made, for instance in the chromatin, that cells become more sensitive to replication stress, and how are developing tissues protected from it when they need to quickly proliferate?).

Some more topical questions, of course, always remain. For instance, in light of the recent study by Munschauer et al., is the NARC1 conserved in mouse tissues, and if so, are any of the observed phenotypes related to NARC1 and the DNA damage regulating factors bound within it? The authors also mention that not all Norad KO mice show the same degree of phenotype. It would be interesting to determine what drives the penetrance of such phenotype, and if in tissues or individuals in which it is less severe, compensatory mechanisms can be identified. Finally, and this maybe just out of curiosity, are Norad KO mice more cancer prone as well? For instance, are blood cancers more prevalent due to the increased aneuploidy, and could this be an important cause of early death?

Ps. I initially named this paper when I saved its PDF “Norad KO pumilio gone mad”. Maybe not appropriate, but since I still like it as a title, I’ll make it my bottomline! :-).

References

- Lee, S. et al. Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* (2016). doi:10.1016/j.cell.2015.12.017
- Tichon, A. et al. A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat. Commun.* (2016). doi:10.1038/ncomms12209
- Munschauer, M. et al. The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature* 561, 132–136 (2018).
- Spiniello, M. et al. HyPR-MS for multiplexed discovery of MALAT1, NEAT1, and NORAD lncRNA protein interactomes. *J. Proteome Res.* 17, 3022–3038 (2018).
- Tichon, A., Perry, R. B. T., Stojic, L. & Ulitsky, I. SAM68 is required for regulation of pumilio by the NORAD long noncoding RNA. *Genes Dev.* (2018). doi:10.1101/gad.309138.117
- Trifunovic, A. et al. Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* (2004). doi:10.1038/nature02517

Tags: aging, lncrna, mitochondria, mouse models, norad

Posted on: 9th October 2018

10. It's not always one or the other: promoter sequence and local environment help gene repression and expression in heterochromatin.

Promoter-intrinsic and local chromatin features determine gene repression in lamina-associated domains

Christ Leemans, Marloes van der Zwalm, Laura Brueckner, Federico Comoglio, Tom van Schaik, Ludo Pagie, Joris van Arensbergen, Bas van Steensel

Preprint posted on November 06, 2018

<https://www.biorxiv.org/content/early/2018/11/06/464081>

Selected by Carmen Adriaens, Clarice Hong

Background

Heterochromatin is often correlated with a lack of transcription. However, it remains unknown whether heterochromatin actively represses genes or whether the genes within heterochromatin are just generally inactive due to the lack of appropriate transcriptional activators. In order to separate these two hypotheses, it is necessary to integrate genes into heterochromatin, or to remove a gene from its native heterochromatic context, and determine whether their intrinsic activity levels change. Furthermore, a small subset of genes within heterochromatic domains are expressed despite what is thought to be a repressive environment. Understanding the mechanisms of expression of these 'escaping' genes may provide insights into heterochromatin-associated repression as well as general transcription. In this preprint, the authors use the peripherally located lamina associated domains (LADs), which are usually inactive and heterochromatic, as a model system to study these questions. They performed a systematic survey of the expression of genes integrated into heterochromatic domains and of the intrinsic activity of heterochromatin genes outside of their native context. Using multiple high-throughput datasets, the authors were able to determine that both local sequence features and heterogeneity in the heterochromatic environment is important in determining heterochromatic gene expression.

What is this preprint about?

One of the main current questions in nuclear biology is what causes the repression of a gene when it is embedded in heterochromatin. In an initial experiment, the authors leverage the data generated from two parallel techniques in the K562 cell line. The first technique, Survey of Regulatory Elements (SuRE, Van Arensbergen et al. 2016), assays the intrinsic promoter activity of genomic elements by a plasmid-based massively parallel reporter assay (MPRA). In this assay, genomic fragments are cloned upstream of a sequence barcode and promoter activity is determined by measuring barcode expression. This allows the authors to determine the intrinsic activity of promoters usually embedded in LADs. In the second dataset generated from Global Nuclear Run On after enrichment for 5'-me7-meGTP-capped RNAs (GRO-cap, Core et al. 2014), a technique that captures and identifies nascent capped RNAs enriching for transcription start sites, the authors then assess the expression levels from these same promoters in their native heterochromatic context. They find that for sequences that show similar SuRE activity, the endogenous expression is, in general, much lower from promoter

and enhancer sequences embedded within LADs versus in inter-LADs (iLAD), suggesting that LADs do indeed form a largely repressive environment.

However, the authors also noted substantial heterogeneity in LAD-embedded promoter activity. Thus, they divided the promoters into three different categories: those with no exogenous (SuRE) and no endogenous (GRO-cap) expression are named inactive promoters; those with exogenous expression levels (high SuRE) but little or no endogenous (low GRO-cap) expression are named repressed promoters; and those with both high exogenous (high SuRE) and endogenous (high GRO-cap) expression are named ‘escaper promoters’ (see Figure below).

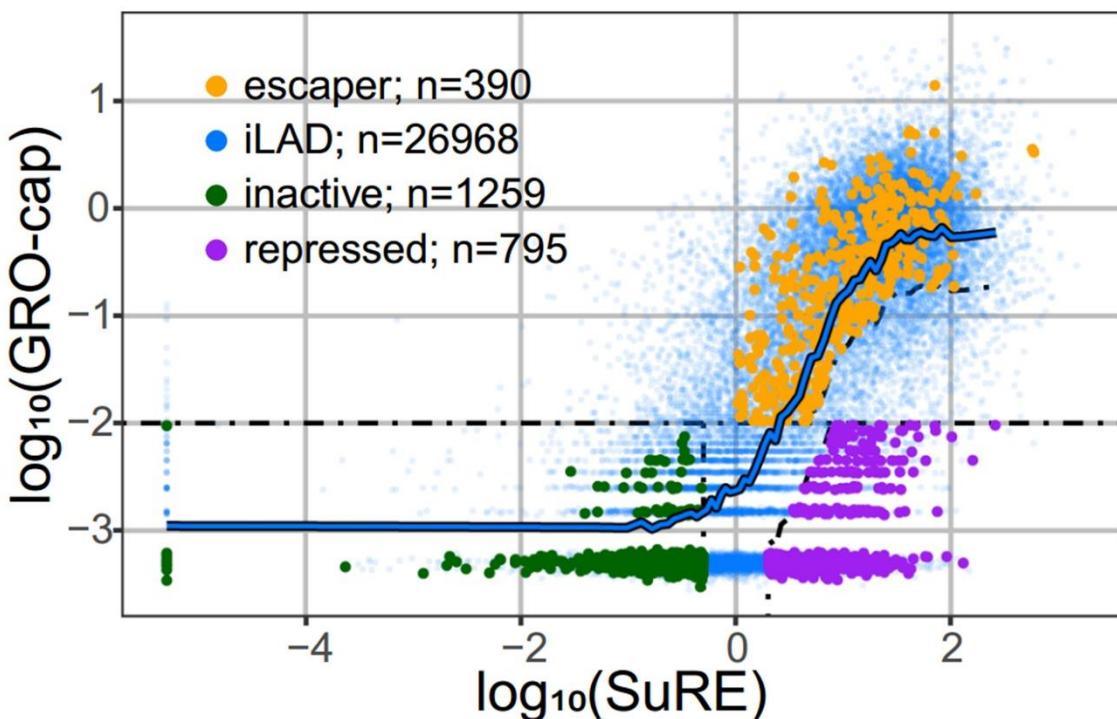


Figure: Endogenous promoter activity as measured by GRO-cap (y-axis) and intrinsic regulatory element activity (SuRE, x-axis) for inter-LAD regulatory elements (blue) and LAD regulatory elements (green, purple and orange). The different categories of promoter elements as described above are indicated in individual colors. This is figure 1C adapted from the preprint published under a CC-BY-NC 4.0 International license.

With these categories, they asked: “Are escapers less sensitive to the repressive LAD environment or are they embedded within a sub-LAD environment that is less repressive?” One way to answer this question is to look at how tightly genes are associated with the nuclear Lamina (NL). This can be measured by Lamin B1 – DamID, a technique that measures how close DNA is to the lamina. With this technique, they found that the escaper regions seem to be locally detached from the nuclear lamina, indicating that they may belong to a sub-LAD region that has a less repressive environment. Intriguingly, escaper promoters generally exhibit weaker endogenous activity than iLAD promoters with the same exogenous activity, suggesting that these promoters are not entirely able to overcome the repressive LAD environment.

In an orthogonal experiment to test the effects of heterochromatin environment vs local sequence features, the authors utilized another high-throughput assay known as Thousands of Reporters Integrated in Parallel (TRIP). The promoter sequences were cloned into a common reporter construct, which was integrated randomly into hundreds of genomic locations, including LADs. Then, the integration sites were mapped by inverse PCR followed by sequencing. Because each reporter has a barcode, the expression of a promoter at a given genomic locus can be analyzed independently, which allows the authors to study the effect of the genomic context on individual promoters. They chose 3 representative promoters each from the repressive and the escaper category and asked: if a repressive promoter is inserted into a less repressive environment, does it behave differently than when it is in its endogenous environment, and vice versa?

With this approach, the authors made several observations. First, when a repressed promoter is located in a LAD, its expression is lower than when integrated in an iLAD, confirming that LADs have a higher repressive potential. Second, this effect is less pronounced for escapers, which suggests that some intrinsic sequence features must determine their expression levels, at least to a certain extent. Third, some promoters are too strong for the LAD environment to be able to repress it, for instance, when the promoter of a housekeeping gene (PGK) was inserted into a LAD, it barely reduced its expression.

Thus, the authors asked, are escaper promoters intrinsically stronger than repressed promoters, and is this what determines their sensitivity to the LAD repressive environment? The authors determined that the expression from the repressive and escaper promoters in an episomal context (a barcoded reporter is overexpressed downstream of the promoter sequences and can be a measure for chromatin-context independent promoter strength) was very similar.

So, in the LADs, it is not promoter sequence that intrinsically determines whether a gene is repressed or escaping. But what does determine the variation in expression levels within LADs?

In order to answer this question, the authors used a statistical learning strategy to look at many epigenomic features at the integration sites in absence of the integrations themselves. They wanted to identify a set of features that is most likely to explain the reporter expression levels. Because they were specifically interested in the heterochromatic features of LADs, they only looked at integration sites within the LADs. With this, they learned that only half of the variance in expression levels can be explained by local chromatin environments for repressed promoters and even less (35%) for escaper promoters. This indicates that escaper promoters are less sensitive to the local chromatin environment than the repressed promoters.

But which features contribute to explaining the expression? Here they found that for both promoter categories, the most striking feature was close association with the nuclear lamina themselves (as defined by DamID-Lamin B1): the higher the association with LaminB1, the higher the NL contact frequency, the lower the reporter activity. On the opposite side, H2A.Z levels were positive predictors of reporter activities. H2A.Z is generally believed to destabilize nucleosomes, and thus increase TF and cofactor accessibility. So, although no one feature could be identified as defining the repressive or escaper status of a promoter, the authors do identify a small subset of features that contributes to LAD promoter activity.

The authors also extended the analysis of differential promoter sensitivity to other types of heterochromatin. For this, they looked at TRIP integration sites focusing on regions with high levels of Polycomb-induced H3K27me3. They found that escaper promoters were still less repressed than repressed promoters, but that the degree of repression is maintained and that it is generally 3 to 5 fold less in polycomb induced heterochromatin than the HC from the LADs.

What do we think?

We really like this paper! Not only does it contain very straightforward questions, it also has a good mix of high-throughput and modelling techniques to address them. This paper reveals an important point which, by our natural instinct to simplify and categorize things, is often forgotten: not always is an answer to a question in biology “one” or “the other” – more often than not a good mix of complex ideas is needed to explain a biological problem. For instance here, it isn’t -just- the packed heterochromatin that explains lower promoter activity, nor is it fully the sequence or the proximity to the nuclear periphery themselves that do. Since the variance isn’t fully explained by the model, are there other features that might explain the results? Are there any local sequence features around the insertion sites that can be identified? In addition, the observations relating to the housekeeping gene PGK promoter – which are really intriguing in itself – reveal important contrasts with less active iLAD and LAD promoters. They tell us something about the bigger picture of these experiments: each promoter, most probably depending on cell type, state, and context, may have its own identity, and may have evolved as such. Philosophically, this is really interesting – the cell has found ways to define how and how often genes are turned off and on in their native context in order to function well and cooperatively for the development of the organism as a whole.

From this work, though, some important questions arise: are the correlations found by the model consequential or causal? For example, is the histone variant H2A.Z incorporated because it needs to avoid that the gene is silenced? How do the lamin proteins really impact on chromatin architecture and compaction? Do they induce it directly, or rather are they just proximal to silent heterochromatin for other reasons? It is also interesting that H3K122ac proximity appears to have opposite effects on escaper vs repressed promoters. Is there any potential explanation for this, given how important it seems to be for escaper promoters? Besides the model, we had some other questions. What determines the distance of a given stretch of DNA to the lamina, and is this somehow evolutionarily fixed? While the numbers of promoters did not provide enough power to perform de novo motif analysis, did the authors try to look for enrichment of specific core promoter motifs, such as TATA or DPE motifs? I would also be curious to know what the contribution of LMNB1, promoter class, and the interaction is to the linear regression model in Figure 3, to understand what the relative strengths of each factor is to expression.

References

- Core et al. (2015) Nat Genet. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers.
Van Arenbergen et al. (2017) Nat Biotechnol. Genome-wide mapping of autonomous promoter activity in human cells.

11.No Atlas without a Map: spatially resolving single cell transcriptomes in tissues using barcoded ‘addresses’ on beads.

Slide-seq: A Scalable Technology for Measuring Genome-Wide Expression at High Spatial Resolution

Samuel G Rodrigues, Robert R Stickels, Aleksandrina Goeva, Carly A Martin, Evan Murray, Charles R Vanderburg, Joshua Welch, Linlin M Chen, Fei Chen, Evan Z Macosko
Preprint posted on February 28, 2019 <https://www.biorxiv.org/content/10.1101/563395v1>

And

High-density spatial transcriptomics arrays for *in situ* tissue profiling

Sanja Vickovic, Goekcen Eraslan, Johanna Klughammer, Linnea Stenbeck, Fredrik Salmen, Tarmo Aijo, Richard Bonneau, Ludvig Bergenstraahle, Joshua Gould, Mostafa Ronaghi, Jonas Frisen, Joakim Lundeberg, Aviv Regev, Patrik L Staahl

Preprint posted on February 28, 2019 <https://www.biorxiv.org/content/10.1101/563338v1>

Selected by Carmen Adriaens

Categories: bioinformatics, molecular biology

Introduction

Significant advances in single-cell technologies have made the study of gene expression at unprecedented resolution mainstream. However, an important challenge remains to simultaneously retrieve spatial and transcriptomic information from the individual cells within a tissue: until now, existing technologies could either be spatially precise and provide low-throughput transcriptomic data (eg. [1], [2]), or provide low spatial resolution but genome-wide expression data [3], [4].

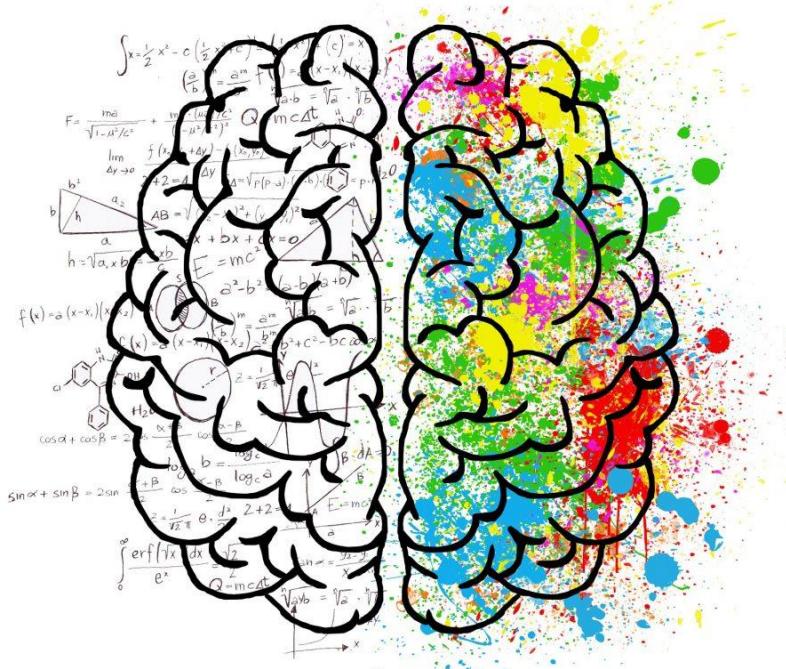
Now, two techniques, preprinted back-to-back on bioRxiv, tackle this issue by the ingenious development of barcoded beads from which spatial information can be recovered at high resolution. The tissue is deposited on these beads and specific locations are issued a spatial address prior to sequencing.

Spatial Transcriptomics 2.0: High Density Spatial Transcriptomics

In a first paper, Vickovic et al. build upon their previously developed technique, Spatial Transcriptomics [3], to increase its resolution from the range of ~100 µm (3-30 cells) to ~2 µm (less than 1 cell) before binning. The technique is based on the production of specific barcodes ligated to beads which, prior to lysis, are coupled to a spatial address by decoding and imaging of the tissue. Thus, once the tissue is lysed, reverse transcribed and sequenced, a transcript carrying a specific barcode can be reassigned to a particular location and histological structure. This technique was named High Density Spatial Transcriptomics (HDST).

One of the main novelties is the packing of the 2 μ m beads in a hexagonal beads-in-well format, which allows for their accurate deposition and compartmentalization of the biological materials under study. The design of this high resolution bead array together with a split-and-pool approach to generate the barcodes ensures that a large number of unique barcodes can be linked to specific locations in the tissue. This linking is done through deposition of a sequential slice of tissue onto a duplicate array, such that each layer and structure can be assigned to a specific location in the array.

As a proof of concept, the authors profiled the mouse main olfactory bulb, which consists of morphologically distinct layers. They asked, can HDST detect the different cell types and assign them to the correct positions? And, does HDST allow for the detection of distinct gene expression patterns within individual regions? The answer is: mostly yes. Although HDST was highly successful at retrieving a vast array of Unique Molecular Identifiers (UMIs) linked to a spatial address and the pooled genes detected by HDST correlated significantly with the genes detected in bulk RNA-seq from the same tissue, the individual data was relatively sparse. However, this could be solved by lightly (eg. 5X) binning the reads of adjacent beads to increase the information available for a specific location and thus, the different cell types in the main olfactory bulb could be reliably identified. By integrating a scRNA-seq dataset with the HDST data, they found that about 30% of the HDST profiles could be assigned to a single cell type, whereas after binning 5X, they obtained sufficient power to detect a single cell type in ~70% of the cases.



Copyright www.sfatulparintilor.ro copyright. All Rights Reserved.

Slide-seq

In a second paper, Rodrigues & Stickels et al. develop another technique to tackle the same problem. They too, employ barcoded beads, but instead of packing them into a hexagonal high-density array, they embed these 10 μ m beads in rubber onto a glass slide followed by determining their individual, random locations by SOLiD sequencing-by-ligation. The RNA was then captured upon lysis of the tissue placed on the beads, followed by slide removal before

library prep and sequencing. As a first check, the authors observed, as in the study above, high correlations between bulk sequencing and Slide-seq in different tissues. In addition, they too intercrossed their data with data from single cell RNA sequencing, and found that, similarly to HDST, the RNA from ~66% of the beads could be attributed to a single cell type, whereas another ~32% showed representation of two cell types.

Next, as a proof of concept, Slide-seq was performed on several tissues, and the authors showed that they could, both in a 2D assay (a single section) and in a series of sections reconstructing a 3D view, detect and locate the different cell types within complex organs such as the hippocampus in the brain. Finally, the authors applied Slide-seq to a traumatic brain injury model, and, at different time points after the tissue damage, described dynamic changes in the types, expression patterns and abundance of the cells present.

What I think about these works & open questions

I believe the scientific community has been eagerly waiting for the development of spatially informative techniques for single-cell transcriptomics. In fact, they are necessary to tackle cell identity and heterogeneity questions in a more integrated and comprehensive manner; a logical next step after the development and broad adoption of single-cell technologies. We need spatial information in order to understand, for instance, cell-to-cell communication during development; or differential responses of tumor cells to drugs, varying cytokines, changing oxygen levels, etc.

The techniques are also both scalable and broadly adaptable for different applications.

However, I do think some open questions remain and there is always room for improvements. For instance, in HDST, the bead size and thus hypothetical resolution is 2 μm , but it appears that at this resolution the data is too sparse for its purpose. I wonder if, in future versions of the technique, this hurdle can be overcome – after all, wouldn't it be incredible to obtain subcellular, spatially resolved expression data in single cells? In contrast, although the size of the beads in Slide-seq is significantly larger (~5x), the data obtained, albeit still relatively sparse as characteristic for most single cell level data [5], seems to be usable without the need for binning to obtain biologically meaningful information at this resolution. In the end, both may have a comparable “working resolution” of ~10-30 μm (1-3 cells).

So far as data sparsity is an issue, maybe a potential solution is to consistently combine high-throughput (dissociated) single-cell analysis with these techniques. Together, their power could be increased by integrating the rather sparse transcriptome information in the spatially informative techniques with more in-depth views on what defines the cellular identities in a standard single cell RNAseq protocol. Crossing-over the data could allow for assignment of cellular identity and location with a deeper understanding of the gene expression profiles of each cell. This could be done concomitantly, but also post-hoc based on sorting with markers found in the initial spatial protocols.

Furthermore, the authors have focused so far on the proof-of-concept to retrieve rather broadly known regions and relatively abundant cell types in the tissues. I am excited to see how, in the

future, these techniques will be used to sensitively detect extremely rare cell types (e.g. stem cells) within known regions of the tissues.

In conclusion, the biology and technologies described here shows the enormous potential of the techniques to make new, multidimensional discoveries. I am convinced they are the next step in the field of single cell research and, when perfected, they have a huge potential to vastly deepen our understanding of cells in their natural context. I really like that both papers address the issue in a totally different, yet conceptually similar way and I am curious to see the further development and fine-tuning of both!

References

- [1] S. Shah, E. Lubeck, W. Zhou, and L. Cai, "In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus.", *Neuron*, vol. 92, no. 2, pp. 342–357, Oct. 2016.
- [2] K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang, "RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells.", *Science*, vol. 348, no. 6233, p. aaa6090, Apr. 2015.
- [3] P. L. Stahl, F. Salmen, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, A. Borg, F. Ponten, P. I. Costea, P. Sahlen, J. Mulder, O. Bergmann, J. Lundeberg, and J. Frisen, "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics.", *Science*, vol. 353, no. 6294, pp. 78–82, Jul. 2016.
- [4] C. Medaglia, A. Giladi, L. Stoler-Barak, M. De Giovanni, T. M. Salame, A. Biram, E. David, H. Li, M. Iannaccone, Z. Shulman, and I. Amit, "Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq," *Science*, vol. 358, no. 6370, p. 1622 LP-1626, Dec. 2017.
- [5] P. Angerer, L. Simon, S. Tritschler, F. A. Wolf, D. Fischer, and F. J. Theis, "Single cells make big data: New challenges and opportunities in transcriptomics," *Curr. Opin. Syst. Biol.*, vol. 4, pp. 85–91, Aug. 2017.

Tags: brain science, new technologies, single cell, slide seq, spatial transcriptomics

Posted on: 20th March 2019