Doctoral theses at NTNU, 2021:234

Helle Samdal

Identification and characterization of long noncoding RNAs involved in the cell cycle

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Helle Samdal

Identification and characterization of long noncoding RNAs involved in the cell cycle

Thesis for the Degree of Philosophiae Doctor

Trondheim, June 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU

Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Computer Science

ISBN 978-82-326-6873-1 (printed ver.)
ISBN 978-82-326-6905-9 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2021:234

Printed by NTNU Grafisk senter

NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

FAKULTET FOR INFORMASJONSTEKNOLOGI OG ELEKTROTEKNIKK

Helle Samdal

Lange ikke-kodende RNA involvert i cellesyklus

For å produsere nye celler må kroppens celler dele seg til to identiske kopier i en prosess som kalles cellesyklus. Dette er en nøye kontrollert prosess for å sikre at eventuelle feil som oppstår under celledeling blir detektert og reparert. Hvis skaden er uopprettelig igangsetter cellen programmert celledød for å hindre at feil overføres til dattercellene. Feil i DNA som gir vekstfremmende egenskaper kan føre til kreft, som i dag er den sykdommen som fører til flest dødsfall i Norge med hele 11041 tilfeller i 2019.

Utviklingen av sekvenserings-teknologi på begynnelsen av 2000-tallet avslørte at mesteparten av det humane genomet blir transkribert og tusenvis av ikke-kodende RNA ble detektert. Lange ikke-kodende RNA er transkripter som er lenger enn 200 nukleotider og blir vanligvis ikke translatert til protein. I dag er det identifisert totalt 84485 ikke-kodende transkript, 48684 av disse er lange ikke-kodende RNA. De siste femten årene er det vist at mange av de lange ikke-kodende RNA har viktige funksjoner i cellene, der de er involvert i flere ulike biologiske prosesser inkludert cellesyklus, genregulering, utvikling, differensiering og X-kromosominaktivering. En stor andel av de lange ikke-kodende RNA er celletype-spesifikke, i tillegg til at de ofte er dysregulert i kreft, noe som tyder på at enkelte av disse kan være mulige biomarkører med prognostisk og terapeutisk potensiale.

Det er kjent at lange ikke-kodende RNA er involvert i regulering av cellesyklus, men det gjenstår fortsatt mye for å kunne identifisere alle lange ikke-kodende RNA som har generelle og patogene roller knyttet til cellesyklus, samt de molekylære mekanismene knyttet til funksjonen. I artikkel 1 benyttet vi en kombinasjon av ulike metoder for å identifisere lange ikke-kodende RNA involvert i cellesyklus. Vi gjorde cellesyklus-synkronisering av en human keratinocytt cellelinje (HaCaT), etterfulgt av RNA-sekvensering og ChIP-sekvensering. Totalt identifiserte vi 1009 gener som ble aktivt transkribert i ulike faser av cellesyklus, hvorav 59 var lange ikke-kodende RNA. For fire av disse, *SNHG26, EPB41L4A-AS1, RP11-132A1.4* og *ZFAS1,* undersøkte vi videre funksjon

og fant at alle fire påvirket vekst og fasefordeling i cellesyklus i flere ulike cellelinjer. Metoden vi benyttet i artikkel 1 viste seg å være svært nyttig for å identifisere cellesyklus-assosierte gener.

I artikkel 2 og 3 foretok vi en grundigere funksjonell evaluering av henoldsvis *SNHG26* og *EPB41L4A-AS1*, som var to av de lange ikke-kodende kandidatene vi identifiserte i artikkel 1.

Siden det er RNA som utøver funksjonen, vil kjennskap til den subcellulære lokaliseringen av *SNHG26* og *EPB41L4A-AS1* kunne gi viktig informasjon om mulige biologiske funksjoner. Vi brukte derfor RNA fluorescence in situ hybridization (FISH) for å bestemme den subcellulære lokaliseringen til begge de lange ikke-kodende kandidatene. Resultatene viste at *SNHG26* hovedsakelig er lokalisert i kjernen, med to til fem høy-intensitets punkter, i tillegg til lav-intensitets punkter i cytoplasma i tre ulike cellelinjer. Når vi analyserte mitotiske HaCaT celler, observerte vi at høy-intensitets punkter forsvant, mens det i stedet oppstod lav-intensitets punkter co-lokalisert med DNA. Vi kan derfor ikke utelukke at *SNHG26* kan ha funksjoner knyttet til for eksempel epigenetisk nedarving. *EPB41L4A-AS1* er lokalisert både i kjerne og cytoplasma i HaCaT celler, med varierende intensitet.

Videre brukte vi CRISPR inhibering og CRISPR aktivering til å henholdsvis nedregulere og øke transkripsjonen til *SNHG26* og *EPB41L4A-AS1* for å se om vi kunne reprodusere effekten på cellesyklus og vekst fra artikkel 1, samt undersøke underliggende mekanismer. I artikkel 2 viste vi at nedregulering av *SNHG26* påvirket både vekst og fasefordeling i cellesyklus på samme måte som i artikkel 1, mens oppregulering av *SNHG26* ga motsatt effekt. I tillegg viste vi at effekten *SNHG26* har på fasefordeling var stabil over tid. Vi viste også at reduksjon i *SNHG26* hadde ulik effekt i ulike celletyper. Dette tyder på at celletype-spesifikke gener og mutasjoner gir en annen fenotypisk effekt eller ulike funksjonelle mekanismer i ulike celletyper. I artikkel 3 viste vi at nedregulering av *EPB41L4A-AS1* påvirket fasefordeling i cellesyklus som i artikkel 1, og at oppregulering av *EPB41L4A-AS1* ga motsatt effekt. Nedregulering og oppregulering av *EPB41L4A-AS1* resulterte i ulik fasefordeling i HaCaT og lungekreft celler (A549), noe som tyder på at også *EPB41L4A-AS1* har celletype-spesifikke funksjoner.

Mange lange ikke-kodende RNA regulerer uttrykket av sine nabogener. For å undersøke om dette også gjaldt våre kandidater brukte vi CRISPR til å endre uttrykket av *SNHG26* og *EPB41L4A-AS1* etterfulgt av RT-qPCR for å måle om utvalgte nabogen ble påvirket. Både *SNHG26* og *EPB41L4A-AS1* påvirket uttrykket av sine nabogen, henholdsvis *TOMM7* og *EPB41L4A,* noe som tyder på en mulig *cis*-regulering. Videre gjorde vi RNA-sekvensering for å få en bedre forståelse

av de funksjonelle mekanismene til *SNHG26* og *EPB41L4A-AS1*. For *SNHG26* identifiserte vi differensielt uttrykte gener som var motsatt uttrykt i respons til CRISPR inhibering og CRISPR aktivering i HaCaT celler. Kjente cellesyklus-assosierte gener var differensielt uttrykt, og felles for flere av disse er at de er indirekte eller direkte regulert av MYC. *MYC* er et onkogen som kan forårsake kreft ved å påvirke cellesyklus og proliferasjon gjennom regulering av ulike gener. En analyse av de fem mest signifikante genene støttet at *SNHG26* er involvert i reguleringen av cellesyklus. Genuttrykks-analyser av CRISPR-mediert nedregulering av *EPB41L4A-AS1* i HaCaT og A549 celler identifiserte henholdsvis 1612 and 1371 differensielt uttrykte gener, og genontologi terminologier inkluderte vekstfremmende signalveier.

I denne avhandlingen benyttet vi en kombinasjon av metoder som er spesielt egnet til å identifisere cellesyklus-assosierte gener. Basert på resultatene foretok vi funksjonelle og mekanistiske analyser av to lange ikke-kodende RNA*,* som viser at både *SNHG26* og *EPB41L4A-AS1* har vekstfremmende egenskaper. Funnene våre støtter at de molekylære mekanismene er interessante for videre undersøkelser i forhold til fremtidig klinisk potensiale.

Kandidat: Helle Samdal
Institutt: Institutt for datateknologi og informatikk
Hoved veileder: Pål Sætrom
Biveileder: Siv Anita Hegre
Finansieringskilder: Kreftforeningen og Forskningsrådet

ACKNOWLEDGEMENTS

This work was carried out at the department of computer science at the Norwegian University of Science and Technology in the period from 2017 to 2021. I appreciate the financial support from the Norwegian Cancer Society and the Research Council of Norway.

First and foremost, I would like to thank my supervisor Professor Pål Sætrom for introducing me to the interesting field of noncoding RNAs and for including me in your research group. Your scientific input and guidance are highly appreciated, and your incredible knowledge is truly inspiring.

To my co-supervisor Siv Anita, I really appreciate your eye for details and for always finding time to help me. Your contributions to the papers and proofreading the thesis were invaluable. Thanks for all the laughs and support through this journey, you are truly a good friend.

I would like to credit my co-authors for a good collaboration. I am especially thankful to Nina-Beate Liabakk, who performed the FACS analysis for all three papers, and Per Arne Aas for sharing his inexhaustible source of knowledge regarding experimental procedures and general know how. Kristin, I really appreciate you volunteered for the night work hours at the cell lab and thanks for being such a great company for the Keystone conferences. Thanks to Caroline for giving me the opportunity to participate in interesting projects, it has been rewarding to collaborate with you.

I am grateful to my colleagues and friends for creating an inspiring and great work environment. Meeting the 11 o`clock lunch gang has been a daily high light. Konika, it has been really fun sharing an office with you. Hilde, I appreciate our stay in Italy and your optimistic state of mind. Lene and Caroline, thanks for proofreading the thesis.

A special thanks to my close friends for taking my mind off work. Especially my childhood friends; Veronica, Guro, Solveig, Hilde, Beate, and Bente, and my friends from the local handball trainer crew; Elin M, Elin Ø, Vibecke, and Anne Marit. I truly appreciate the time we spend together.

I am forever grateful to my wonderful family. To my parents for helping us with the kids and the house whenever needed, and for your endless love and support. And my brother; I appreciate the relaxing times spent together at Røros. To my mother-in law, Karin, for flying in from Narvik and helping Raymond with the kids whenever I was away for work and conferences.

Finally, I would like to thank my dear Raymond for always being there for me and making me happy, and especially for taking such wonderful care of our girls. To my sweet children, Ronja and Anna, for reminding me that there are much more important things in life than work. No matter what happens, I will always have my family to come home to.

Trondheim, February 2021

TABLE OF CONTENTS

LIST OF PAPERS

Paper 1

Joint changes in RNA, RNA polymerase II, and promoter activity through the cell cycle identify non-coding RNAs involved in proliferation

Siv Anita Hegre*, **Helle Samdal***, Antonin Klima, Endre B. Stovner, Kristin G. Nørsett, Nina-Beate Liabakk, Lene Christin Olsen, Konika Chawla, Per Arne Aas, and Pål Sætrom

*) The first two authors should be regarded as joint First Authors.

Submitted manuscript

Paper 2

The G2-phase enriched lncRNA SNHG26 is necessary for proper cell cycle progression and proliferation

Helle Samdal, Siv Anita Hegre, Konika Chawla, Nina-Beate Liabakk, Per Arne Aas, Bjørnar Sporsheim, Pål Sætrom

Submitted manuscript

Paper 3

The lncRNA *EPB41L4A-AS1* regulates gene expression and exerts cell type-dependent effects on cell cycle progression

Helle Samdal, Siv Anita Hegre, Konica Chawla, Nina-Beate Liabakk, Per Arne Aas, Bjørnar Sporsheim, and Pål Sætrom

Manuscript is available at bioRxiv (BIORXIV/2021/430566)

ADDITIONAL CONTRIBUTIONS DURING PHD

Basal level of autophagy and MAP1LC3B-II as potential biomarkers for DHA-induced cytotoxicity in colorectal cancer cells

Helle Samdal, Malin A Sandmoe, Lene C Olsen, EAH Jarallah, Therese S Høiem, Svanhild A Schønberg, Caroline H H Pettersen

FEBS J, 2018. 285(13): p. 2446-2467. DOI: 10.1111/febs.14488

Establishment of a patient derived xenograft model of colorectal cancer in CIEA NOG mice and exploring Smartfish liquid diet as a source of omega-3 fatty acids

Helle Samdal, Lene C Olsen, Knut S Grøn, Elin S Røyset, Therese S Høiem, Ingunn Nervik, Pål Sætrom, Arne Wibe, Svanhild A Schønberg and Caroline H H Pettersen

In review, February 2021

ABBREVATIONS

AGO2	Argonaute 2
ASO	Antisense oligonucleotide
CDK	Cyclin-dependent kinase
CDC	Cell division cycle
CHART	Capture Hybridization Analysis of RNA Targets
ChIRP	Chromatin Isolation by RNA purification
ChIP-seq	Chromatin immunoprecipitation followed by sequencing
CHK1	Checkpoint kinase 1
CKI	Cyclin-dependent kinase inhibitor
CRISPR	Clustered regularly interspaced short palindromic repeats interference
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
dCAS9	Dead CAS9
ENCODE	Encyclopedia of DNA Elements
eRNA	Enhancer RNA
FIRRE	Functional intergenic repeating RNA element
G0 phase	Zero gap phase, resting stage
G1 phase	First gap phase
G2 phase	Second gap phase
GAS5	Growth arrest-specific 5
GO	Gene ontology
gRNA	Guide RNA
GRO-seq	Global nuclear run-on sequencing
H3K27me3	Histone H3 lysine 27 trimethylation

H3K4me3	Histone H3 lysine 4 trimethylation
H3K36me3	Histone H3 lysine 36 trimethylation
HOXD	Homeobox D
HOTAIR	HOX antisense intergenic RNA
KEGG	Kyoto encyclopedia of genes and genomes
K	Keratin
LAST	LncRNA-Assisted Stabilization of Transcripts
lincRNA	Long intergenic noncoding
LNA	Locked nucleic acid
LOF	Loss-of-function
lncRNA	Long noncoding RNA
M phase	Mitosis phase
MALAT1	Metastasis Associated Lung Adenocarcinoma Transcript 1
MA-linc1	Mitosis Associated Long Intergenic NonCoding RNA 1
MCM2	Mini-chromosome maintenance protein 2
MDM2	Mouse double minute 2
miRNA	MicroRNA
ncRNA	Noncoding RNA
NF-YA	Nuclear Transcription Factor Y Subunit Alpha
PANDA	p21-associated ncRNA DNA damage -activated
P-bodies	Processing bodies
PCGEM1	Prostate-Specific Transcript
PCNA	Proliferating cell nuclear antigen
pncRNA	Promoter-associated noncoding RNA
PRC2	Polycomb repressive complex

piRNA	piwi-interacting RNA
PLK	Polo-like kinases
Pol II	RNA polymerase II
PURPL	p53 upregulated regulator of p53 levels
Rb	Retinoblastoma protein
RBP	RNA-binding proteins
RIP-Seq	RNA immunoprecipitation followed by sequencing
RISC	RNA-induced silencing complex
RNAi	RNA interference
RNA-FISH	Fluorescent in situ hybridization targeting ribonucleic acid molecules
RNA-seq	RNA sequencing
RNP	Ribonucleoprotein
RT-qPCR	Quantitative Real-Time PCR
S phase	Synthesis phase
SAC	Spindle assembly checkpoint
siRNA	Small interfering RNA
SNHG	Small nucleolar RNA host gene
subRNAseq	Subcellular RNA-seq
TLS	Translocated in liposarcoma
TOMM7	Translocase of Outer Mitochondrial Membrane 7
TUG1	Taurine-upregulated gene 1
WEE1	Wee1-like protein kinas

1 INTRODUCTION

1.1. Cell cycle

Cells need to undergo a cell cycle to grow and divide into two daughter cells. The cell cycle consists of four main phases: Gap 1 (G1), Synthesis (S), Gap 2 (G2), and Mitosis (M) (Figure 1). The interphase (G1, S, and G2) is where the cell spends most of the time, during which it grows, replicates its chromosomes and some of its organelles, and prepares to divide. After leaving interphase, the cell undergoes mitosis where it provides two copies of its genetic material to the resulting daughter cells in a process divided into prophase, prometaphase, metaphase, anaphase and telophase (reviewed in [1]). The chromosomes condense during prophase, and the spindle forms in prometaphase. During metaphase the spindle attaches to the chromosomes, followed by segregation in anaphase. Two nuclei are formed during telophase when the chromatids are sufficiently separated before cytokinesis produces two cells and cell division can be repeated (reviewed in [2]). The cell can also leave the cell cycle and enter a resting stage, Gap 0 (G0) (reviewed in [1]).

1.1.1. Cell cycle regulation

The cell cycle is highly regulated to ensure detection of genetic damage and prevention of uncontrolled cell division. The cell cycle progression is controlled by checkpoints in the different stages. These checkpoints detect whether a cell contains damaged DNA or lack any components necessary for normal replication and induce cell cycle arrest to keep the cell from replicating before the damage is repaired. If the damage is irreparable, the cell induces apoptosis to prevent the damage to be transferred to the daughter cells (reviewed in [3]).

The cyclins and the cyclin-dependent kinases (CDKs), a family of serine/threonine kinases, are the two key regulatory components of the cell cycle. The CDKs have catalytic activity and are activated by the binding of specific cyclins. Activated CDKs phosphorylate target proteins causing activation or inactivation followed by proper entry into the next phase of the cell cycle [4]. The levels of cyclins are highly regulated both by synthesis and by their destruction by ubiquitin-dependent proteolysis. The activity of CDKs can be inhibited by the binding of Cyclin kinase inhibitors (CKIs) or by inhibitory tyrosine phosphorylation [5] (reviewed in [1, 3]).

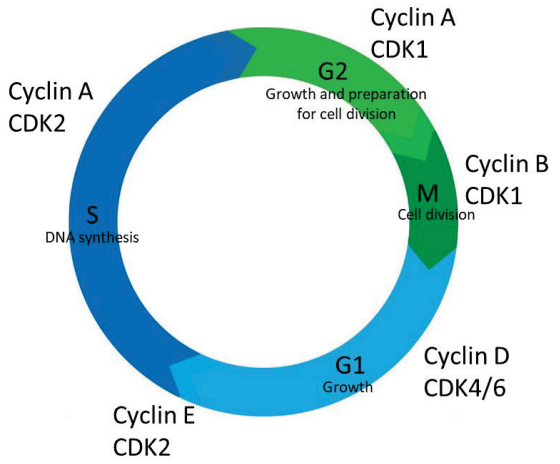


Figure 1: The key regulatory cyclins and CDKs in the cell cycle phases they regulate (reviewed in [1]).

There are several checkpoints affected by both external and internal factors. The main checkpoints include the restriction point, the G1 checkpoint, the G2/M checkpoint, and the mitosis-associated spindle assembly checkpoint (SAC). The restriction point is a key checkpoint regulating the G1 to S transition, where cells that lack all material necessary for replication are detected. If the cell is not allowed to proceed, it enters the resting state called the G0 phase. The majority of cells that pass the restriction point end up completing the cell cycle. DNA replication in the presence of DNA damage is delayed by the G1 checkpoint, which is mainly controlled by the p53/p21/mouse double minute 2 (MDM2) pathway. DNA integrity and correct replication are controlled for at the G2/M checkpoint, and detected errors are repaired. If the damage is irreparable, the cell undergoes apoptosis. The SAC controls spindle formation and correct attachment of the chromosomes to the mitotic spindle, and spindle defects are detected and corrected to ensure proper separation of sister chromatids (reviewed in [3, 6]).

Different cyclins are associated with distinct stages of the cell cycle while the level of CDK*s* is more constant. In early G1 phase the D-type cyclins (D1, D2, and D3) activate CDK4 and CDK6 which phosphorylate members of the Rb (retinoblastoma) protein family, resulting in the release of E2F transcription factors and subsequent transcription and activation of E- and A-type cyclins [7]. Cyclin E (cyclins E1 and E2) bind and activate CDK2 causing passage through the restriction point and progression to S phase of the cell cycle. For the cell to progress through S phase, CDK2 binds cyclin A. A-type cyclins bind CDK1 at the end of S phase, and both CDK2-cyclin A and CDK1-

cyclin A complexes are involved in the transition from S to G2 phase. The level of B-type cyclins increases during G2 while the A-type cyclins are degraded. CDK1 binds to B-type cyclins, and CDK1-cyclin B complexes are involved in both the regulation of G2 to M transition and the progression through mitosis (reviewed in [8, 9]).

There are two families of CKI: the inhibitors of CDK4 (INK4) and Kinase inhibitory protein (Kip) family. The Kip family includes p21, p27, and p57, and these can inhibit several CDKs. Meanwhile, the INK4 family consists of p15, p16, p18, and p19, which regulate CDK4 and CDK6 (reviewed in [10]). The cyclin-dependent kinase inhibitor p21 is a potent CKI and regulates the cell cycle progression from G1 to S phase through the binding and inhibition of the activity of cyclin-CDK2, -CDK1, and -CDK4/6 complexes (reviewed in [11]).

There are also other key proteins involved in the regulation of cell proliferation such as proliferation cell nuclear antigen (PCNA), marker of proliferation Ki-67, and mini-chromosome maintenance protein 2 (MCM2). PCNA is a DNA clamp that acts as a scaffold for proteins involved in DNA replication, recombination, and repair. It is necessary for replication and is an important regulator of the cell cycle through its interaction with cell cycle proteins. The expression of PCNA varies during the cell cycle and peaks in S phase where it interacts with CDK2-cyclin A. Abruption of this interaction by the competitive binding of p21 causes DNA replication arrest. Due to its fluctuating expression during cell cycle, and its absence in G0 phase, Ki-67 is a useful marker for determining the proliferating fraction of a cell population (reviewed in [12]). Cell cycle arrest is induced by the blocking of Ki-67, although little is known about its mechanistic function. MCM2 is part of the MCM protein family, which is necessary for initiation of replication and replication elongation (reviewed in [12]). The expression of MCM2 is elevated in proliferating cells, and it is essential for DNA replication during the G1-phase of the cell cycle [13]. Due to their role in proliferation, PCNA, Ki-67, and MCM2 are used as markers for proliferation with potential as prognostic markers in different cancers [13] (reviewed in [12]).

1.1.2. Cell cycle and cancer

Cancer is characterized by cells that divide uncontrollably with the ability to spread through the body and infiltrate healthy cells. Mutations can cause cancer by either accelerating the division rate or by inhibiting growth controlling mechanisms. The cancerous cells can develop into a tumor, and as they continue dividing over time, new growth promoting mutations may increase the cells' ability to proliferate and evade normal growth control (reviewed in [14, 15]). Cancer

initiation and progression is caused by several processes including aberrant intracellular signaling pathways and intercellular communication affecting the tumor microenvironment (reviewed in [14, 15]). In 2000 Hanahan and Weinberg proposed six acknowledged hallmarks of cancer which contribute to the multistep development of cancer; (1) sustained proliferative signaling, (2) resisting cell death, (3) evading growth suppressors, (4) replicative immortality, (5) induced angiogenesis, and (6) invasion and metastasis (reviewed in [14]). Genome instability and inflammation are underlying conditions that promote the development of the abovementioned hallmarks. Two additional hallmarks emerged in an updated paper from Hanahan and Weinberg in 2011: reprogramming of energy metabolism and evading immune destruction. In addition, the surrounding microenvironment contributes to the development and complexity of the tumor (reviewed in [15]).

Several of the components of the CDK4/6-RB pathway, which is necessary for the progression through G1 phase and transition to S phase, are frequently mutated in cancers. The cyclin D1 gene (*CCND1*) is often amplified in human cancers [16]. Mutations in *CCND1* can cause constitutive activation of the CDK4/6 kinases, which are considered the major onco-drivers among CDKs (reviewed in [17]). The Rb protein is a tumor suppressor that regulates the transition from G1 to S phase [18], where it binds and deactivates transcription factors from the E2F family and prevents the cell from entering S phase. Loss of Rb causes uncontrolled cell division and is frequently mutated in human cancers (reviewed in [19]).

The G1 checkpoint is mainly controlled by the p53/p21/Mdm2 pathway [20], and mutations in any of these components are frequently observed in cancers. p53 is a tumor suppressor and is activated in response to different cellular and external stresses. Mutations in the gene that encodes p53 (*TP53*) are detected in about 50% of all cancers (reviewed in [21]). The activity of p53 is negatively regulated by the E3 ubiquitin ligase MDM2, which targets p53 for proteasomal degradation. Inhibition of MDM2 stabilizes p53 and causes accumulation. Meanwhile, MDM2 is amplified in about 7% of human cancers, which increases the negative regulation of P53 and compromises the function of the checkpoint [22]. p53 regulates the expression of several cell cycle regulators, including p21, which is a known inhibitor of cyclin E. Cyclin E is necessary for the entry into mitosis (reviewed in [23, 24]).

At the G2/M checkpoint the DNA integrity and correct replication are controlled. CDK1 is activated by the binding of cyclin A2 and cyclin B, which is essential for the entry into mitosis.

CDK1 is required for tumorigenesis yet is rarely mutated in cancers. Checkpoint kinase 1 (CHK1) is activated in response to DNA damage, and initiates G2 arrest by causing inactivation of CDK2 and CDK1 through the phosphorylation of cell division cycle (CDC) phosphatases CDC25A, CDC25B, and CDC25C or Wee1-like protein kinase (WEE1). The CHK1-mediated cell cycle arrest is especially important to impair proliferation in cancer cells that lack a functional p53 (reviewed in [24]).

The Polo-like kinases (PLKs) are checkpoint kinases, and PLK1 is involved in the maturation of centrosomes by regulation of the centrosomal localization of Aurora A [25]. PLK1 is also an activator of cyclin B-CDK1 complexes, and is necessary for the entry into mitosis after DNA damage-induced G2 phase arrest [26]. Aurora A is crucial for normal centrosome and spindle formation and is also an activator of PLK1 and thus promotes mitotic entry in response to DNA damage-induced phase G2 arrest. Meanwhile, Aurora B controls cytokinesis and is involved in chromosome condensation and orientation (reviewed in [24]).

Cell cycle regulators like CDKs, checkpoint kinases (CHK1 and WEE1), cyclins, and PLK1 and Aurora kinases (Aurora A and Aurora B) are often overexpressed in cancers and represent attractive targets for cancer therapy. Selective inhibitors of CDK4 and CDK6 have benefited patients in clinical studies on head and neck squamous cell carcinomas, breast-, melanoma, and non-small-cell lung cancers. Also, inhibitors of PLK1 and compounds targeting Aurora A have shown encouraging results in clinical studies of some cancers (reviewed in [24]).

1.2. Long noncoding RNAs

Several recent studies suggests that several noncoding RNAs (ncRNAs) are involved in cell cycle regulation and tumorigenesis [27-31]. Analysis of the transcriptome using RNA sequencing in the early 2000s revealed tens of thousands of RNAs transcribed from the mammalian genome [32, 33]. By integrating data from several experiments and sequencing techniques, the Encyclopedia of DNA Elements (ENCODE) project found that only about 1.2% of the transcripts were protein coding [34]. A large portion of the transcripts had little, if any coding capacity. The noncoding RNA molecules that are longer than 200 nucleotides are known as long noncoding RNAs (lncRNAs), and these have little if any coding capacity. Similar to messenger RNAs (mRNAs), most lncRNAs are transcribed by RNA Polymerase II (Pol II), 5' capped and have a polyA tail in the 3' end (reviewed in [35]). Highly conserved lncRNAs are believed to be involved in common biological processes across different lineages such as fetal development, while less conserved

lncRNAs may contribute to interspecies and individual differences in function and phenotypes. Noncoding genes are generally less conserved across species than protein-coding genes [36], which has been used as an argument against biologic function. However, lack of nucleotide homology is not in itself proof for lack of function, as there are other important features for evaluating lncRNA conservation such as RNA 2D structures, transcript exon-intron structures, function, and expression from syntenic loci (reviewed in [37, 38]). There are a wide variety of biological processes reported for lncRNAs including genomic imprinting, X-inactivation, development, differentiation, and cell cycle regulation (reviewed in [39]). To date, there are a total of 60656 genes and 84485 transcripts registered in the ENCODE project (GENCODE v35). Of these, 17957 are annotated lncRNA genes and 19954 protein coding genes, giving rise to 48684 and 84485 different transcripts, respectively. Moreover, 7569 and 14767 genes are small noncoding genes and pseudogenes, respectively.

1.2.1. Classification

Noncoding RNAs can be divided into two large groups, short RNAs and lncRNAs, based on their size. Short RNAs consist of less than 200 nucleotides and include piwi-interacting RNAs (piRNAs), short-interfering RNAs (siRNAs), and microRNAs (miRNAs). With an average size of 22 nucleotides, miRNAs are post-transcriptional repressors through base-pairing with mRNAs with similar sequence. Other regulating small RNAs are the siRNAs, which repress gene expression by cleaving target mRNAs and are a useful tool to study gene function. The piRNAs form complexes through the interactions with members of the piwi protein family and are involved in post-transcriptional and epigenetic silencing of transposable elements (reviewed in [40]). There are no universal classification of lncRNAs, but several approaches have been proposed based on length, location compared to known annotated genes, transcript properties, regulatory elements, subcellular localization, and functions (reviewed in [41]). The broad classification is commonly based on the localization of lncRNAs with respect to protein-coding genes in the genome. Long intergenic noncoding RNAs (lincRNAs) represent more than half of human lncRNA transcripts and do not overlap any other annotated genes (reviewed in [42]). Meanwhile, intragenic lncRNAs overlap other protein-coding genes, and are further classified based on whether they overlap protein-coding genes in the sense- or antisense direction, in intergenic- or intronic areas, and whether they have a bidirectional transcription (Figure 2, reviewed in [41]).

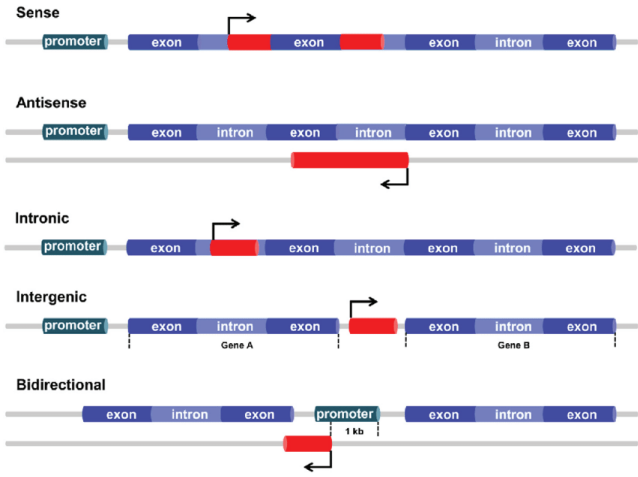


Figure 2: Classification of lncRNAs based on their localization in the genome. Reprinted from [43] with permission from KeAi Publishing.

1.2.2. Subcellular localization

RNA is the functional product of lncRNA genes and must be localized at the site of action, hence the subcellular localization of lncRNAs provides important insight regarding their biogenesis and biological function (Figure 3). Nuclear lncRNAs are enriched in the chromatin, nucleolus, or in sub-nuclear compartments, and are most likely involved in transcriptional regulation. An exclusively nuclear localization rules out biological functions connected to the cytoplasm including translation into small peptides and acting as microRNA sponges [44] (reviewed in [45]). In a review from Ling-Ling, known functional evaluated lncRNAs were classified into three broad categories based on their subcellular localization; (1) absolutely nuclear localized at their site of transcription (in *cis*), (2) mainly nuclear localized but needs to relocate from their site of transcription (in *trans*), and (3) those that localize in the cytoplasm (reviewed in [39]). LncRNAs that accumulate at their site of transcription in *cis* can regulate local transcription by recruiting chromatin modifiers or organizers, transcription factors, or scaffolding effector proteins to the promotor. *Cis*-acting lncRNAs can also modulate gene expression in *trans* through 3D organization of the chromosomes or by acting as decoys. Meanwhile, the other group of nuclear localized lncRNAs needs to relocate from the site of transcription to modulate the gene expression in *trans.* Cytoplasmic lncRNAs may affect gene expression by interfering with protein post-translational modifications, or by acting as decoys for miRNAs and proteins. In addition, proteomic analysis revealed that some lncRNAs encode functional micropeptides [46] (reviewed

in [39]). A study from Cabili *et.al*. used RNA FISH to determine the cellular localization of several lncRNAs in three different cell lines, where they observed a more varied localization pattern and cell-to-cell variability than proposed by Ling-Ling [44]. The majority of the examined lncRNAs were localized in both cytoplasm and nucleus (40%) and the second largest group was predominantly nuclear, with no foci (26%). The smallest group was localized in the cytoplasm (5.5%), while the remaining groups of nuclear localized lncRNAs were divided into lncRNAs with large nuclear foci (13%), or large nuclear foci with surrounding scattered single molecules (15.5%). The presence of bright nuclear foci may be connected with roles involving chromatin regulation, or that the lncRNA is localized in a sub-nuclear compartment [44].

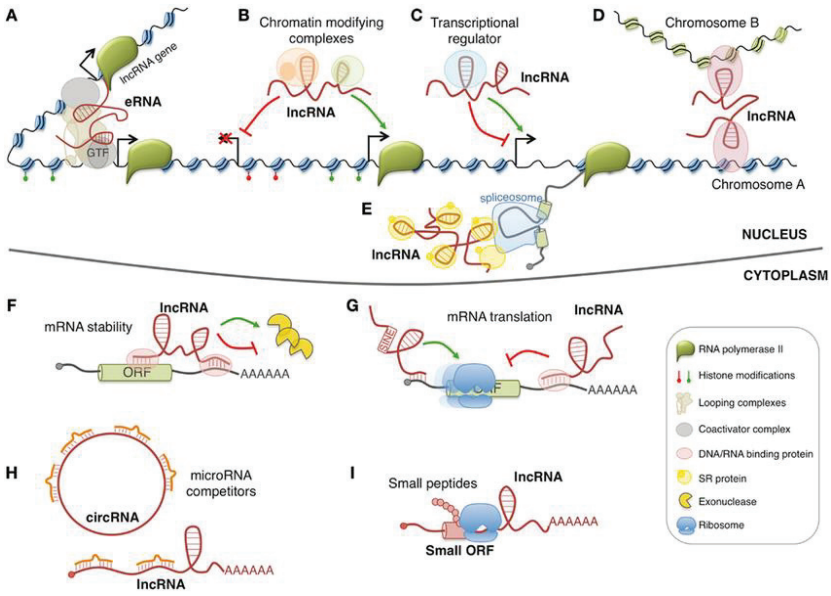


Figure 3: The variety of functions of lncRNAs localized in the nucleus and cytoplasm. LncRNAs localized in the nucleus can regulate transcription by acting as enhancer RNA (eRNA) (A), by recruiting chromatin modifying complexes (B), by regulating the activity of transcription factors (C), by acting on the spatial conformation of chromosomes (D), or by influencing pre-mRNA splicing (E). The lncRNAs located in the cytoplasm can regulate mRNA stability and availability (F), translation (G), or act as miRNA sponges (H). A few lncRNAs can be translated into small peptides (I). Reprinted from [47] with permission from Morlando, Ballarino and Fatica.

1.2.3. Tissue-specific expression

Compared to protein coding genes, the expression and function of lncRNAs are more tissue-, cell type-, and context-dependent, providing important clues regarding biological function. A study from Derrien *et.al*. analyzed RNA sequencing (RNA-seq) data from several different human tissues where 65% of protein coding genes were found in all tissues compared to only 11% of lncRNAs [48]. In another study, 78% of lincRNAs were defined as tissue-specific and 12% were expressed in all tissues. The brain and testis contain most tissue specific lncRNAs [49]. This high specificity of many lncRNAs implies biological relevance in different biological contexts, in both normal and pathological development. The dysregulation of these tissue-specific lncRNAs in disease may be used as biomarkers for diagnosis, disease progression, and as therapeutic targets. Meanwhile, ubiquitously expressed lncRNAs imply involvement in cellular maintenance and housekeeping functions necessary for cell survival. Moreover, the lncRNAs expressed in all tissues were higher expressed and characterized with several features including a more compact gene structure, a strict regulation at transcriptional, post-translational, and epigenetic levels, a strong regulation of enhancers, and they had a high conservation [49].

1.2.4. Gene regulation

With the ability to bind DNA, RNA, and proteins, lncRNAs can modulate gene expression at the transcriptional, post-transcriptional, translational, and epigenetic level. They can function as signals for transcription, decoys for transcription factors, scaffolds for the assembly of proteins and for the formation of functional ribonucleoproteins (RNPs), as well as guides for chromatin modifying components (reviewed in [50]).

At the transcriptional level, lncRNAs affect chromatin remodeling, the transcription of regulatory elements, or the binding of regulatory elements to promoters and enhancers. LncRNAs can directly modulate gene expression near their site of transcription in cis or distally in trans. Some lncRNAs are described as eRNAs, a class of regulating lncRNAs that facilitates the interaction between promoters and enhancers, or activates the transcription at the promoter (reviewed in [51]). A study from Ørom *et.al.* demonstrated through loss-of-function experiments that 7 of 12 lncRNAs affected the expression of their nearby genes [52]. An example of such regulation is Mitosis Associated Long Intergenic NonCoding RNA 1 (*MA-linc1*), a nuclear lncRNA that regulates the exit of M phase by repressing the transcription of the neighboring gene Purine Rich Element Binding Protein A *(PURA)* in *cis* [53]. While, Functional Intergenic Repeating RNA Element (*FIRRE*) is an example of a lncRNA that is localized at the X chromosome, but establishes five trans-

chromosomal contacts from its transcription site [54]. Several lncRNAs are able to affect chromatin structure by targeting chromatin modifying proteins such as the Polycomb repressive complex 2 (PRC2) (reviewed in [55]). HOX antisense intergenic RNA (*HOTAIR*) is a nuclear localized lincRNA commonly found dysregulated in several cancers. *HOTAIR* can act as a guide in trans where it relocates from its site of transcription and binds PRC2, causing genome-wide histone H3 lysine 27 methylation (H3K27me3) and subsequent repression of genes in the homeobox D (HOXD) cluster [56, 57]. Other studies have reported that lncRNAs can modulate transcription by affecting the localization and the activity of transcription factors. For instance *PURPL* lncRNA regulates the cellular pool of p53 by interacting with p53 stabilizing protein MYBBP1A in colorectal cancer cells [58] (reviewed in [55]).

Cytoplasmic localized lncRNAs affect the post-transcriptional regulation at several levels including pre-mRNA splicing, mRNA turnover, and -translation. Some lncRNAs have been reported to modulate splicing by interacting with splicing factors or by forming duplexes with pre-mRNA molecules (reviewed in [59]). Metastasis Associated Lung Adenocarcinoma Transcript 1 (*MALAT1*) is a nuclear localized lncRNA with oncogenic functions that affects splicing by modulating the distribution of splicing factors in nuclear speckle domains [60]. Some lncRNAs are termed competing endogenous RNAs, or miRNA sponges, and reduce the repressive effect of microRNAs by competing for the binding to miRNA targets. One example is *SNHG16* which functions as an oncogene in several cancers by sponging different miRNA, including in breast cancer where it induced cell migration by binding miR-98 and thereby inhibiting its regulation of E2F5 [61] (reviewed in [62]). Moreover, lncRNAs can give rise to miRNAs, and directly repress the expression of target transcripts (reviewed in [63]). Another regulatory mechanism is affecting mRNA turnover, exemplified by the LncRNA-Assisted Stabilization of Transcripts (*LAST*) which acts as a *CCND1* mRNA stabilizer by protecting against nuclease targeting [64].

RNA-binding proteins (RBPs) are involved in transcriptional regulation, as well as in mRNA transport, storage, turnover, and translation, and their activity can be modulated by lncRNAs (reviewed in [55]). A lncRNA that functions as a decoy negatively regulates other regulatory factors including RBPs, miRNAs, transcription factors, and modifying complexes by presenting alternative binding sites, thereby limiting their availability (reviewed in [65]). The lncRNA p21-associated ncRNA DNA damage-activated (*PANDA*) acts as a molecular decoy by interacting with transcription factor Nuclear Transcription Factor Y Subunit Alpha (NF-YA) causing impaired pro-apoptotic function in response to DNA damage [66]. Moreover, lncRNAs can modulate gene

expression by acting as molecular scaffolds by providing platforms where they assemble catalytic complexes and other regulatory factors to targeted genomic locations (reviewed in [65]).

1.3. LncRNAs in cell cycle and cancer

Mutations in the noncoding regions of the genome are involved in the pathology of several diseases, including cancer [67]. Due to their regulatory role, lncRNAs are frequently dysregulated in cancers, where they regulate the expression of several tumor suppressors and oncogenes at the transcriptional, post-transcriptional, and epigenetic level, affecting processes such as apoptosis, proliferation, angiogenesis, and activation of tissue invasion and metastasis of cancer cells. The functional mechanisms of some of these cancer-associated lncRNAs have been extensively studied, including growth arrest-specific transcript 5 (*GAS5), HOTAIR,* and *MALAT1* (Figure 4). The expression of *GAS5* is reduced in several cancers, where a low level is associated with poor prognosis. *GAS5* acts as a tumor suppressor by inducing apoptosis and growth arrest [68] (reviewed in [69]). *HOTAIR* is known as a regulator of chromatin state and can modulate the gene expression at several levels. Functional studies have reported that *HOTAIR* is involved in the pathogenesis of several cancers including gastric-, colorectal-, prostate-, liver-, breast-, cervical-, and ovarian cancer. The expression of *HOTAIR* is elevated in cancer compared to normal tissue, and it acts as an oncogene by positively regulating initiation, progression, and poor prognosis in cancer (reviewed in [70]). *MALAT1* is highly conserved and ubiquitously expressed at high levels across all tissues. Despite these characteristics, knock-out mouse studies reported that loss of *MALAT1* did not affect normal physiology. Meanwhile, *MALAT1* has been implicated in the pathology of several diseases including cancer, where its expression is associated with tumor progression and metastasis. Overexpression of *MALAT1* is detected in several cancers including, breast-, colorectal-, lung-, and liver cancer. Several modes of action have been reported for *MALAT1* including a direct or indirect regulation of transcription and alternative pre-mRNA splicing. Thus, the mechanisms of action and functional roles of *MALAT1* seems to vary depending on the cell type, localization in nuclear speckles and physiological context (reviewed in [71]).

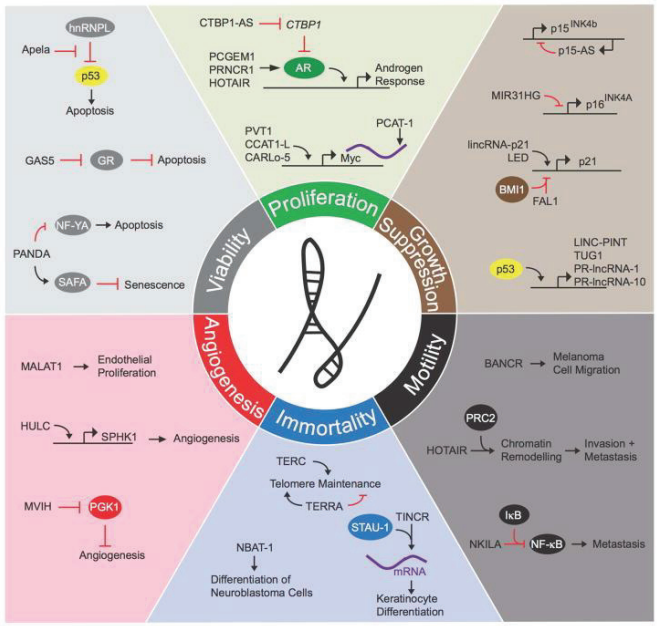


Figure 4: LncRNAs connected to the different hallmarks of cancer: proliferation, growth suppression, motility, immortality, angiogenesis, and viability. Reprinted from [72] with permission from Elsevier.

1.3.1. Cell cycle-associated lncRNAs

Several cell cycle lncRNAs have been identified, although only a few have been functionally characterized (Figure 5). The cell cycle is tightly regulated by cyclins, CDKs, CDKIs, Rb, p53, and E2F transcription factors. The expression level of these cell cycle-associated regulators is further controlled by lncRNAs at the epigenetic-, transcriptional-, and post-transcriptional level (reviewed in [73]). *CCND1* is negatively regulated in response to DNA damage through the recruitment and binding of a RBP, translocated in liposarcoma (TLS), in *cis* by the promoter-associated lncRNA (*pncRNA-D*) [74, 75]. *GAS5* is important for normal cell cycle progression in T-cells [76], and overexpression can cause cell cycle arrest in cancers. The expression level of *GAS5* is lower in stomach cancer compared to normal stomach, where downregulation of *GAS5* increase the turnover of Y-Box Binding 1 protein, thus decreasing the YBX1-mediated activation of p21, abolishing G1 phase cell cycle arrest [27]. Similarly, the expression of *GAS5* in pancreatic cancer is reduced compared to control, and downregulation increase the progression through G1 to S phase, probably through the regulation of CDK6 [77]. *MALAT1* is an oncogene involved in progression of more than one phase in the cell cycle, and its expression is important for the

control of several cell cycle-associated regulators. Depletion of *MALAT1* activates p53 and downstream pathways, causing G1 arrest. *MALAT1* is also involved in the G2 to M progression through the positive regulation of the oncogenic transcription factor Myb-related protein B [78] (reviewed in [71]). Another cell cycle-associated lncRNA is taurine upregulated gene 1 (*TUG1*), which has elevated expression in several cancers. A study on pancreatic cancer demonstrated that downregulation of *TUG1* resulted in G1/G0 phase arrest and reduced proliferation and apoptosis in two pancreatic cancerous cell lines [79].

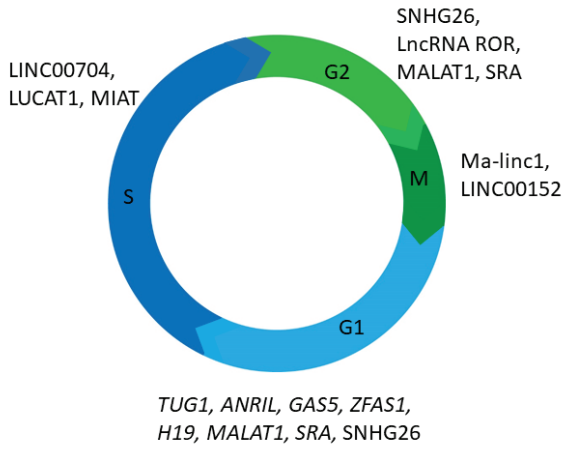


Figure 5: Cell cycle-associated lncRNAs and the phases they regulate [27, 29, 53, 80] (reviewed in [73]).

1.4. Perturbation of lncRNA expression

Due to the cell- and tissue specificity of many lncRNAs, they are considered potent diagnostic and therapeutic targets for cancer patients (reviewed in [72]). All types of nucleic acid therapy rely on delivery vectors for the entry into the cell's cytoplasm or nucleus, which represents a limitation regarding biosafety and immunogenicity. The development of different types of nanotechnology-based delivery systems may eventually overcome this problem (reviewed in [81]).

1.4.1. Perturbation strategies

The rapid development of RNA targeting strategies represents a huge potential in the modulation of lncRNA expression for cancer therapeutic proposes. Prior knowledge regarding the subcellular localization of the lncRNA of interest is necessary to select the most effective technology, mainly due to various access to protein factors necessary for degradation in the

different cellular compartments. SiRNAs are most effective for the downregulation of cytoplasmic localized lncRNAs, while synthetic oligonucleotides (antisense oligonucleotides; ASOs) are more effective for lncRNAs with nuclear localization. Meanwhile, both technologies are able to deplete both cytoplasmic and nuclear localized lncRNAs [82] (Reviewed in [72]). Clustered regularly interspaced short palindromic repeats interference (CRISPR) connected to a nuclease is an effective tool to knock out a target gene. Meanwhile, CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa) modulate gene expression at the transcriptional level, which probably represent a more flexible and transformable tool with less off-target effects than traditional nuclease CRISPR, siRNAs and ASOs [83].

1.4.2. Antisense oligonucleotides

How ASOs modulate gene expression or pre-mRNA splicing is determined by their chemistry, binding sequence, and target. ASOs form an RNA-DNA hybrid through Watson-Crick base pairing, followed by RNase H-mediated RNA degradation (reviewed in [84]). Locked nucleic acids (LNAs) are the most commonly used antisense method, which often includes a LNA modification within a RNA:DNA heteroduplex that induces RNase H-mediated degradation of the target RNA (reviewed in [85]). The LNA modification increases the binding strength to the target, which makes it possible to manufacture shorter ASOs with a higher potential for effective *in vivo* delivery and the ability to avoid detection of the innate immune system. Shorter sequences and higher binding potency may however increase the off-target effects due to increased binding ability towards unintended targets [82]. Currently, more than 100 ASOs are being tested in clinical trials (reviewed in [81]).

1.4.3. RNA interference

RNA interference (RNAi) is a commonly used natural way to regulate gene expression in most eukaryotic cells and includes the specific inhibition of translation or degradation of homologous transcripts by siRNAs and miRNAs [81]. Synthetic siRNAs can be produced from oligonucleotides and use the natural cellular pathway for RNAi-mediated gene repression to exert their function. The siRNA binds the Argonaute 2 (AGO2) protein in the RNA-Induced Silencing Complex (RISC) where a single strand is exposed to bind target RNA followed by degradation or reduced translation of target transcripts. Both ASOs and siRNAs are common tools to reduce gene expression, but the effectiveness depends on the abundance of necessary protein factors that are available in the different cellular compartments. There are contradictory reports regarding the effectiveness of siRNAs depending on AGO2-RISC mediated cleavage. Some reports that

AGO2-RISC is mainly cytoplasmic, and that siRNAs are non-functional or less effective for nuclear localized targets. Meanwhile, studies have detected AGO2-RISC in the nucleus, and successfully depleted nuclear RNAs by using siRNA [82, 86]. More than 30 different siRNAs have passed clinical trials, and the first therapeutic siRNA, Patisiran, was approved by the US Food and Drug Administration and the European Commission for treatment of polyneuropathy in hereditary transthyretin-mediated amyloidosis [87] (reviewed in [81, 88]).

1.4.4. CRISPR

CRISPR is a highly effective gene editing system which consists of a CRISPR-associated endonuclease (such as Cas9) and a short guide RNA (gRNA). The gRNA is a synthetic RNA that contains a user defined sequence which guides Cas9 to a genomic location. CRISPR/Cas9 creates a double strand break and is traditionally used in loss-of-function studies. Off-target effects like unintended cutting outside the target sequence and cytotoxicity caused by the double strand break, are potential drawbacks [89]. CRISPRi and CRISPRa can modulate gene expression at the transcriptional level and are more flexible and transformable tools with less off-target effects than the traditional endonuclease CRISPR/Cas9. The creation of a nuclease dead Cas9 (dCas9) results in the incapacity to cleave target DNA, meanwhile the ability to bind DNA based on the gRNA targeting sequence is retained. To effectively increase (CRISPRa) or decrease (CRISPRi) the expression of a downstream target gene, dCas9 can be fused with transcriptional activators or repressors and targeted to the promoter area of the gene of interest (Figure 6) [83, 90]. A study successfully used CRISPRi to modulate the expression of a specific well-characterized gene in mice, indicating that this method can be used for loss-of-function studies *in vivo* targeting specific cell types that are easily targeted for viral vectors [91].

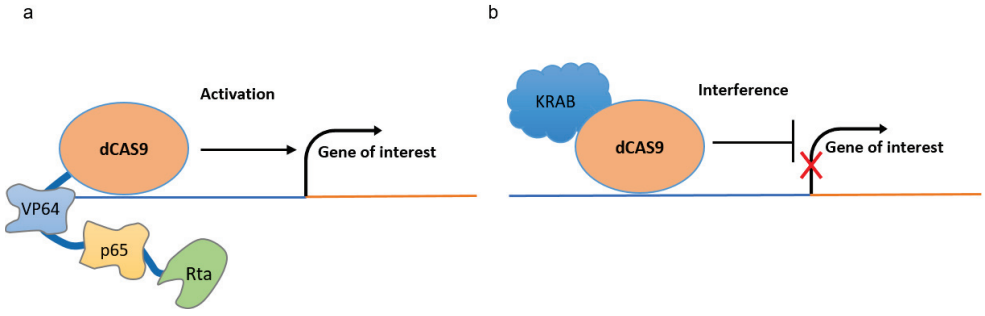


Figure 6: Nuclease dead CAS9 (dCAS9) fused with three activating domains VP64, p65, and Rta for CRISPR activation (a), and a repressive domain, KRAB, for CRISPR interference (b).

1.5. Functional characterization of lncRNAs

Differential expression analysis between patients and healthy controls have led to the identification of lncRNAs associated with a variety of diseases such as cancers [92], Alzheimer's disease [93], and Multiple Sclerosis [94]. To prove potential biological functions of these lncRNAs, it is common to include coding-noncoding co-expression analysis to identify whether any of the protein-coding RNAs have the same expression pattern as the lncRNAs of interest, suggesting a potential mRNA regulation target; guilt by association [95]. The dysregulated protein-coding genes with correlating expression patterns are then further annotated and analyzed for overrepresentation in pathways or gene ontology terms using public databases such as the Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG). RNAi, ASO or CRISPR- mediated loss-of-function experiments followed by gene expression profiling and functional assays are commonly used to evaluate the biological relevance and function of lncRNAs of interest. Furthermore, loss of function studies in mouse models are commonly used to study the biological relevance at an organismal level [96] (reviewed in [97]). For a final confirmation of biological function, lncRNA-knockout followed by RNA rescue experiments should be performed to investigate whether the knockout phenotype can be rescued upon ectopic expression of the lncRNA (reviewed in [38]).

1.5.1. Detecting lncRNAs

RNA-seq is among the most common methods for detecting lncRNAs. Since almost 40% of the human transcriptome is non-polyadenylated [98], a ribosomal depletion step may be advisable, allowing detection of both polyadenylated and non-polyadenylated transcripts. The remaining RNA is converted to cDNA using random hexamers and fragmented followed by high-throughput sequencing typically producing 30–400 base pair reads (reviewed in [99]). Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is another method used for the detection of lncRNAs, where genomic maps of the epigenome are generated by sequencing the sites that are occupied by histones with specific modifications. The presence of histone H3 lysine 4 trimethylation (H3K4me3) and H3K27me3 are known marks for gene activation and repression, respectively (reviewed in [100]). Moreover, genes transcribed by Pol II are recognized by the presence of H3K4me3 at the promoters and histone H3 lysine 36 trimethylation (H3K36me3) along the length of the transcribed region [101]. Guttman *et.al*. used ChIP-seq to search for H3K4me3 and H3K36me3 domains in four mouse cell types and identified approximately 1600 lincRNAs, where 95% of these displayed a high evolutionary conservation

[102]. While RNA-seq measures the steady-state level of RNA expression, global nuclear run-on sequencing (GRO-seq) is a very sensitive method used to identify newly synthesized transcripts selected based on the interaction with transcriptionally engaged polymerases [103]. GRO-seq is an especially useful tool to study ncRNAs with low expression or a short half-life [104].

1.5.2. Identifying cell-type specific expression

Studies have reported that the expression level of lncRNAs are much lower and tissue specific compared to protein coding genes [49, 105], and that tissue-specific lncRNAs have a lower relative expression than widely expressed lncRNAs [49]. The half-lives of different lncRNAs vary, although on average they are similar to the half-lives of mRNAs [106]. Meanwhile, the lower relative expression of tissue-specific lncRNAs may be a by-product of their expression in a few specific cells in a complex tissue. In fact, heterogeneous tissue types such as testis, brain, skin, and lungs have the most tissue-specific protein coding genes and lncRNAs [49]. Single-cell RNA-seq can be used to detect and analyze the expression of lncRNAs in individual cells, which provides a more accurate expression level compared to bulk tissue RNA-seq. High-throughput sequencing of single cells is a useful approach for identifying lncRNAs expressed in a small subset of cells and through different stages of development [107].

1.5.3. Identifying subcellular localization

The localization of RNA can be determined by absolute and relative approaches. The most common relative single-gene approach is Quantitative Real-Time PCR (RT-qPCR) of RNA extracts from subcellular compartments [108]. Fluorescent in situ hybridization targeting ribonucleic acid molecules (RNA FISH) is another single-cell approach that can be used to determine the subcellular localization of lncRNAs. Sequence-specific fluorescent probes are hybridized to the lncRNA of interest in fixated or live cells and visualized by imaging. RNA FISH may provide information about absolute molecule numbers per cell and of potential mechanisms of action based on the subcellular localization pattern of the lncRNA of interest [44] (reviewed in [109]). Subcellular RNA-seq (subRNAseq) is a high-throughput, quantitative method capable of whole-genome localization mapping of RNA extracted from fractionated cells. Subcellular localization data generated from subRNAseq of several cell lines are available at lncATLAS [110].

1.5.4. Identifying interaction partners

RNA immunoprecipitation (RIP)-seq is a combination of immunoprecipitation and RNA-seq used for the detection of RNAs associated with a RBP of interest [111]. For Chromatin Isolation by

RNA Purification (ChIRP), multiple biotin-tagged oligonucleotide probes are used to recognize and bind target RNA, followed by the isolation of the chromatin complexes containing the RNA of interest [112]. Capture Hybridization Analysis of RNA Targets (CHART) is a similar method, where the main difference is the use of an RNase-H mapping assay instead of oligonucleotide probes, used to identify potential hybridization sites [113]. Sequencing or mass spectrometry can be used to identify DNA, RNA, or proteins from these chromatin complexes [114] (reviewed in [109]).

2. AIMS OF THE STUDY

The main aim of the study was to identify and functionally characterize lncRNAs involved in the cell cycle.

The following subgoals were addressed in papers 1-3:

Paper 1

To identify cell cycle-associated lncRNAs by correlating the cyclic expression profile with changes in Pol II occupancy or promoter activity as measured by H3K4me3 in cell cycle synchronized cells.

Paper 2

To further evaluate the biological function and mechanism of the lncRNA candidate *SNHG26*, which demonstrated the highest correlation between expression profile and signal strength of H3K4me3 and Pol II from paper 1.

Paper 3

To further characterize the biological function of the lncRNA candidate *EPB41L4A-AS1*, which had a similar ChIP-seq characteristic as the known cell cycle-associated lncRNA *ZFAS1* from paper 1.

3. SUMMARY OF PAPERS

Paper 1

RNA-seq combined with ChIP-seq is a useful method for detecting transcriptional fluctuations by correlating gene expression with changes in histone modifications. To identify novel cell cycle-associated lncRNAs we performed cell cycle synchronization of HaCaT cells followed by total RNA-seq and ChIP-seq.

We detected 1803 genes with a cyclic expression profile, and 94 of these were lncRNAs. To further identify which of these cyclic genes displayed transcriptional fluctuations, we correlated their expression profile with the level of Pol II and H3K4me3 occupancy at the promoter region. We detected a total of 1009 genes whose expression and Pol II and/or H3K4me3 signals correlated and varied through the cell cycle. Of these, 59 were lncRNAs. Several of these cell cycle-associated genes were already known from the literature, and we validated both the expression profile and ChIP-seq signal of Pol II of *CCNB1, TOP2A,* and *CCNE2* by using RT-qPCR and ChIP-qPCR, respectively.

To further validate whether our proposed model is useful for the detection of cell cycle-associated genes, we selected four lncRNAs for siRNA-mediated loss-of-function studies. Knockdown of *AC005682.5, RP11-132A1.4*, *ZFAS1*, and *EPB41L4A-AS1* affected cell cycle phase distributions in HaCaT cells and reduced proliferation in multiple cell lines, which supports that our multi-omics method is well suited for identifying lncRNAs involved in the cell cycle.

Paper 2

SNHG26 (*AC005682.5*) displayed the highest correlation between the RNA-seq expression profile and Pol II and H3K4me3 signals from cell cycle synchronized HaCaT cells and was selected for further characterization by including other gene perturbation methods and functional assays. We used RNA FISH to determine the subcellular localization of *SNHG26* in HaCaT, A549, and LS411N cells. To modulate the expression of *SNHG26* we used siRNA, ASO, or CRISPR-inhibition and activation (CRISPRi and CRISPRa, respectively) to investigate how *SNHG26* affected mitochondrial stress, viability, proliferation, and gene expression.

SNHG26 was localized in the nucleus where it appears as bright foci, as well as low-intensity spots in the nucleus and cytoplasm, in HaCaT, A549, and LS411N cells. Downregulation of

SNHG26 affected the expression of its neighbor mRNA Translocase of Outer Mitochondrial Membrane 7 (*TOMM7),* mitochondrial stress, cell viability, and cell cycle phase distribution, and these effects were post-transcriptional. Moreover, the effect on cell cycle phase distribution was cell type-specific and stable over time. We were able to reverse these phenotypic effects by upregulation of *SNHG26* expression.

RNA-seq analysis revealed differentially expressed genes between CRISPRa compared to CRISPRi of *SNHG26* in HaCaT cells, suggesting trans-regulatory abilities. Several of these were known cell cycle-associated and MYC-responsive genes, including *MCM10, CDCA7, TP53, CDKN2A, CCNF,* and *CDK8*. Moreover, the enriched gene ontology terms for the five differentially expressed genes in the opposite direction between CRISPRi and CRISPRa included cell cycle G1/S transition, G2/M checkpoints, and DNA replication. As *SNHG26* is a direct MYC-target, we propose that it affects cell cycle progression and proliferation through the regulation of MYC-responsive genes.

Paper 3

EPB41L4A-AS1 was one of the cell cycle-associated candidates and was selected for further study based on a similar correlation between RNA-seq and ChIP-seq as for the already known cell cycle-associated lncRNA *ZFAS1*. For an in-depth characterization of *EPB41L4A-AS1* connected to cell cycle functions, we used CRISPRi and CRISPRa to modulate the expression of *EPB41L4A-AS1* at the transcriptional level followed by cell cycle and XTT viability assays. We used RNA FISH to determine the subcellular localization of *EPB41L4A-AS1*, and the effect on its antisense mRNA was evaluated using RT-qPCR. To determine whether it affected the gene expression we performed RNA-seq of CRISPRi modified HaCaT and A549 cells transduced with an *EPB41L4A-AS1*-specific gRNA or a non-target gRNA.

EPB41L4A-AS1 was localized both in nucleus and cytoplasm in HaCaT cells, where it appeared as bright foci as well as low-intensity spots. In line with previous results from paper 1 we demonstrated that *EPB41L4A-AS1* affected the cell cycle phase distribution in a similar way, with a decreased percentage of cells in the G1 phase and an increase in S and G2/M phases for CRISPRi-mediated repression of *EPB41L4A-AS1* in HaCaT cells. Moreover, we got the opposite effect for CRISPRa-mediated upregulation, which strengthens this effect on cell cycle phase distribution. We also included A549 cells, and interestingly CRISPRi and CRISPRa had the opposite effect on phase distribution compared to HaCaT cells, implying cell type-specific functions of *EPB41L4A-AS1.* Finally, *EPB41L4A-AS1* affected the expression of its antisense

mRNA *EPB41L4A,* in addition to the expression of several genes involved in proliferative signaling, suggesting *cis*- and *trans* regulatory abilities.

4. DISCUSSION

4.1. A useful method for identifying cell cycle-associated lncRNAs

In this study we propose a multi-omics method to detect transcriptional fluctuations for the identification of lncRNAs enriched for cell cycle functions; cell cycle synchronization followed by RNA-seq and ChIP-seq. The combination of markers for transcriptional activity used for ChIP-seq is important to detect transcriptional fluctuations. A study by Mikkelsen *et.al*. reported that promoters exclusively marked with H3K4me3 commonly regulate genes with general housekeeping functions, while bivalent promoters (containing both H3K4me3 and H3K27me3) generally have more complex expression patterns [101]. Moreover, actively transcribed genes by Pol II are recognized by the presence of H3K4me3 at the promoters and H3K36me3 along the length of the transcribed region. The level of gene expression was highly correlated to the level of H3K36me3, which was identified as strongly enriched in the regions of the actively transcribed genes in mouse embryonic stem cells, neural progenitor cells, and embryonic fibroblasts [101]. In another study they used the H3K4me3 and H3K36me3 domains to detect about 1600 highly conserved lincRNAs in four types of mouse cells [102]. In addition to ChIP-seq, there are also other possible strategies for detecting transcriptional fluctuations during cell cycle. One group used a combination of GRO-seq, RNA-seq, and ChIP-seq of histone 3 lysine 27 acetylation (H3K27ac) and histone 3 lysine 4 methylation (H3K4me2) to include regulatory sequences in synchronized human breast cancer cells [104].

In our study we included H3K4me3, H3K27me3, and Pol II to detect transcriptional fluctuations in the genes with a cyclic expression profile. As expected, we found a clear correlation between the expression level of genes and the signal strength of Pol II, H3K4me3, and H3K27me3. Moreover, the strength of both Pol II and H3K4me3 signals was higher in highly expressed cell cycle genes compared to lowly expressed genes, and the signals also differed between the different phases of the cell cycle. Although lowly expressed cell cycle-associated genes had a higher H3K27me3 signal than highly expressed genes, the signal strength did not differ between the cell cycle phases, neither was there any increased correlation between the H3K27me3 signal and expression level of the cyclic genes compared to the non-cyclic genes. Another study used H3K4me3 and H3K27me3 to study how these marks behave in bivalent genes during the cell cycle in pluripotent cells and upon differentiation in bivalent genes. In line with our results, they

reported that H3K4me3 occupancy varied during the cell cycle, while the signal of H3K27me3 remained stable [115].

For the ChIP-seq analysis, the transcriptional start site of the longest transcript variant was selected by default. As exemplified for *PCNA*, which has two transcripts with the shortest being transcribed in our dataset, only the longest transcript was included in the ChIP-seq analysis. This approach was a potential weakness due to the generation of false negatives.

The monoclonal antibody used for capturing Pol II during ChIP recognizes the YSPTSPS (Ser5) repeat in the B1 subunit of stalled Pol II. Pol II often enters a stalled state after the initiation of transcription [116], where only the Ser5 of the carboxyl-terminal end of the stalled Pol II complex is phosphorylated. An additional set of proteins is required for Pol II to escape this stalled state and initiate the transcriptional elongation (reviewed in [117]). For our study design, this means that we only detect genes occupied with the stalled state of Pol II, which is not necessarily a mark for active transcription [116] (reviewed in [117]). To identify the actively transcribed cyclic genes, we included H3K4me3 together with Pol II. Consequently, we detected genes with high correlation between expression profile and H3K4m3 and/or Pol II intensity, including genes with already known cell cycle functions, which we also validated by using RT-qPCR. Adding to the strength of this combination of methods for identifying cell cycle-associated genes, four of the lncRNA candidates were experimentally validated to affect both cell cycle and proliferation.

4.2. Gene perturbation methods: siRNA, ASO, and CRISPR

Both siRNAs and ASOs can downregulate non-target genes if the sequence complementarity is sufficient. Off-target effects occur when the gene perturbation method affects the expression of non-target genes, causing non-specific phenotypes [118, 119]. A study by Stojic *et.al*. quantified and compared off-target effects mediated by ASOs, siRNAs, and CRISPRi (dCas9–KRAB) by differential expression analysis of RNA-seq data from Hela cells compared to a variety of negative controls [85]. For siRNAs and ASOs, the sequences determined which non-target genes were affected. Meanwhile, the transfection of dCas9–KRAB to polyclonal Hela cells had very few off-target effects, while the sub-cultivation of single clones from this population affected the expression of non-target genes, resulting in a reproducible transcriptional signature [85]. The off-target effects observed by using siRNAs are often caused by miRNA-like effects of hybridization and silencing of transcripts with some sequence complementarity [118]. Hybridization of ASOs to non-target transcripts may reduce the translation of protein coding

genes or trigger degradation of transcripts. Meanwhile, sequence alignment algorithms can be applied to predict and prevent potential sequence-specific off-target interactions [119]. In addition to sequence-specific off-target effects that differ between different siRNAs and ASOs, there are also transfection-specific off-target effects that differ between the different gene perturbation methods. For instance, synthetic siRNAs compete with endogenous miRNAs for proteins, such as AGO2, and other factors necessary to exert their function. Consequently, transfecting siRNAs into the cells may affect the miRNA-mediated gene regulation and cause subsequent off-target effects [120].

To limit method- and sequence-specific off-target effects from affecting our conclusions, we included different gene-perturbation strategies for the functional characterization of the lncRNA candidates. For paper 1 we used two different siRNAs to target different sequences in the four lncRNA candidates for the proliferation assays. We also included four different cell lines to investigate whether effects were consistent across cell lines. Moreover, we reproduced an effect on cell cycle distribution from paper 1 for both *SNHG26* and *EPB41L4A-AS1* by using different gene perturbation methods and an additional cell line. For *SNHG26* we used ASO, four siRNAs with different target sequence, and CRISPRi to downregulate the expression, and for *EPB41L4A-AS1* we used two different sequence-specific siRNAs and gRNAs to target different sequences. Adding to the strength of our experimental results, we also used CRISPRa to upregulate the expression of both lncRNAs followed by functional assays.

4.3. Challenges connected to functional characterization of lncRNAs

As several lncRNAs have been identified during the last decade, there are challenges connected to the multitude of transcripts, incomplete classifications, and functional characterizations. Only a few lncRNAs are conserved at the sequence level, which seems contradictory with the functional importance that has been reported for many lncRNAs. Studies suggest that there is interspecies conservation of secondary structure, which may represent a better indication of biological function for lncRNAs. Meanwhile, the secondary structures of lncRNAs are difficult to predict. Another challenge is that lncRNAs can be alternatively spliced, which leads to the production of several isoforms from the same genetic loci, with possible different biological functions (reviewed in [121]).

The tissue- and cell type-specific expression of lncRNAs adds to the complexity of lncRNAs in terms of biological role and relevance, which might vary according to where they are expressed and the developmental stage. As lncRNAs can exhibit cell type and tissue-specific roles in different organs, the functional characterization should be done in the appropriate biological contexts. Cell type- and tissue-specific lncRNAs may have a low expression because they are expressed in a small proportion of cells within a tissue type and adequate sequencing depth is recommended. RNA-seq of bulk-tissue may dilute, or not even detect, lncRNAs with highly specialized functions that are expressed in a smaller proportion of specific cells. A study combining single-cell seq and RNA-seq from bulk tissue revealed that several cell type-specific lncRNAs that had a low expression level in whole tissue were highly expressed in individual cells, supporting that the low expression level detected for several lncRNAs are a consequential failure of them being expressed in only a few cells [107]. In line with this, ubiquitously expressed lncRNAs have a higher expression level than cell type-specific lncRNAs from the whole transcriptome [49].

We used HaCaT, A549, and LS411N cells as model systems to characterize the biological function of *SNHG26* and *EPB41L4A-AS1*. Due to the tissue- and cell type-specificity of the majority of lncRNAs, they may have different cellular functions depending on the investigated cell type. In paper 2 we demonstrated that *SNHG26* had different effects on the cell cycle phase distribution in HaCaT and A549 cells compared to LS411N cells. In paper 3 *EPB41L4A-AS1* affected the phase distribution differently in HaCaT compared to A549 cells. Thus, cell type-specific differences in addition to the epigenetic and mutational status of the cell lines probably influence the functional phenotype. This is especially relevant for characterization of functions connected to cell cycle progression and proliferation, as signaling pathways associated with these processes commonly harbor various mutations in immortalized and cancerous cell lines. For instance, *HOTAIR* is a much studied lncRNA commonly found dysregulated in several cancers where its expression correlates with prognosis and patient overall survival. Mechanistically, *HOTAIR* seems to be consistent, yet it affects different cellular processes depending upon cancer type [122]. Consequently, a functional characterization should involve several cell- and tissue types, in addition to *in vivo* experiments, which is both costly and time consuming. HaCaT cells have a mutated p53 (mutp53) with mutation in two alleles (H179Y and R282Q). These missense mutations in p53 mediate gain-of-function and enhance the proliferative properties of the cells. UVB related mutations of p53 are very common in squamous cell carcinomas. Even though

HaCaT cells harbor mutp53 with a UVB signature, they are non-transformed keratinocytes unable to generate tumors in mice [123, 124]. HaCaT cells have a very high expression level of mutp53, and knockdown of mutp53 supposedly reduce the cellular growth rate [123]. As p53 is a key regulator of several signaling pathways connected to cell cycle and proliferation, its mutational status may affect the phenotypic outcome of functional assays.

5. FUTURE PERSPECTIVES

Our work demonstrates that cell cycle synchronization followed by RNA-seq and ChIP-seq is a useful combination of methods for identifying cell cycle-associated lncRNAs. Furthermore, we validate the cell cycle and proliferative role of four of these identified cell cycle-associated lncRNAs. In addition, in paper 2 and paper 3 we perform an in-depth characterization of two of the cell cycle-associated lncRNA candidates.

We were able to validate the cell cycle and proliferative role of both *SNHG26* and *EPB41L4A-AS1*, although we still lack the full mechanistic insight into their mode of action. Interestingly, several of the cell cycle-associated lncRNAs detected in paper 1, including *SNHG26, RP11-132A1.4, EPB41L4A-AS1*, and *ZFAS1,* have previously been identified as direct MYC targets by others. As *MYC* participates in several gene regulatory networks and regulates at least 15% of the genome, it is an important decision maker for the outcome of multiple cellular processes. Especially c-MYC is involved in the tumorigenesis of several cancers and is activated in more than 50% of human cancers. Consequently, the expression level of *c-MYC* is highly controlled. As for the recent evidence that several lncRNAs are involved in the MYC-mediated regulation, both as regulators and effectors, this opens for investigations regarding their specific roles in these signaling networks. Due to the dysregulation of several lncRNAs in human cancers and the involvement in tumorigenesis, their mechanistic role connected to MYC-signaling should be of scientific interest, especially within the area of cancer research.

To be certain that the observed biological effect was caused by the lncRNA transcript, knockout followed by RNA-rescue is considered a gold standard and a final step to determine biological function (reviewed in [38]). For our experiments we used CRISPRa to upregulate the expression of the lncRNAs and observed an adverse effect. Meanwhile, to conclude that a phenotype was caused by a *trans*-acting lncRNA, the phenotype should be rescued by an ectopic expression from an independent transgene. This approach will separate between RNA-dependent effects and possible effects from the genetic loci where the RNA is transcribed.

HaCaT cells are considered a suitable model to investigate immunological and inflammatory responses of skin diseases like psoriasis [125]. Primary human keratinocytes are commonly used in studies of inflammatory skin diseases. The drawback with primary keratinocytes is that they require supplementation of growth factors to survive *in vitro*, they have a variable plating efficiency, a short lifetime in culture and their characteristics change with increasing number of

passages, which make them an insufficient model system for long-term monitoring. Like primary keratinocytes, HaCaT cells can differentiate in culture by regulating the Ca^{2+} concentration in the growth serum. The different stages of differentiation are recognized by the expression of unique keratin markers such as Keratin (K) 5 and K14 in the basal layer and K1 and K10 for the more differentiated cells. The ability of HaCaT cells to switch between different stages of differentiation, is a very useful characteristic for the investigation of inflammatory mechanisms and signaling pathways [125, 126]. HaCaT cells cultivated in low Ca^{2+} growth medium at about 80% confluence produce the same cytokines and express all the main surface markers and functional characteristics of isolated primary keratinocytes [125].

The key characteristics of psoriasis are enhanced proliferation of basal keratinocytes and the aberrant differentiation of spinous and granular keratinocytes [125]. The interplay between keratinocytes and immune cells causes an imbalance between differentiation and proliferation of keratinocytes in patients with psoriasis (reviewed in [127]). Recent studies investigating the expression profiles of lncRNAs in psoriatic patients compared to healthy controls found several differentially expressed lncRNAs, including both novel and annotated lncRNAs [128, 129]. In a lncRNA microarray study performed on psoriasis tissue and paired control tissue, a co-expression network analysis predicted several *cis*- and *trans*-regulated targets of lncRNAs, suggesting a dysregulation of mRNAs and lncRNAs as contributors to psoriasis. Despite the lack of functional studies, the tissue-specific expression and epigenetic profiles of lncRNAs in psoriatic patients suggest a pivotal role in the pathogenesis of psoriasis, and as potential diagnostic and therapeutic biomarkers [130]. There are no published studies elucidating the function of these lncRNAs, which represents a major knowledge gap in the psoriasis research field. By selecting cell cycle-associated lncRNA from paper 1 based on skin tissue-specificity and a differential expression in psoriasis patients compared to healthy controls, lncRNA's role in disease progression, as tools for disease stratification, and as therapeutic targets in psoriasis may be elucidated.

6. REFERENCES

1. Vermeulen, K., D.R. Van Bockstaele, and Z.N. Berneman, *The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer.* Cell Prolif, 2003. **36**(3): p. 131-49.
2. McIntosh, J.R., *Mitosis.* Cold Spring Harb Perspect Biol, 2016. **8**(9).
3. Barnum, K.J. and M.J. O'Connell, *Cell cycle regulation by checkpoints.* Methods Mol Biol, 2014. **1170**: p. 29-40.
4. Morgan, D.O., *Principles of CDK regulation.* Nature, 1995. **374**(6518): p. 131-4.
5. Harper, J.W., et al., *Inhibition of cyclin-dependent kinases by p21.* Mol Biol Cell, 1995. **6**(4): p. 387-400.
6. Bower, J.J., et al., *Patterns of cell cycle checkpoint deregulation associated with intrinsic molecular subtypes of human breast cancer cells.* NPJ Breast Cancer, 2017. **3**: p. 9.
7. Lundberg, A.S. and R.A. Weinberg, *Functional inactivation of the retinoblastoma protein requires sequential modification by at least two distinct cyclin-cdk complexes.* Mol Cell Biol, 1998. **18**(2): p. 753-61.
8. Satyanarayana, A. and P. Kaldis, *Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms.* Oncogene, 2009. **28**(33): p. 2925-39.
9. Malumbres, M. and M. Barbacid, *Mammalian cyclin-dependent kinases.* Trends Biochem Sci, 2005. **30**(11): p. 630-41.
10. Law, M.E., et al., *Cyclin-Dependent Kinase Inhibitors as Anticancer Therapeutics.* Mol Pharmacol, 2015. **88**(5): p. 846-52.
11. Gartel, A.L. and S.K. Radhakrishnan, *Lost in transcription: p21 repression, mechanisms, and consequences.* Cancer Res, 2005. **65**(10): p. 3980-5.
12. Jurikova, M., et al., *Ki67, PCNA, and MCM proteins: Markers of proliferation in the diagnosis of breast cancer.* Acta Histochem, 2016. **118**(5): p. 544-52.
13. Guzinska-Ustymowicz, K., et al., *Correlation between proliferation markers: PCNA, Ki-67, MCM-2 and antiapoptotic protein Bcl-2 in colorectal cancer.* Anticancer Res, 2009. **29**(8): p. 3049-52.
14. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.

15. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
16. Beroukhim, R., et al., *The landscape of somatic copy-number alteration across human cancers.* Nature, 2010. **463**(7283): p. 899-905.
17. Choi, Y.J. and L. Anders, *Signaling through cyclin D-dependent kinases.* Oncogene, 2014. **33**(15): p. 1890-903.
18. Goodrich, D.W., et al., *The retinoblastoma gene product regulates progression through the G1 phase of the cell cycle.* Cell, 1991. **67**(2): p. 293-302.
19. Burkhart, D.L. and J. Sage, *Cellular mechanisms of tumour suppression by the retinoblastoma gene.* Nat Rev Cancer, 2008. **8**(9): p. 671-82.
20. Agarwal, M.L., et al., *p53 controls both the G2/M and the G1 cell cycle checkpoints and mediates reversible growth arrest in human fibroblasts.* Proc Natl Acad Sci U S A, 1995. **92**(18): p. 8493-7.
21. Toledo, F. and G.M. Wahl, *Regulating the p53 pathway: in vitro hypotheses, in vivo veritas.* Nat Rev Cancer, 2006. **6**(12): p. 909-23.
22. Momand, J., et al., *The MDM2 gene amplification database.* Nucleic Acids Res, 1998. **26**(15): p. 3453-9.
23. Nag, S., et al., *The MDM2-p53 pathway revisited.* J Biomed Res, 2013. **27**(4): p. 254-71.
24. Otto, T. and P. Sicinski, *Cell cycle proteins as promising targets in cancer therapy.* Nat Rev Cancer, 2017. **17**(2): p. 93-115.
25. De Luca, M., P. Lavia, and G. Guarguaglini, *A functional interplay between Aurora-A, Plk1 and TPX2 at spindle poles: Plk1 controls centrosomal localization of Aurora-A and TPX2 spindle association.* Cell Cycle, 2006. **5**(3): p. 296-303.
26. van Vugt, M.A., A. Bras, and R.H. Medema, *Polo-like kinase-1 controls recovery from a G2 DNA damage-induced arrest in mammalian cells.* Mol Cell, 2004. **15**(5): p. 799-811.
27. Liu, Y., et al., *lncRNA GAS5 enhances G1 cell cycle arrest via binding to YBX1 to regulate p21 expression in stomach cancer.* Sci Rep, 2015. **5**: p. 10159.
28. Wang, C., et al., *Long noncoding RNA EMS connects c-Myc to cell cycle control and tumorigenesis.* Proc Natl Acad Sci U S A, 2019. **116**(29): p. 14620-14629.
29. Yildirim, O., et al., *S-phase Enriched Non-coding RNAs Regulate Gene Expression and Cell Cycle Progression.* Cell Rep, 2020. **31**(6): p. 107629.

30. Zhang, Y., et al., *Genome-wide study of ER-regulated lncRNAs shows AP000439.3 may function as a key regulator of cell cycle in breast cancer.* Oncol Rep, 2017. **38**(5): p. 3227-3237.
31. Gasri-Plotnitsky, L., et al., *A novel lncRNA, GASL1, inhibits cell proliferation and restricts E2F1 activity.* Oncotarget, 2017. **8**(14): p. 23775-23786.
32. Kapranov, P., et al., *Large-scale transcriptional activity in chromosomes 21 and 22.* Science, 2002. **296**(5569): p. 916-9.
33. Carninci, P., et al., *The transcriptional landscape of the mammalian genome.* Science, 2005. **309**(5740): p. 1559-63.
34. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
35. Marchese, F.P., I. Raimondi, and M. Huarte, *The multidimensional mechanisms of long noncoding RNA function.* Genome Biol, 2017. **18**(1): p. 206.
36. Necsulea, A., et al., *The evolution of lncRNA repertoires and expression patterns in tetrapods.* Nature, 2014. **505**(7485): p. 635-40.
37. Diederichs, S., *The four dimensions of noncoding RNA conservation.* Trends Genet, 2014. **30**(4): p. 121-3.
38. Gao, F., et al., *Reverse-genetics studies of lncRNAs-what we have learnt and paths forward.* Genome Biol, 2020. **21**(1): p. 93.
39. Chen, L.L., *Linking Long Noncoding RNA Localization and Function.* Trends Biochem Sci, 2016. **41**(9): p. 761-772.
40. Chu, C.Y. and T.M. Rana, *Small RNAs: regulators and guardians of the genome.* J Cell Physiol, 2007. **213**(2): p. 412-9.
41. Jarroux, J., A. Morillon, and M. Pinskaya, *History, Discovery, and Classification of lncRNAs.* Adv Exp Med Biol, 2017. **1008**: p. 1-46.
42. Ransohoff, J.D., Y. Wei, and P.A. Khavari, *The functions and unique features of long intergenic non-coding RNA.* Nat Rev Mol Cell Biol, 2018. **19**(3): p. 143-157.
43. Hermans-Beijnsberger, S., M. van Bilsen, and B. Schroen, *Long non-coding RNAs in the failing heart and vasculature.* Noncoding RNA Res, 2018. **3**(3): p. 118-130.
44. Cabili, M.N., et al., *Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution.* Genome Biology, 2015. **16**(1): p. 20.

45. Singh, D.K. and K.V. Prasanth, *Functional insights into the role of nuclear-retained long noncoding RNAs in gene expression control in mammalian cells.* Chromosome Res, 2013. **21**(6-7): p. 695-711.
46. Anderson, D.M., et al., *A micropeptide encoded by a putative long noncoding RNA regulates muscle performance.* Cell, 2015. **160**(4): p. 595-606.
47. Morlando, M., M. Ballarino, and A. Fatica, *Long Non-Coding RNAs: New Players in Hematopoiesis and Leukemia.* Front Med (Lausanne), 2015. **2**: p. 23.
48. Derrien, T., et al., *The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.* Genome Res, 2012. **22**(9): p. 1775-89.
49. Jiang, C., et al., *Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs.* Oncotarget, 2016. **7**(6): p. 7120-33.
50. Wang, K.C. and H.Y. Chang, *Molecular mechanisms of long noncoding RNAs.* Mol Cell, 2011. **43**(6): p. 904-14.
51. Kim, T.K., M. Hemberg, and J.M. Gray, *Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers.* Cold Spring Harb Perspect Biol, 2015. **7**(1): p. a018622.
52. Orom, U.A., et al., *Long noncoding RNAs with enhancer-like function in human cells.* Cell, 2010. **143**(1): p. 46-58.
53. Bida, O., et al., *A novel mitosis-associated lncRNA, MA-linc1, is required for cell cycle progression and sensitizes cancer cells to Paclitaxel.* Oncotarget, 2015. **6**(29): p. 27880-90.
54. Hacisuleyman, E., et al., *Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre.* Nat Struct Mol Biol, 2014. **21**(2): p. 198-206.
55. Sun, Q., Q. Hao, and K.V. Prasanth, *Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression.* Trends Genet, 2018. **34**(2): p. 142-157.
56. Gupta, R.A., et al., *Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis.* Nature, 2010. **464**(7291): p. 1071-6.
57. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.* Cell, 2007. **129**(7): p. 1311-23.
58. Li, X.L., et al., *Long Noncoding RNA PURPL Suppresses Basal p53 Levels and Promotes Tumorigenicity in Colorectal Cancer.* Cell Rep, 2017. **20**(10): p. 2408-2423.
59. Romero-Barrios, N., et al., *Splicing regulation by long noncoding RNAs.* Nucleic Acids Res, 2018. **46**(5): p. 2169-2184.

60. Tripathi, V., et al., *The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation.* Mol Cell, 2010. **39**(6): p. 925-38.
61. Cai, C., et al., *SNHG16 contributes to breast cancer cell migration by competitively binding miR-98 with E2F5.* Biochem Biophys Res Commun, 2017. **485**(2): p. 272-278.
62. Zhang, X., et al., *Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels.* Int J Mol Sci, 2019. **20**(22).
63. He, R.Z., D.X. Luo, and Y.Y. Mo, *Emerging roles of lncRNAs in the post-transcriptional regulation in cancer.* Genes Dis, 2019. **6**(1): p. 6-15.
64. Cao, L., et al., *LAST, a c-Myc-inducible long noncoding RNA, cooperates with CNBP to promote CCND1 mRNA stability in human cells.* Elife, 2017. **6**.
65. Balas, M.M. and A.M. Johnson, *Exploring the mechanisms behind long noncoding RNAs and cancer.* Noncoding RNA Res, 2018. **3**(3): p. 108-117.
66. Hung, T., et al., *Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters.* Nat Genet, 2011. **43**(7): p. 621-9.
67. Maurano, M.T., et al., *Systematic localization of common disease-associated variation in regulatory DNA.* Science, 2012. **337**(6099): p. 1190-5.
68. Mourtada-Maarabouni, M., et al., *GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer.* Oncogene, 2009. **28**(2): p. 195-208.
69. Ji, J., et al., *The role of long non-coding RNA GAS5 in cancers.* Cancer Manag Res, 2019. **11**: p. 2729-2737.
70. Tang, Q. and S.S. Hann, *HOTAIR: An Oncogenic Long Non-Coding RNA in Human Cancer.* Cell Physiol Biochem, 2018. **47**(3): p. 893-913.
71. Arun, G., D. Aggarwal, and D.L. Spector, *MALAT1 Long Non-Coding RNA: Functional Implications.* Noncoding RNA, 2020. **6**(2).
72. Schmitt, A.M. and H.Y. Chang, *Long Noncoding RNAs in Cancer Pathways.* Cancer Cell, 2016. **29**(4): p. 452-463.
73. Kitagawa, M., et al., *Cell cycle regulation by long non-coding RNAs.* Cell Mol Life Sci, 2013. **70**(24): p. 4785-94.
74. Wang, X., et al., *Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription.* Nature, 2008. **454**(7200): p. 126-30.

75. Cui, W., et al., *Arginine methylation of translocated in liposarcoma (TLS) inhibits its binding to long noncoding RNA, abrogating TLS-mediated repression of CBP/p300 activity.* J Biol Chem, 2018. **293**(28): p. 10937-10948.
76. Mourtada-Maarabouni, M., et al., *Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5).* J Cell Sci, 2008. **121**(Pt 7): p. 939-46.
77. Lu, X., et al., *Downregulation of gas5 increases pancreatic cancer cell proliferation by regulating CDK6.* Cell Tissue Res, 2013. **354**(3): p. 891-6.
78. Tripathi, V., et al., *Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB.* PLoS Genet, 2013. **9**(3): p. e1003368.
79. Hui, B., et al., *Overexpressed long noncoding RNA TUG1 affects the cell cycle, proliferation, and apoptosis of pancreatic cancer partly through suppressing RND3 and MT2A.* Onco Targets Ther, 2019. **12**: p. 1043-1057.
80. Notzold, L., et al., *The long non-coding RNA LINC00152 is essential for cell cycle progression through mitosis in HeLa cells.* Sci Rep, 2017. **7**(1): p. 2265.
81. Weng, Y., et al., *Improved Nucleic Acid Therapy with Advanced Nanoscale Biotechnology.* Mol Ther Nucleic Acids, 2020. **19**: p. 581-601.
82. Lennox, K.A. and M.A. Behlke, *Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides.* Nucleic Acids Res, 2016. **44**(2): p. 863-77.
83. Gilbert, L.A., et al., *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* Cell, 2013. **154**(2): p. 442-51.
84. Rinaldi, C. and M.J.A. Wood, *Antisense oligonucleotides: the next frontier for treatment of neurological disorders.* Nat Rev Neurol, 2018. **14**(1): p. 9-21.
85. Stojic, L., et al., *Specificity of RNAi, LNA and CRISPRi as loss-of-function methods in transcriptional analysis.* Nucleic Acids Res, 2018. **46**(12): p. 5950-5966.
86. Robb, G.B., et al., *Specific and potent RNAi in the nucleus of human cells.* Nat Struct Mol Biol, 2005. **12**(2): p. 133-7.
87. Adams, D., et al., *Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis.* N Engl J Med, 2018. **379**(1): p. 11-21.
88. Hoy, S.M., *Patisiran: First Global Approval.* Drugs, 2018. **78**(15): p. 1625-1631.

89. Gilbert, L.A., et al., *Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation.* Cell, 2014. **159**(3): p. 647-61.
90. Chavez, A., et al., *Highly efficient Cas9-mediated transcriptional programming.* Nat Methods, 2015. **12**(4): p. 326-8.
91. MacLeod, R.S., et al., *Effective CRISPR interference of an endogenous gene via a single transgene in mice.* Sci Rep, 2019. **9**(1): p. 17312.
92. Zhang, X., et al., *Comprehensive analysis of differentially expressed lncRNAs as diagnostic and prognostic markers for colorectal cancer.* Exp Ther Med, 2019. **18**(6): p. 4481-4489.
93. Shi, Y., et al., *Transcriptomic Analyses for Identification and Prioritization of Genes Associated With Alzheimer's Disease in Humans.* Front Bioeng Biotechnol, 2020. **8**: p. 31.
94. Santoro, M., et al., *Expression Profile of Long Non-Coding RNAs in Serum of Patients with Multiple Sclerosis.* J Mol Neurosci, 2016. **59**(1): p. 18-23.
95. Han, Y., et al., *Analysis of Differentially Expressed lncRNAs and mRNAs for the Identification of Hypoxia-Regulated Angiogenic Genes in Colorectal Cancer by RNA-Seq.* Med Sci Monit, 2019. **25**: p. 2009-2015.
96. Guay, C., et al., *Loss-of-function of the long non-coding RNA A830019P07Rik in mice does not affect insulin expression and secretion.* Sci Rep, 2020. **10**(1): p. 6413.
97. Goff, L.A. and J.L. Rinn, *Linking RNA biology to lncRNAs.* Genome Res, 2015. **25**(10): p. 1456-65.
98. Cheng, J., et al., *Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution.* Science, 2005. **308**(5725): p. 1149-54.
99. Wang, H.V. and J.A. Chekanova, *An Overview of Methodologies in Studying lncRNAs in the High-Throughput Era: When Acronyms ATTACK!* Methods Mol Biol, 2019. **1933**: p. 1-30.
100. Hyun, K., et al., *Writing, erasing and reading histone lysine methylations.* Exp Mol Med, 2017. **49**(4): p. e324.
101. Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-60.
102. Guttman, M., et al., *Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.* Nature, 2009. **458**(7235): p. 223-7.

103. Core, L.J., J.J. Waterfall, and J.T. Lis, *Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.* Science, 2008. **322**(5909): p. 1845-8.

104. Sun, M., et al., *Discovery, Annotation, and Functional Analysis of Long Noncoding RNAs Controlling Cell-Cycle Gene Expression and Proliferation in Breast Cancer Cells.* Mol Cell, 2015. **59**(4): p. 698-711.

105. Cabili, M.N., et al., *Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses.* Genes Dev, 2011. **25**(18): p. 1915-27.

106. Clark, M.B., et al., *Genome-wide analysis of long noncoding RNA stability.* Genome Res, 2012. **22**(5): p. 885-98.

107. Liu, S.J., et al., *Single-cell analysis of long non-coding RNAs in the developing human neocortex.* Genome Biol, 2016. **17**: p. 67.

108. Wang, Y., W. Zhu, and D.E. Levy, *Nuclear and cytoplasmic mRNA quantification by SYBR green based real-time RT-PCR.* Methods, 2006. **39**(4): p. 356-62.

109. Fang, Y. and M.J. Fullwood, *Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer.* Genomics Proteomics Bioinformatics, 2016. **14**(1): p. 42-54.

110. Mas-Ponte, D., et al., *LncATLAS database for subcellular localization of long noncoding RNAs.* RNA, 2017. **23**(7): p. 1080-1087.

111. Zhao, J., et al., *Genome-wide identification of polycomb-associated RNAs by RIP-seq.* Mol Cell, 2010. **40**(6): p. 939-53.

112. Chu, C., et al., *Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions.* Mol Cell, 2011. **44**(4): p. 667-78.

113. Simon, M.D., et al., *The genomic binding sites of a noncoding RNA.* Proc Natl Acad Sci U S A, 2011. **108**(51): p. 20497-502.

114. Li, W., et al., *Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation.* Nature, 2013. **498**(7455): p. 516-20.

115. Grandy, R.A., et al., *Genome-Wide Studies Reveal that H3K4me3 Modification in Bivalent Genes Is Dynamically Regulated during the Pluripotent Cell Cycle and Stabilized upon Differentiation.* Mol Cell Biol, 2016. **36**(4): p. 615-27.

116. Guenther, M.G., et al., *A chromatin landmark and transcription initiation at most promoters in human cells.* Cell, 2007. **130**(1): p. 77-88.

117. Wu, J.Q. and M. Snyder, *RNA polymerase II stalling: loading at the start prepares genes for a sprint.* Genome Biol, 2008. **9**(5): p. 220.
118. Jackson, A.L., et al., *Expression profiling reveals off-target gene regulation by RNAi.* Nat Biotechnol, 2003. **21**(6): p. 635-7.
119. Kamola, P.J., et al., *In silico and in vitro evaluation of exonic and intronic off-target effects form a critical element of therapeutic ASO gapmer optimization.* Nucleic Acids Res, 2015. **43**(18): p. 8638-50.
120. Lam, J.K., et al., *siRNA Versus miRNA as Therapeutics for Gene Silencing.* Mol Ther Nucleic Acids, 2015. **4**: p. e252.
121. Cao, H., C. Wahlestedt, and P. Kapranov, *Strategies to Annotate and Characterize Long Noncoding RNAs: Advantages and Pitfalls.* Trends Genet, 2018. **34**(9): p. 704-721.
122. Zhang, J., et al., *Long non-coding RNA HOTAIR in carcinogenesis and metastasis.* Acta Biochim Biophys Sin (Shanghai), 2014. **46**(1): p. 1-5.
123. Martynova, E., et al., *Gain-of-function p53 mutants have widespread genomic locations partially overlapping with p63.* Oncotarget, 2012. **3**(2): p. 132-43.
124. Cordani, N., et al., *Mutant p53 subverts p63 control over KLF4 expression in keratinocytes.* Oncogene, 2011. **30**(8): p. 922-32.
125. Colombo, I., et al., *HaCaT Cells as a Reliable In Vitro Differentiation Model to Dissect the Inflammatory/Repair Response of Human Keratinocytes.* Mediators Inflamm, 2017. **2017**: p. 7435621.
126. Deyrieux, A.F. and V.G. Wilson, *In vitro culture conditions to study keratinocyte differentiation using the HaCaT cell line.* Cytotechnology, 2007. **54**(2): p. 77-83.
127. Woo, Y.R., D.H. Cho, and H.J. Park, *Molecular Mechanisms and Management of a Cutaneous Inflammatory Disorder: Psoriasis.* Int J Mol Sci, 2017. **18**(12).
128. Yan, J., et al., *Long noncoding RNA expression profile and functional analysis in psoriasis.* Mol Med Rep, 2019. **19**(5): p. 3421-3430.
129. Gupta, R., et al., *Landscape of Long Noncoding RNAs in Psoriatic and Healthy Skin.* J Invest Dermatol, 2016. **136**(3): p. 603-609.
130. Tsoi, L.C., et al., *Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin.* Genome Biol, 2015. **16**: p. 24.

Paper I

Joint changes in RNA, RNA polymerase II, and promoter activity through the cell cycle identify non-coding RNAs involved in proliferation

Siv Anita Hegre[1,*], Helle Samdal[2,*], Antonin Klima[2], Endre B. Stovner[2,3], Kristin G. Nørsett[2,**], Nina Beate Liabakk[1], Lene Christin Olsen[1,4,5], Konika Chawla[1,4], Per Arne Aas[1], and Pål Sætrom[1,2,3,4,‡]

(1) Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology (NTNU), NO-7491 Trondheim, Norway

(2) Department of Computer Science, Norwegian University of Science and Technology (NTNU), NO-7491 Trondheim, Norway

(3) K.G. Jebsen Center for Genetic Epidemiology, Norwegian University of Science and Technology (NTNU), 7491, Trondheim, Norway

(4) Bioinformatics core facility - BioCore, Norwegian University of Science and Technology (NTNU), 7491, Trondheim, Norway

(5) The Central Norway Regional Health Authority, St. Olavs Hospital HF, Norway

‡) To whom correspondence should be addressed. Tel: +47 72573011; Email: pal.satrom@ntnu.no.

*) The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

**) Present address: Department of Biomedical Laboratory Science, Norwegian University of Science and Technology (NTNU), NO-7491 Trondheim, Norway

Paper II

The G2-phase enriched lncRNA *SNHG26* is necessary for proper cell cycle progression and proliferation

Helle Samdal[2], Siv Anita Hegre[1], Konika Chawla[1,4], Nina Beate Liabakk[1], Per Arne Aas[1], Bjørnar Sporsheim[1,5,6], and Pål Sætrom[1, 2,3,4, ‡]

1. Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, 7491 Trondheim, Norway
2. Department of Computer Science, Norwegian University of Science and Technology, 7491 Trondheim, Norway
3. K.G. Jebsen Center for Genetic Epidemiology, Norwegian University of Science and Technology, 7491 Trondheim, Norway
4. Bioinformatics core facility - BioCore, Norwegian University of Science and Technology, 7491 Trondheim, Norway
5. Cellular & Molecular Imaging Core Facility – CMIC, Norwegian University of Science and Technology, 7491 Trondheim, Norway
6. Central Administration, St Olavs Hospital, The University Hospital in Trondheim

‡) To whom correspondence should be addressed. Tel: +47 73594796; Email: pal.satrom@ntnu.no.

Paper III

The lncRNA *EPB41L4A-AS1* regulates gene expression and exerts cell type-dependent effects on cell cycle progression

Helle Samdal[2], Siv Anita Hegre[1], Konika Chawla[1,4], Nina-Beate Liabakk[1], Per Arne Aas[1], Bjørnar Sporsheim[1,5,6], and Pål Sætrom[1, 2,3,4, ‡]

1. Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, 7491 Trondheim, Norway
2. Department of Computer Science, Norwegian University of Science and Technology, 7491 Trondheim, Norway
3. K.G. Jebsen Center for Genetic Epidemiology, Norwegian University of Science and Technology, 7491 Trondheim, Norway
4. Bioinformatics core facility - BioCore, Norwegian University of Science and Technology, 7491 Trondheim, Norway
5. Cellular & Molecular Imaging Core Facility – CMIC, Norwegian University of Science and Technology, 7491 Trondheim, Norway
6. Central Administration, St Olavs Hospital, The University Hospital in Trondheim

‡) To whom correspondence should be addressed. Tel: +47 73594796; Email: pal.satrom@ntnu.no.

Abstract

The long non-coding RNA (lncRNA) *EPB41L4A-AS1* is aberrantly expressed in various cancers and has been reported to be involved in metabolic reprogramming and as a repressor of the Warburg effect. Although the biological relevance of *EPB41L4A-AS1* is evident, its functional role seems to vary depending on cell type and state of disease. By combining RNA sequencing and ChIP sequencing of cell cycle synchronized HaCaT cells we previously identified *EPB41L4A-AS1* to be one of 59 lncRNAs with potential cell cycle functions. Here, we demonstrate that *EPB41L4A-AS1* exists as bright foci and regulates gene expression in both *cis* and *trans*. Specifically, we find that *EPB41L4A-AS1* positively regulates its sense overlapping gene *EPB41L4A* and influences expression of hundreds of other genes, including genes involved in cell proliferation. Finally, we show that *EPB41L4A-AS1* affects cell cycle phase distribution, though these effects vary between cell types.

Introduction

According to the Encyclopedia of DNA elements (ENCODE) project (GENCODE v35) there are 17957 annotated lncRNA genes giving rise to 46977 different transcripts. LncRNAs are characterized as transcripts more than 200 nucleotides long with none or little coding potential. Most of them are transcribed by RNA polymerase II (Pol II), poly-adenylated, 5'- capped, and spliced. Functional evaluation of lncRNAs demonstrate that they are involved in gene regulation at the transcriptional, epigenetic, and translational level, where they can interact with DNA, RNA, and proteins to exert their function. LncRNAs can broadly be classified based on whether they regulate gene expression around their transcription site in *cis*, or if they leave their site of transcription and exert their function elsewhere in *trans* [1]. Several lncRNAs are involved in the development and pathogenesis of diseases, including different types of cancer, degenerative, autoimmune, and cardiovascular diseases. Although their biological relevance is undisputable, only a small part of the annotated lncRNAs have been functionally evaluated.

As RNA is the functional product of lncRNAs, their subcellular localization provides information about their potential biological role. Nuclear localized lncRNAs have possible roles connected to chromatin organization, transcriptional regulation, RNA processing, or nuclear domains. Meanwhile, cytoplasmic lncRNAs can modulate gene regulation by acting as decoys for microRNAs (miRNAs) and proteins, or affect signal transduction by interfering with protein post-translational modifications [2]. LncRNAs are more cell type-specific than mRNAs, and both their subcellular localization and biological role can vary according to cell/tissue and stage of development [2]. This characteristic increases the complexity and adds to the challenge in determining the biological role of lncRNAs. For instance, Metastasis Associated Lung Adenocarcinoma Transcript 1 (*MALAT1*) is one of the best-studied lncRNAs, and it is implicated in several diseases such as diabetes, atherosclerosis, and cancer. *MALAT1* is involved in different cellular processes including differentiation, tumor progression, hypoxia, inflammation, and stress. There are accumulating studies demonstrating an oncogenic role of *MALAT1* [3], and interestingly, a few studies reporting a tumor-suppressor role [4-7].

EPB41L4A-AS1 is identified as dysregulated in several gene expression studies of different diseases, suggesting a possible role as biomarker and a therapeutic target [8, 9]. There are a few functional studies investigating the biological role and molecular mechanisms of *EPB41L4A-AS1* [9-12] (Hegre and Samdal *et.al*. unpublished). Two functional studies report that *EPB41L4A-AS1* is involved in the regulation of glycolysis and glutaminolysis, and acts as a suppressor of the Warburg effect in the placental villus and in cancer cells [11, 12]. Another functional study demonstrates that mir-146a

inhibits the expression of *EPB41L4A-AS1*, and overexpression of *EPB41L4A-AS1* affects both proliferation and cell cycle phase distribution in bone marrow-derived mesenchymal stem (BMSC) cells [10].

In a previous study from our group we combined RNA sequencing (RNA-seq) data with ChIP-sequencing (ChIP-seq) signal from cell cycle synchronized HaCaT cells where we identified 94 lncRNAs with a cyclic expression profile. Of these, 59 had a positive correlation to Pol II or histone 3 lysine 4 trimethylation (H3K4me3) signals, consistent with being actively transcribed at specific phases of the cell cycle (Hegre and Samdal *et.al.* unpublished). Among these lncRNAs, we selected *EPB41L4A-AS1* for further functional evaluation. We observed a decreased proliferation in HaCaT, A549, and DLD1 cells in response to siRNA-mediated downregulation of *EPB41L4A-AS1*. In addition, siRNA-mediated knockdown of *EPB41L4A-AS1* led to a reduction of cells present in the G1 phase and an enrichment of cells present in the G2/M phase of the cell cycle in HaCaT cells, indicating a possible role in the G2/M progression.

Here, we used RNA FISH and determined that the subcellular localization of *EPB41L4A-AS1* is both nuclear and extranuclear. We used CRISPR interference and CRISPR activation (CRISPRi and CRISPRa, respectively) to modulate the expression of *EPB41L4A-AS1* at the transcriptional level, followed by viability and cell cycle assays. In line with our previous study, we find that CRISPR-mediated downregulation affects cell cycle phase distribution. The effect on phase distribution was different between HaCaT and A549 cells, suggesting that *EPB41L4A-AS1* has cell type-specific roles. Moreover, *EPB41L4A-AS1* has *cis*-regulatory abilities, and is a positive regulator of its antisense mRNA neighbor *EPB41L4A.* CRISPR-mediated upregulation of *EPB41L4A-AS1* resulted in the opposite effect on phase distribution and *EPB41L4A* expression level. Finally, we used RNA-seq to provide mechanistic insight into the functional role of *EPB41L4A-AS1.* The analysis suggests that *EPB41L4A-AS1* affects the expression of multiple genes in HaCaT and A549 cells, and gene ontology (GO) terms include several growth promoting signaling pathways.

Material and Methods

Cell culture

Cell lines were obtained from the American Type Culture Collection (ATCC) and cultivated in a humidified incubator in 5% CO_2 at 37°C. HaCaT, A549, and Hek293T cells were all cultured in Dulbecco's modified Eagle's medium (DMEM) (Sigma-Aldrich, D6429) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich, F7524) and 2 mM L-Glutamine (Sigma-Aldrich, G7513). For XTT viability assays we used DMEM without phenol red (Thermo Fisher Scientific, 21063029). All cell lines were sub-cultivated at about 70 % confluence at least twice a week.

Single molecule RNA fluorescence *in situ* hybridization (RNA FISH)

We used the software available through Stellaris Probedesigner to design the oligonucleotide set for RNA FISH. For the *EPB41L4A-AS1* we used masking 3, length 20, spacing 2 which resulted in 36 oligonucleotides. Stellaris™ probesets (Biosearch Technologies) were conjugated to a Quasar670 dye in the 3' end. Hybridization and staining were performed as prescribed in the Stellaris protocol for adherent cells. GAPDH was used as a predesigned control. The different probe sequences are listed in Supplementary Table 1.

Imaging

To image the cells, we used a Zeiss Laser TIRF 3 fluorescence microscope (Zeiss), equipped with a α-Plan-Apochromat 100x/1.46 oil-immersion objective. We used a Zeiss 81 HE DAPI/FITC/Rh/Cy5 filter, DAPI was excited by LED-module 365 nm (Zeiss Colibri) and Quasar670 was excited by 644 nm wavelength laser. The images were acquired by either a Hamamatsu EMCCD EMX2 or a Hamamatsu ORCA-Fusion camera at 16 bit and at a voxel size of 100 x 100 x 220 nm^3 (EMCCD) or 129 x 129 x 220 nm size (ORCA-Fusion). We used SVI Huygenes Professional (version 18.10) for image deconvolution and image analysis was done using Fiji (version 1.52t) [13]. Presented images are maximum intensity projections of 34 Z-stack slices (7.26 µm) of the cell.

Plasmids and cloning

We followed the Zhang labs protocol (https://media.addgene.org/data/plasmids/52/52961/52961-attachment_B3xTwla0bkYD.pdf) for guide RNA (gRNA) design and cloning of the gRNA between the two BsmBI restriction sites. Lenti-sgRNA puro was a gift from Brett Stringer (Addgene plasmid # 104990). The hU6-F (5'-GAGGGCCTATTTCCCATGATT-3') was used to sequence the RNA to validate gRNA insert. For CRISPRa we used pXPR_120 [14] with multiple activating domains, VP64, P65 and Rta, and for CRISPRi we used pHR-SFFV-dCas9-BFP-KRAB [15], a gift from John Doench & David Root (Addgene plasmid # 96917) and Stanley Qi & Jonathan Weissman (Addgene plasmid # 46911), respectively.

Lentiviral production

We plated Hek293T cells 24 hours before transfection with Lenti-sgRNA puro containing the different gRNAs or pHR-SFFV-dCas9-BFP-KRAB or pXPR_120 together with the packaging plasmids psPAX.2 and pMD2.G. Lipofectamine 2000 (Invitrogen™, 11668019) was used as a transfection reagent. We replaced the culture medium 8 hours after transfection to decrease toxicity. We collected the viral supernatant 72 hours after transfection, centrifuged at 1800 g for 5 min at 4°C, and filtered it through a 45 µM filter. The virus was stored at -80℃.

Generating stable dCas9 expressing cell lines

Both HaCaT and A549 cells were transfected by adding lentiviral titer of pXPR_120 or pHR-SFFV-dCas9-BFP-KRAB together with 8 µg/ml polybrene in DMEM. We replaced the medium 48 hours after transduction. After transduction, 10 µg/ml Blasticidin (Invivogen) was used to select for cells that had incorporated pXPR_120 into their genome. We changed the medium after three days, while selection was continued for about seven days. To keep the selection pressure, we sub-cultivated the pXPR_120 cells with 5 µg/ml Blasticidin. We used a FACS Aria II cell sorter (BD Bioscience) to select cells transduced with pHR-SFFV-dCas9-BFP-KRAB.

Guide RNA (gRNA) design

We used the E-CRISPR online tool (http://www.e-crisp.org/E-CRISP) for gRNA design [16]. The sequence targeted -50 to + 200 base pairs (bp) relative to the transcriptional start site (TSS) of *EPB41L4A-AS1* (hg38_dna range=chr5:112160660-112160910). The exact location of TSS was determined using Fantom5 web resource (ZENBU 3.0.1) [17, 18]. To avoid off-target effects, we used Basic local alignment Search Tool to discard any sequences that were located within annotated genes other than target. The sequences of different gRNAs are listed in Supplementary Table S2.

gRNA transductions

The cells were transduced with target-specific lentiviral gRNAs or a non-target gRNA together with 8 µg/ml polybrene approximately 20 hours after seeding at about 50% confluence. The multiplicity of infection (MOI) was 2. We added 2 µg/ml puromycin (Invivogen, ant-pr-1) to the growth media 24 hours after transduction to select for resistant cells containing the gRNA and harvested the cells 72 hours after selection was added. We included gRNA specific for *MALAT1* and *SCLA41* as positive controls for CRISPRi and CRISPRa, respectively.

Short interfering (siRNA)-mediated knockdown

All cells were transfected with siRNAs for a 20 nM final concentration using Lipofectamine RNAimax (Invitrogen™, 13778030) when plated, according to the manufacturer's protocol. Cells were harvested after 48 and/or 72 hours after transfection at about 70% confluence. MISSION® siRNA Universal Negative Control #1 (Sigma, SIC001) was used as control for siRNAs. The producers and sequences of siRNAs are listed in Supplementary Table S3.

Quantitative Real-Time PCR (RT-qPCR)

The *mir*Vana miRNA Isolation Kit (Invitrogen™, AM1560) was used to isolate total RNA according to manufacturer's protocol, followed by DNase treatment with TURBO DNA-*free*™ (Invitrogen™, AM1907). RNA concentration and quality were measured on a NanoDrop ND-1000 UV-Vis Spectrophotometer. RNA was converted to cDNA under standard conditions with random hexamer primers using TaqMan™ Reverse Transcription reagents (Invitrogen™, N8080234). Quantitative RT PCR reactions were prepared with Syber select master mix (Applied Biosystems, 4472919). The sequences of the different primers are listed in Supplementary Table S4. Relative expression levels were calculated using the comparative C_T method ($2^{-\Delta\Delta CT}$ method, [19]), and expression data were normalized to GAPDH.

Viability assay

We performed TACS XTT Cell viability Assay (R&D Systems™, 4891025K) to investigate how down- and upregulation of *EPB41L4A-AS1* affected the viability, measured as the level of metabolic activity, in HaCaT cells. We transduced HaCaT cells with a non-target gRNA or an *EPB41L4A-AS1*-specific gRNA and harvested after selection with puromycin. Cells were then seeded in triplicates for each condition in a 96-well tray 96 hours after transduction. We measured absorbance 24, 48, and 72 hours after seeding, 4 hours after XTT was added, following the manufacturer's protocol. We also performed cell counting using Moxi z mini automated cell counter (ORFLO Technologies) to investigate how CRISPRi-mediated downregulation of *EPB41L4A-AS1* affected proliferation in HaCaT cells. Cells were seeded in triplicates for each condition in a 24-well tray and counted 48 and 72 hours after transfection. Each well was washed twice with preheated PBS and trypsinated for 8-10 min before the cells were resuspended in preheated growth medium and counted.

Cell cycle and fluorescence-activated cell sorting (FACS) analysis

For FACS analysis of the cell cycle phase distribution we harvested the cells 96 hours after gRNA transduction, by washing twice with preheated PBS and trypsinating for 8-10 min. Then we collected the cells using ice-cold PBS supplemented with 3% FBS, centrifuged the cells at 4°C for 5 min, and removed the supernatant. Cells were resuspended in 100 µl cold PBS, fixed in 1 ml ice-cold 100% methanol and stored at 4°C until DNA measurement. The cells were washed with cold PBS and

incubated with 200 µl of DNase-free RNAse A in PBS (100 µg/ml) for 30 min at 37°C before DNA staining with 200 µl of Propidium Iodide (PI, Sigma) (50 µg/ml) at 37°C for 30 min. We performed cell cycle analyses by using a BD FACS Canto flow cytometer (BD Biosciences). The excitation maximum of PI is 535 nm and the emission maximum 617 nm. PI stained cells were excited with the blue laser (488 nm), and the PI fluorescence was detected in the PE channel (578 nm). We used FlowJo software to quantify the cells in each phase, and the percentage of cells that was assigned to G1, S, and G2/M phases was calculated.

Total RNA sequencing (RNA-seq)

We used the *mir*Vana miRNA Isolation Kit (ThermoFisher Scientific, AM1560) to isolate total RNA, according to the manufacturer's protocol. RNA concentration was measured on a Qubit (Thermo fisher), whereas integrity and stability of the RNA samples were assessed by using an Agilent 2100 Bioanalyzer (Agilent Technologies). The ribosomal RNA (rRNA) was removed using RiboCop rRNA Depletion Kit for Human/Mouse/Rat V2 (Lexogen, 037), and total RNA-seq libraries were prepared using CORALL (Lexogen, 095), according to the manufacturer's instructions. The samples were sequenced using the Illumina NextSeq 500 Sequencing System.

RNA-seq data analysis

We filtered and trimmed FASTQ files (fastp v0.20.0) and generated transcript counts using quasi alignment (Salmon v1.3.0) to the transcriptome reference sequence (Ensembl, GRCh38 release 92). The transcript sequences were imported into the R statistical software and aggregated to gene counts using the tximport (v1.14.0) bioconductor package. The genes with expression of less than 1 count per million in at least 50% of the samples were filtered out to compare gene expression between samples. We transformed count matrices by using the bioconductor package limma [20] combined with voom transformation [21]. Differentially expressed genes were identified using the toptable function of limma and we used the Benjamini-Hochberg method to correct for multiple testing. Genes with p-values < 0.05 were considered significantly differentially expressed.

Results

***EPB41L4A-AS1* is localized both in and outside the nucleus**

The relative concentration index (RCI) from RNA-seq analyses of cytoplasmic and nuclear fractions in multiple cell lines suggest that *EPB41L4A-AS1* is mainly localized in the cytoplasm (Figure 1A; average RCI of 1.12). To address the question of *EPB41L4A-AS1*'s cellular localization in detail, we imaged HaCaT cells following RNA FISH against *EPB41L4A-AS1*. The data showed that in HaCaT cells, *EPB41L4A-AS1* appears as bright foci and low-intensity spots both in and outside the nucleus (Figure 1B; Supplementary Figure S1). Cross-sections of the nucleus confirmed that some bright foci co-localized with the DNA (Figure 1C).

Generation of stable dCAS9 expressing cell lines

HaCaT is a spontaneously transformed human keratinocyte cell line and A549 is an adenocarcinoma human alveolar basal epithelial cell line. We successfully modified both cell lines to express catalytically inactive CAS9, a nuclease dead CAS9 (dCAS9), fused with either the Krüppel-associated box (KRAB) repressing domain or multiple activating domains to use as tools for transcriptional inhibition (CRISPRi) and activation (CRISPRa), respectively. The same two gRNAs were used for both

up- and downregulation of *EPB41L4A-AS1*, although for A549, only gRNA 1 was able to downregulate the expression of *EPB41L4A-AS1* (Figure 2A).

***EPB41L4A-AS1* affects cell cycle phase distribution**

In a previous study by our group, we reported that siRNA-mediated knockdown of *EPB41L4A-AS1* affected proliferation and cell cycle phase distribution (Hegre and Samdal *et.al.* unpublished). To investigate whether this effect was consistent between different gene perturbation methods, and not just an effect of the post-transcriptional downregulation, we used CRISPRi and CRISPRa to repress and activate the transcriptional level of *EPB41L4A-AS1,* respectively.

We investigated whether CRISPR-mediated modulation of *EPB41L4A-AS1* affected the viability of HaCaT cells by measuring the metabolic activity using an XTT viability assay. We observed a slight decrease in metabolic activity in HaCaT cells transduced with *EPB41L4A-AS1*-specific gRNAs compared to a non-specific gRNA control after 72 hours (Figure 2B), yet there were no significant effect between timepoints when we used CRISPRa to upregulate *EPB41L4A-AS1* (Supplementary Figure S2).

In line with previous results with siRNAs, we observed a decreased percentage of cells present in the G1 phase, and an increase in S and G2/M phases in response to CRISPRi-mediated downregulation of *EPB41L4A-AS1* in HaCaT cells (Figure 2C). As expected, the opposite was observed when *EPB41L4A-AS1* was upregulated, except for the G2/M phase, which displayed a slight enrichment of HaCaT cells in response to CRISPRa (Figure 2C). We used a hierarchical, linear model to calculate the combined effect on cell cycle phase distribution between CRISPRi and CRISPRa, requiring opposite contributions of the two assays and assuming a random effect for each assay. In HaCaT cells we observed a difference in cell cycle phase distribution for the G1 and S phases and these were consistent between the two assays. These changes were verified with another *EPB41L4A-AS1*-specific gRNA (Supplementary Figure S3). For the cell cycle assays we also included A549 cells to investigate whether the effect on cell cycle phase distribution was consistent across different cell lines. Compared with HaCaT cells, CRISPRi/a of *EPB41L4A-AS1* had an opposite effect in A549 cells, where downregulation of *EPB41L4A-AS1* led to a slight increase of cells present in the G1 phase, while the S and G2/M phases were not affected. Upregulation of *EPB41L4A-AS1* resulted in the opposite effect with a decreased percentage of A549 cells in the G1 phase. The number of cells in S phase was slightly increased, while the G2/M phase distribution was unaffected (Figure 2C).

EPB41L4A-AS1* regulates the expression of *EPB41L4A

EPB41L4A-AS1 is localized at chromosome 5 (hg38 chr5:112,160,526-112,164,818) and is transcribed from the opposite strand of *EPB41L4A,* where they partly overlap in a tail-to-tail orientation (Figure 3A). According to tissue expression data from the Genotype-Tissue Expression (GTEx) project, *EPB41L4A-AS1* is expressed across all tissues (Figure 3A; Supplementary Figure S4). Several antisense lncRNAs are known to regulate the expression of the protein coding gene that is overlapping on their opposite strand, and other neighboring genes, in *cis* [16]. To address whether *EPB41L4A-AS1* affects the expression of *EPB41L4A*, we used CRISPRi and CRISPRa-modified HaCaT cells transduced with an *EPB41L4A-AS1*-specific gRNA or a non-target gRNA. CRISPRi-mediated downregulation of *EPB41L4A-AS1* reduced the expression of *EPB41L4A* with two different gRNAs, whereas CRISPRa-mediated upregulation of *EPB41L4A-AS1* increased the expression of *EPB41L4A* (Figure 3B).

We used two different *EPB41L4A-AS1*-specific siRNAs (Figure 3C) to investigate whether the effect of *EPB41L4A-AS1* on *EPB41L4A* was transcriptional or post-transcriptional. In line with the CRISPRi-mediated downregulation, both siRNAs knocked down *EPB41L4A-AS1* (Figure 3C) and caused

a significant downregulation of *EPB41L4A* (Figure 3D). Combined, these results suggest that *EPB41L4A-AS1* positively regulates *EPB41L4A* at the post-transcriptional level.

RNA sequencing identifies genes affected by CRISPRi of *EPB41L4A-AS1*

To get a mechanistic insight into how *EPB41L4A-AS1* affects viability and cell cycle phase distribution, we performed RNA-seq of CRISPRi-modified HaCaT and A549 cell lines transduced with *EPB41L4A-AS1*-specific gRNA or a non-target control gRNA (Figure 4A). Consistent with our RT-qPCR analyses, *EPB41L4A* was downregulated in both cell lines (p-value ≤ 0.05) in response to CRISPRi-mediated downregulation of *EPB41L4A-AS1* (Figure 4B). We identified a total of 1612 and 1371 differentially expressed genes with p-value ≤ 0.05, where 8 and 13 of these genes had an adjusted p-value ≤ 0.05, in HaCaT and A549 cells, respectively (Figure 4C). Moreover, in the joint analysis of HaCaT and A549 cells we identified 1439 differentially expressed genes with p-value ≤ 0.05 (Figure 4C; Supplementary Dataset S1). Gene ontology (GO) enrichment analyses of the differentially expressed genes in HaCaT, A549, or the joint analysis of CRISPRi vs. control showed enrichment of signaling pathways (KEGG and Reactome) associated with proliferation, including NTRK, ERK, and ARMS for the upregulated genes namely *TMEM189-UBE2V1*, *KIDINS220*, *C1orf167-AS1*, *ZSWIM8-AS1*, and *RMRP* (Figure 4D).

Discussion

Our data show that *EPB41L4A-AS1* is localized as bright foci both in and outside the nucleus in HaCaT cells. Consistent with the nuclear foci having a role in gene regulation, we found that *EPB41L4A-AS1* positively regulates the expression of its antisense mRNA *EPB41L4A*, as CRISPR-mediated up- and downregulation of *EPB41L4A-AS1*, up- and downregulates *EPB41L4A*, respectively. Moreover, this effect is post-transcriptional, as we found that siRNA-mediated knockdown of *EPB41L4A-AS1* also downregulates *EPB41L4A*.

In line with our previous study, CRISPR-mediated downregulation of *EPB41L4A-AS1* expression affected cell cycle phase distribution with a reduction of HaCaT cells in the G1 phase, while CRISPRa-mediated upregulation led to an increased percentage of cells in the G1 phase. The effect on cell cycle phase distribution is cell type-specific, as CRISPR-mediated decrease and increase of *EPB41L4A-AS1* expression in A549 cells resulted in an increase and decrease of cells in G1 phase, respectively. Finally, we found that downregulation of *EPB41L4A-AS1* affects the gene expression of multiple genes in both HaCaT and A549 cells, indicating that *EPB41L4A-AS1* has both *cis*- and *trans* regulatory abilities. Several of the affected genes are involved in signaling pathways associated with proliferation, possibly explaining *EPB41L4A-AS1*'s proliferation phenotype.

The subcellular localization of *EPB41L4A-AS1* was both nuclear and extranuclear in HaCaT cells. The extranuclear localization appears to be cytoplasmic, which is in line with *EPB41L4A-AS1* encoding a small protein, TIGA1. TIGA1 is translated from the first exon of *EPB41L4A-AS1,* and a study demonstrated that ectopic expression of TIGA1 reduced colony-formation and growth of the lung cancer cell line EBC-1 in soft agar [22]. Whereas the extranuclear foci suggests that TIGA1 also is present in HaCaT cells, our previous (Hegre and Samdal et al., unpublished) and current results indicate that downregulating *EPB41L4A-AS1* negatively affects HaCaT proliferation. Consequently, either TIGA1 has cell type-dependent effects on proliferation or *EPB41L4A-AS1* affects proliferation through other mechanisms beside TIGA1.

A likely such mechanism is *EPB41L4A-AS1* directly regulating nuclear gene expression, as we show that *EPB41L4A-AS1* localizes to nuclear DNA and that *EPB41L4A-AS1* positively regulates the expression of its antisense mRNA *EPB41L4A.* EPB41L4A is a part of a membrane-associated protein

superfamily that contains a FERM (Four-point-one, Ezrin, Radixin, Moesin) domain. These proteins are important for embryonic development and have regulatory functions involving intracellular trafficking, signal transductions, and cytoskeletal rearrangement [23]. Moreover, *EPB41L4A* is a target gene of the Wnt/β-catenin pathway [24], a pathway that regulates embryonic development and adult tissue homeostasis, in addition to proliferation and differentiation [25]. Depending on their genomic configuration, sense-antisense genes can share regulatory DNA elements that jointly control their gene expression. However, for *EPB41L4A-AS1* and *EPB41L4A* the distance between their transcription start sites (TSS) is 258751 bb (Figure 3A). Although the two TSSs could still be close in three-dimensional space in the nucleus, our result showing that siRNA-mediated knockdown of *EPB41L4A-AS1* also downregulates *EPB41L4A* strongly supports that the regulation is post-transcriptional and caused by the RNA of *EPB41L4A-AS1*. Indeed, *EPB41L4A-AS1*'s bright foci that co-localize with the DNA in HaCaT cells are consistent with an accumulation of *EPB41L4A-AS1* at its site of transcription.

As mentioned above, we found that CRISPR-mediated downregulation of *EPB41L4A-AS1* slightly reduced the viability in HaCaT cells, which supports results from our recent study where siRNA-mediated knockdown of *EPB41L4A-AS1* reduced proliferation of the colorectal cancer (CRC) cell line DLD1, as well as in HaCaT and A549 cells (Hegre and Samdal et.al. unpublished). In line with our results, Bin et. al. found that knockdown of *EPB41L4A-AS1* in the CRC cell lines HCT116 and SW620 reduced proliferation, while overexpression of *EPB41L4A-AS1* had the opposite effect. Moreover, the expression of *EPB41L4A-AS1* was higher in CRC tissue compared to normal tissue. They concluded that *EPB41L4A-AS1*functions as an oncogene by regulating the Rho/ROCK pathway [26]. Another study identified *EPB41L4A-AS1* as a possible hub lncRNA in colorectal cancer [9], indicating a potential role as a prognostic marker.

We further validated that *EPB41L4A-AS1* is involved in the regulation of cell cycle progression by using CRISPRi to downregulate *EPB41L4A-AS1* in HaCaT and A549 cells. In addition, we strengthened the evidence by using CRISPRa to upregulate *EPB41L4A-AS1,* which affected the cell cycle phase distribution in the opposite direction. Moreover, the effect on cell cycle phase distribution was different in HaCaT and A549 cells, suggesting cell type-specific functions of *EPB41L4A-AS1.* Other recent studies indicate that the expression of *EPB41L4A-AS1* and its biological role seems to vary in different types of cells, tissues, and cancer stages. A study from Rao et.al. found a correlation between low expression of *EPB41L4A-AS1* and early-stage breast cancer [8], whereas two other studies reported an upregulation of *EPB41L4A-AS1* in cervical and ovarian cancer [27, 28]. One of the studies identified *EPB41L4A-AS1* as a central hub in a sub-network in cervical cancer. Analysis indicated an interaction between *EPB41L4A-AS1* and the genes *CCND2* and *VDAC1* via different microRNAs, proposing a potential role as a biomarker for disease progression and as a therapeutic target [27]. In the study from Zhou et.al, *EPB41L4A-AS1* was categorized as a risky lncRNA linked to stage progression and poor clinical outcome in ovarian cancer, and might be a potential candidate for classification of patients into subgroups [28].

In some functional studies, *EPB41L4A-AS1* was reported to be involved in metabolic reprogramming and as a repressor of the Warburg effect in placental tissue of miscarriage [12] and in cancer cells (cervical, breast, bladder and liver) [11]. Liao et.al. found that *EPB41L4A-AS1* was regulated by p53 at the transcriptional level by direct binding to the promoter in Hela and HepG2 cells. Also, stable knockdown of *EPB41L4A-AS1* in HeLa cells increased tumor colony formation, a process that was dependent on the presence of glutamine metabolism. Furthermore, an increased therapeutic effect of a glutaminase inhibitor in *EPB41L4A-AS1* knockdown cells was confirmed in vivo in mice [11].

Another study identified *EPB41L4A-AS1* as a direct *MYC* target gene, which was downregulated in response to MYC overexpression in P493-6 human B-cells [29].

As *EPB41L4A-AS1* has been detected as dysregulated in several cancers, with functions affecting several cellular processes, including metabolic reprogramming, tumor colony formation, gene regulation, cell cycle, and proliferation, it probably has *trans* regulatory abilities. We found that downregulation of *EPB41L4A-AS1* resulted in differentially expressed genes in both A549 and HaCaT cells, including several genes involved in proliferation, such as *SMURF1, F2RL1, CPNE3,* and *KIDINS220*. Moreover, GO terms included several growth-promoting pathways, including NTRK, ERK, and ARMS mediated signaling.

In summary, *EPB41L4A-AS1* is located both in the nucleus and cytoplasm in HaCaT cells. We demonstrate that *EPB41L4A-AS1* has a slight effect on cellular metabolic activity and that it exhibits cell type-specific functions as it affects the cell cycle phase distribution differently in HaCaT and A549 cells. Finally, *EPB41L4A-AS1* regulates the expression of its antisense mRNA *EPB41L4A* through post-transcriptional effects and downregulating *EPB41L4A-AS1* affects more than a thousand other genes in HaCaT and A549 cells, including genes involved in proliferation and growth-promoting signaling pathways.

Acknowledgement

This work was supported by the Norwegian Cancer Society (grant number 2278701) and the Research Council of Norway (grant number 230338). The library preparation and sequencing were provided by the Genomics Core Facility (GCF), Norwegian University of Science and Technology (NTNU). GCF is funded by the Faculty of Medicine and Health Sciences at NTNU and Central Norway Regional Health Authority. The light microscopy was provided by the Cellular and Molecular Imaging Core Facility (CMIC), NTNU. CMIC is funded by the Faculty of Medicine at NTNU and Central Norway Regional Health Authority.

Author contribution

HS planned and performed imaging and wet lab experiments, contributed with statistical analysis, and wrote the manuscript. SAH contributed with statistical analysis, generated all the figures and was a major contributor in editing the manuscript. KC analyzed the RNA-seq data. NBL performed all FACS analysis. PAA contributed with the planning and design of CRISPR experiments. BS contributed to the planning of RNA FISH imaging and performed the image analysis. PS contributed to the study design, supervised the project, and edited the manuscript. All authors read and approved the final manuscript.

Declaration of interest

The authors declare that they have no competing interests.

Supplementary Data

Supplementary file 1: Supplementary Figures and Tables

Supplementary dataset 1: RNA-seq identified a total of 1439 differentially expressed genes with *p*-value ≤ 0.05 in the joint analysis of HaCaT and A549 cells.

References

1. Kopp, F. and J.T. Mendell, *Functional Classification and Experimental Dissection of Long Noncoding RNAs.* Cell, 2018. **172**(3): p. 393-407.
2. Chen, L.L., *Linking Long Noncoding RNA Localization and Function.* Trends Biochem Sci, 2016. **41**(9): p. 761-772.
3. Ji, Q., et al., *MALAT1 regulates the transcriptional and translational levels of proto-oncogene RUNX2 in colorectal cancer metastasis.* Cell Death Dis, 2019. **10**(6): p. 378.
4. Kim, J., et al., *Long noncoding RNA MALAT1 suppresses breast cancer metastasis.* Nat Genet, 2018. **50**(12): p. 1705-1715.
5. Kwok, Z.H., et al., *A non-canonical tumor suppressive role for the long non-coding RNA MALAT1 in colon and breast cancers.* Int J Cancer, 2018. **143**(3): p. 668-678.
6. Arun, G. and D.L. Spector, *MALAT1 long non-coding RNA and breast cancer.* RNA Biol, 2019. **16**(6): p. 860-863.
7. Arun, G., D. Aggarwal, and D.L. Spector, *MALAT1 Long Non-Coding RNA: Functional Implications.* Noncoding RNA, 2020. **6**(2).
8. Deva Magendhra Rao, A.K., et al., *Identification of lncRNAs associated with early-stage breast cancer and their prognostic implications.* Mol Oncol, 2019. **13**(6): p. 1342-1355.
9. He, M., Y. Lin, and Y. Xu, *Identification of prognostic biomarkers in colorectal cancer using a long non-coding RNA-mediated competitive endogenous RNA network.* Oncol Lett, 2019. **17**(3): p. 2687-2694.
10. Cui, P., et al., *miR-146a interacting with lncRNA EPB41L4A-AS1 and lncRNA SNHG7 inhibits proliferation of bone marrow-derived mesenchymal stem cells.* J Cell Physiol, 2019.
11. Liao, M., et al., *LncRNA EPB41L4A-AS1 regulates glycolysis and glutaminolysis by mediating nucleolar translocation of HDAC2.* EBioMedicine, 2019. **41**: p. 200-213.
12. Zhu, Y., et al., *Overexpression of lncRNA EPB41L4A-AS1 Induces Metabolic Reprogramming in Trophoblast Cells and Placenta Tissue of Miscarriage.* Mol Ther Nucleic Acids, 2019. **18**: p. 518-532.
13. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
14. Najm, F.J., et al., *Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens.* Nat Biotechnol, 2018. **36**(2): p. 179-189.
15. Gilbert, L.A., et al., *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* Cell, 2013. **154**(2): p. 442-51.
16. Heigwer, F., G. Kerr, and M. Boutros, *E-CRISP: fast CRISPR target site identification.* Nat Methods, 2014. **11**(2): p. 122-3.
17. Lizio, M., et al., *Gateways to the FANTOM5 promoter level mammalian expression atlas.* Genome Biol, 2015. **16**: p. 22.
18. Lizio, M., et al., *Update of the FANTOM web resource: expansion to provide additional transcriptome atlases.* Nucleic Acids Res, 2019. **47**(D1): p. D752-D758.
19. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.
20. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
21. Law, C.W., et al., *voom: Precision weights unlock linear model analysis tools for RNA-seq read counts.* Genome Biol, 2014. **15**(2): p. R29.
22. Yabuta, N., et al., *Isolation and characterization of the TIGA genes, whose transcripts are induced by growth arrest.* Nucleic Acids Res, 2006. **34**(17): p. 4878-92.
23. Tepass, U., *FERM proteins in animal morphogenesis.* Curr Opin Genet Dev, 2009. **19**(4): p. 357-67.
24. Ishiguro, H., et al., *Isolation and characterization of human NBL4, a gene involved in the beta-catenin/tcf signaling pathway.* Jpn J Cancer Res, 2000. **91**(6): p. 597-603.
25. MacDonald, B.T., K. Tamai, and X. He, *Wnt/beta-catenin signaling: components, mechanisms, and diseases.* Dev Cell, 2009. **17**(1): p. 9-26.
26. Bin, J., et al., *Long noncoding RNA EPB41L4A-AS1 functions as an oncogene by regulating the Rho/ROCK pathway in colorectal cancer.* J Cell Physiol, 2020.
27. Chen, P., et al., *Comprehensive analysis of aberrantly expressed long noncoding RNAs, microRNAs, and mRNAs associated with the competitive endogenous RNA network in cervical cancer.* Mol Med Rep, 2020. **22**(1): p. 405-415.

28. Zhou, M., et al., *Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lncRNA biomarkers in human ovarian cancer.* Oncotarget, 2016. **7**(11): p. 12598-611.
29. Hart, J.R., et al., *MYC regulates the non-coding transcriptome.* Oncotarget, 2014. **5**(24): p. 12543-54.

Figures

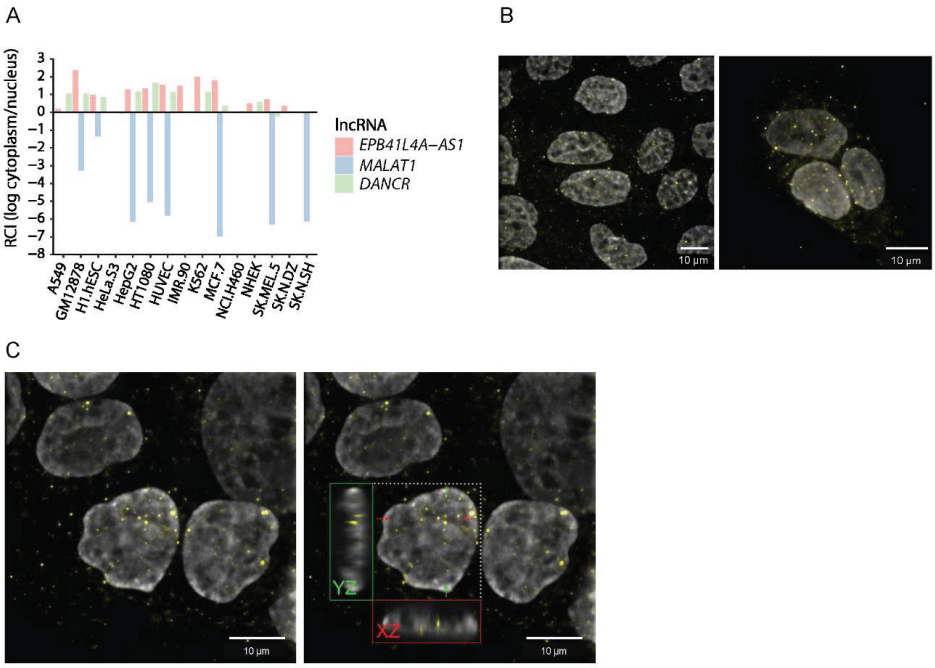


Figure 1: *EPB41L4A-AS1* has a nuclear and extranuclear localization

(A) The subcellular localization of *EPB41L4A-AS1*, *MALAT1,* and *DANCR.* Data are from the lncAtlas (https://lncatlas.crg.eu/) and display the subcellular localization based on the RCI of RNA between cytoplasm and nucleus. Values > 0 indicate cytoplasmic enrichment. (B) RNA FISH showing the localization of EPB41L4A-AS1 (yellow) in fixed HaCaT cells stained with DAPI (grey). Images are presented as maximum intensity projections of 34 Z-stack slices (7.26 µm) of the cell. (C) Cross section of the nucleus of fixed HaCaT cells shows *EPB41L4A-AS1* (yellow) localized with DNA (DAPI, grey). Images are presented as maximum intensity projections of 34 Z-stack slices (7.26 µm) of the cell. The red and green boxes show the signals from the cell planes marked by the red and green arrows projected onto the XZ and YZ planes, respectively.

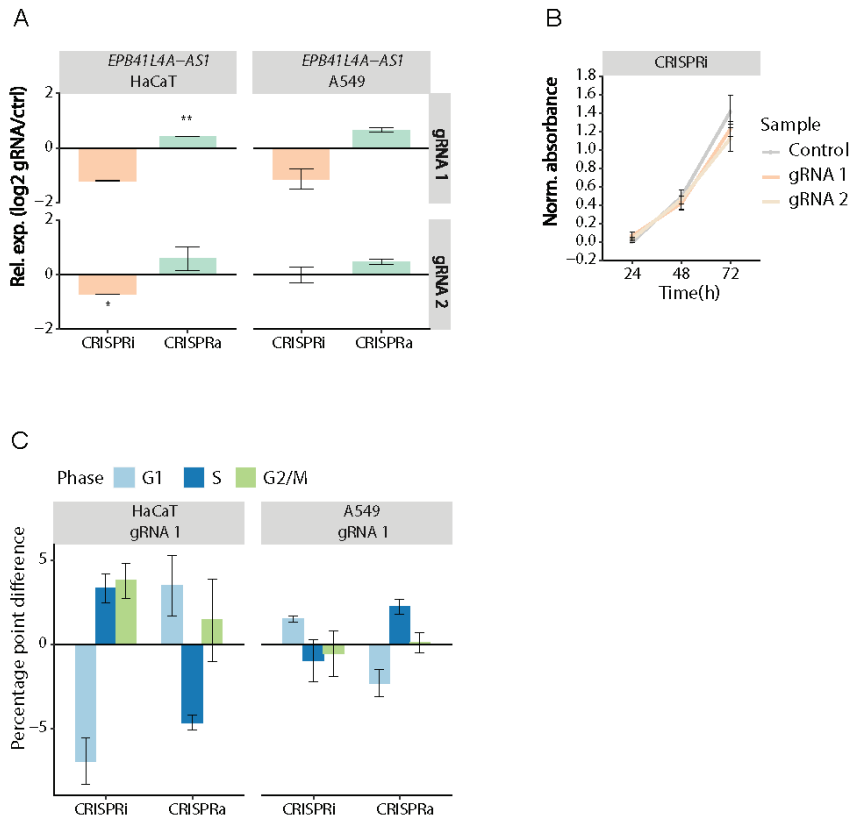


Figure 2. *EPB41L4A-AS1* affects cell cycle distribution

(A) The relative expression level of *EPB41L4A-AS1* as measured by RT-qPCR in response to CRISPRi/a of *EPB41L4A-AS1* in HaCaT and A549 cells. Data are presented as fold change expressions of cells transduced with gRNA 1 and gRNA 2 relative to control gRNA. Bars and error bars are mean and standard error of mean (SEM) of two independent replicates. Significant differences were determined by Student's *t*-test, (unpaired, two-tailed) assuming equal variances (*$p \leq 0.05$, **$p \leq 0.01$). (B) Effect of CRISPRi of *EPB41L4A-AS1* on the metabolic activity in HaCaT cells, as measured by XTT. Data are presented as normalized absorbance ($A_{465\ nm} - A_{630\ nm}$). Bars and error bars are mean and SEM of three independent replicates. *P*-values were determined by Student's *t*-test, (paired, two-tailed) assuming equal variances (gRNA 1 $p = 0.111$ and $p = 0.474$, gRNA 2 $p = 0.784$ and $p = 0.092$ for 48 h and 72 h, respectively). (C) The distribution of cells in G1, S, and G2/M cell cycle phases in response to CRISPRi/a of *EPB41L4A-AS1* in HaCaT and A549 cells. Data are the difference in percentages of G1, S, and G2/M cells between cells transduced with target-specific gRNA 1 to those transduced with a control gRNA. Bars and error bars are mean and SEM of two or more independent replicates. ANOVA *p*-values were calculated from a hierarchical, linear model. HaCaT: G1: 1.7e-04, S: 0.052, and G2/M: 0.095; A549: G1: 0.011, S: 0.023, and G2/M: 0.59.

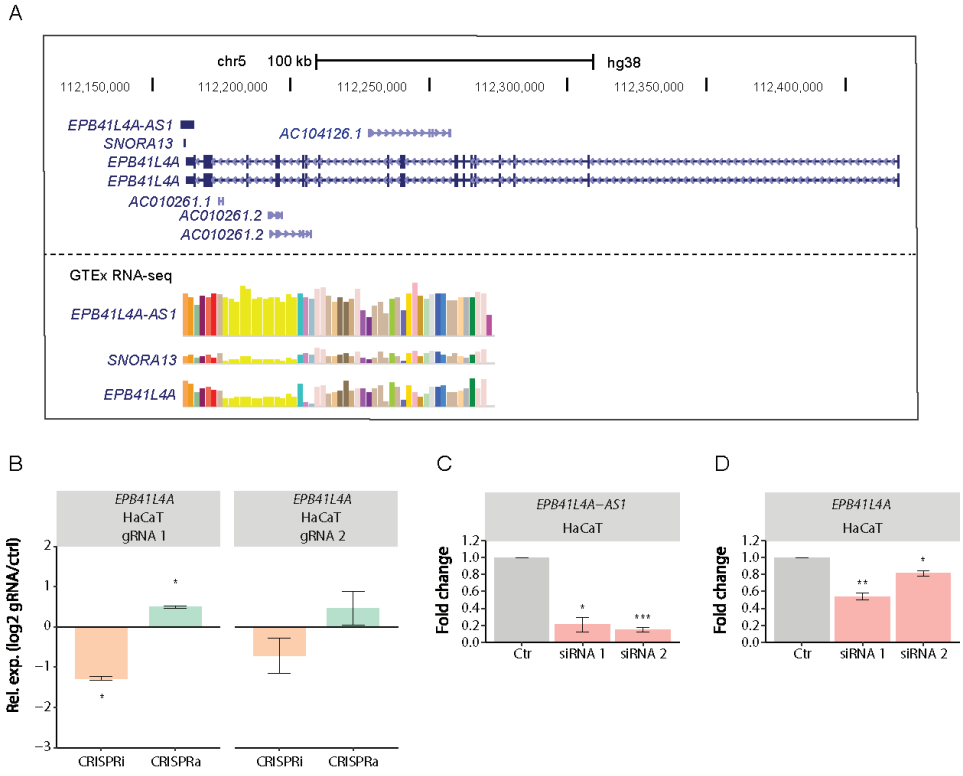


Figure 3. *EPB41L4A-AS1* is a positive regulator of *EPB41L4A*

(A) The genomic loci of *EPB41L4A-AS1* from the UCSC Genome Browser (https://genome.ucsc.edu/). Tissue expression data (transcript per kilobase million, TPM) are from the GTEx project. (B) The relative expression level of *EPB41L4A* as measured by RT-qPCR in response to CRISPRi/a of *EPB41L4A-AS1* in HaCaT cells. Data are presented as fold change expressions of cells transduced with target-specific gRNA 1 and gRNA 2 relative to control gRNA. Bars and error bars are mean and SEM of two independent replicates. Significant differences were determined by Student's t-test, (unpaired, two-tailed) assuming equal variances (*$p \leq 0.05$). (C) Percentage downregulation of *EPB41L4A-AS1* by using two siRNAs with different target sequences. Data are presented as fold change expressions of *EPB41L4A-AS1* following siRNA treatment relative to control-treated cells as measured by RT-qPCR in HaCaT cells. Bars and error bars are mean and SEM of three independent replicates. Significant differences were determined by Student's *t*-test, (unpaired, two-tailed) assuming equal variances (*$p \leq 0.05$; ***$p \leq 0.001$). (D) The relative expression level of *EPB41L4A* as measured by RT-qPCR in response to siRNA-mediated knockdown of *EPB41L4A-AS1* in HaCaT cells. Bars and error bars are mean and SEM of three independent replicates. Significant differences were determined by Student's *t*-test, (unpaired, two-tailed) assuming equal variances (*$p \leq 0.05$; **$p \leq 0.01$).

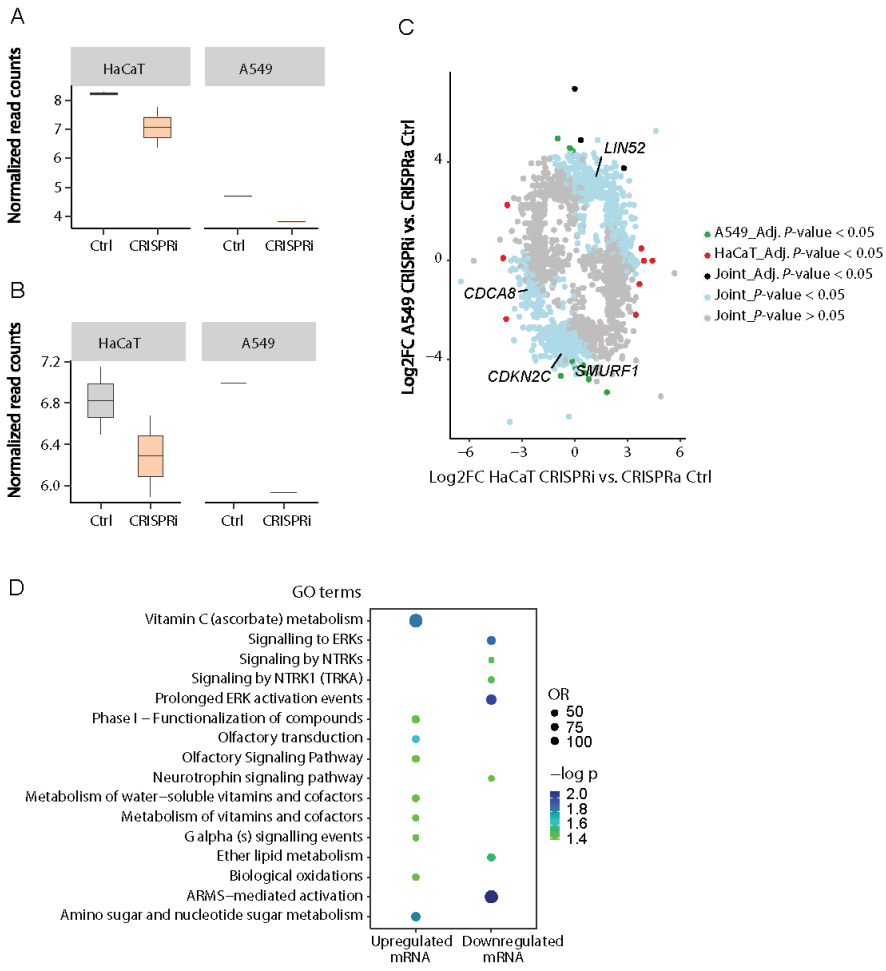
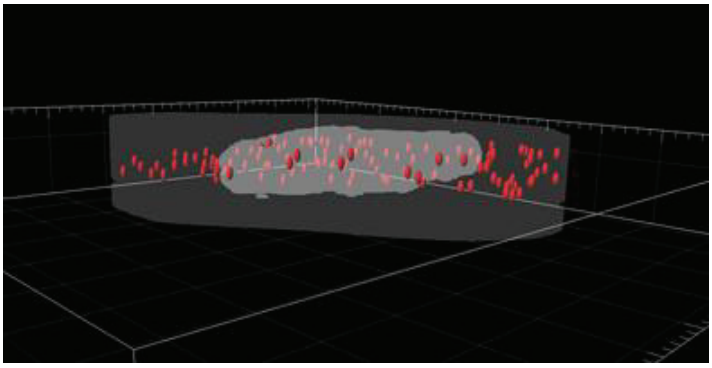


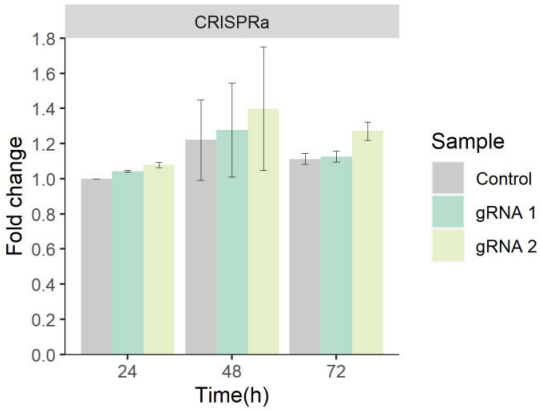
Figure 4. *EPB41L4A-AS1* affects gene expression

(A) Box plot showing RNA-seq expression levels of *EPB41L4A-AS1* in CRISPRi-modified HaCaT (n=2) and A549 cells (n=1) transduced with an *EPB41L4A-AS1*-specific gRNA compared to a non-specific gRNA control. Data are presented as normalized read counts. Median is indicated by the central line in the boxes and whiskers indicate the first and third quartiles (25th and 75th percentiles). (B) Box plot showing RNA-seq expression levels of *EPB41L4A* in CRISPRi-modified HaCaT (n=2) and A549 cells (n=1) transduced with an *EPB41L4A-AS1*-specific gRNA compared to a non-specific gRNA control. Data are presented as normalized read counts. Median is indicated by the central line in the boxes and whiskers indicate the first and third quartiles (25th and 75th percentiles). (C) Differentially expressed genes in CRISPRi compared to control in the individual analysis of HaCaT and A549 cells. There were 2871 differentially expressed genes colored according to *p*-values in individual and joint analysis (CRISPRi vs. Control in the joint data from HaCaT and A549 cell lines). (D) GO enrichment analysis of differentially expressed genes (adjusted *p*-value < 0.05) in either A549, or HaCaT, or in the joint analysis, combined by requiring the same effect direction in each comparison. The terms are from KEGG and Reactome. The colors represent the negative log10 of adjusted *p*-values, so dark blue color represents more significant GO terms. Size of the dots represent the odds ratio (OR).

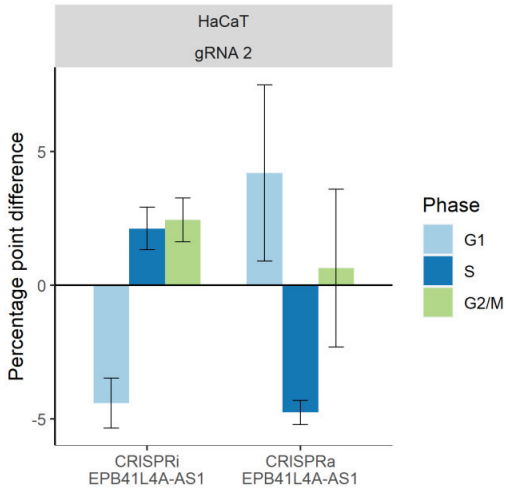
Supplementary Figures



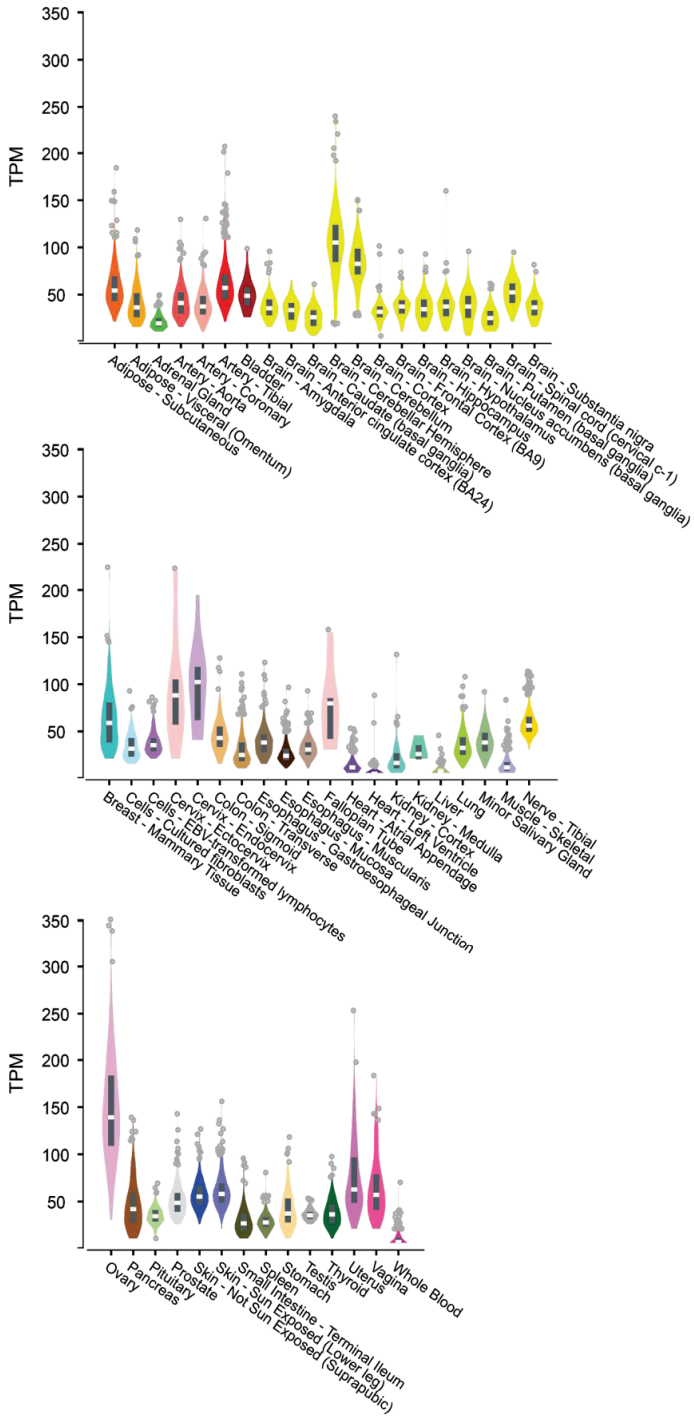
Supplementary Figure S1: Cross-sections of the nucleus in HaCaT cells show *EPB41L4A-AS1* (red) localized with DNA (DAPI, grey).



Supplementary Figure S2: Effect of CRISPRa of *EPB41L4A-AS1* on the metabolic activity in HaCaT cells, as measured by XTT. Data are presented as normalized absorbance ($A_{465\ nm} - A_{630\ nm}$). Bars and error bars are mean and standard error of mean (SEM) of three independent replicates.



Supplementary Figure S3: Distribution of HaCaT cells in G1, S, and G2/M cell cycle phases in response to CRISPRi/a of *EPB41L4A-AS1*. Data are the difference in percentages of G1, S, and G2/M cells between cells transduced with target-specific gRNA 2 to those transduced with a control gRNA. Bars and error bars are mean and SEM of two or more independent replicates. ANOVA *p*-values were calculated from a hierarchical, linear model: G1: 6.2e-04, S: 0.009, and G2/M: 0.173.



Supplementary Figure S4: Tissue expression (transcript per kilobase million, TPM) for *EPB41L4A-AS1* (ENSG00000224032.6) from the Genotype-Tissue Expression (GTEx) project.

Supplementary Tables

Supplementary Table S1: Custom Stellaris FISH probes conjugated to a Quasar670 dye in the 3′ end.

Probe sequences
cacgtattttgggacccgac
aggagcccaaaggactttag
gagcatatgatcagtgctgg
agaggaggtatccagacatc
gatgtgtggccagttctttc
gtcaccaaagggaggaggag
gagggagaagtgccagtgag
ttctacaagtggaaggaccc
aagaaaggacaggcttccgt
acagcaagctcgcaccaaaa
cagaggacccagagcaaaaa
agtcgactcaccagactcag
ttccaaacgtgggaacgcag
ctggcatagtcgatgatgta
aaccaggcttatctggagaa
aggccttttcactgacgaaa
gggcaagcataaagtcagtt
gcacgctgcttttaagttat
gcacgtggaaacagctgtaa
tgtagggtaggtgatcactc
ctacacttcaggcatccatc
agatggtattcagctagcag
ttcaggtcacctttatgctc
caaaatcctcaccaaggaca
ggcaactcttgtgaagatca
tcagactcccagaaatttca
caggtaattatgtcagtact
aatgtgtatttacagactcc
ttattcacatcctcactgtc
acacaaatgccaaagtgcat
gtttctggcagttaatcact
tcctcccttataacttttta
ggtgacagcagtgaatactg
acacaagctattttaaggct
aagtacatttctctttgggc
tagccacgatttttgagtca

Supplementary Table S2: Oligos used for guide RNA (gRNA) cloning.
gRNAs were ordered from Sigma-Aldrich

Gene primer	Oligo sequences (gRNA sequences in bold)	CRISPRi/a
neg_ctr_F	CACCG**TGCGATGGGGGGGTGGGTAGC**	Negative control (ctr)
neg_ctr_R	AAAC**GCTACCCACCCCCCCATCGCA**C	
MALAT1 _F	CACCG**CAGCCCGAGACTTCTGTAA**	Positive ctr CRISPRi
MALAT1 _R	AAAC**TTACAGAAGTCTCGGGCTG**C	
SLC4A1 _F	CACCG**TCAGGAGAACCATGGGGACC**	Positive ctr CRISPRa
SLC4A1_R	AAAC**GGTCCCCATGGTTCTCCTGA**C	
Epb_AS1_1_F (gRNA1)	CACCG**GCGTCTCGCCATAGCGCAG**	CRISPRi/a
Epb_AS1_1_R	AAAC**CTGCGCTATGGCGAGACGC**C	
Epb_AS1_2_F (gRNA2)	CACCG**TGACTCGGGCTGAGAACCC**	CRISPRi/a
Epb_AS1_2_R	AAAC**GGGTTCTCAGCCCGAGTCA**C	

F = forward, R = reverse

Supplementary Table S3: siRNAs used in RNA interference experiments.

Target lncRNA	Alias	Cat. No.	Producer	Sense sequence
SiRNA neg control	Ctr	SIC001	Sigma	
EPB41L4A-AS1	E1	CTM-431274	Dharmacon	ACAGUGAGGAUGUGAAUAAUU
EPB41L4A-AS1	E2	CTM-431275	Dharmacon	GUGAGGAUGUGAAUAAAUAUU

Supplementary Table S4: Primers used for RT-qPCR.
Qiagen QuantiTect Primer Assays (249900) were used for mRNA expression analysis and Qiagen RT^2 lncRNA PCR assays (330701) were used for lncRNA expression analysis.

Gene	Primer assay	Cat.no
GAPDH	Hs_GAPDH_1_SG QuantiTect Primer Assay	QT00079247
EPB41L4A	Hs_EPB41L4A_1_SG QuantiTect Primer Assay	QT00101003
SLC4A1	Hs_SLC4A1_1_SG QuantiTect Primer Assay	QT00068502
MALAT1	RT^2 lncRNA qPCR Assay for Human MALAT1	LPH18065A
EPB41L4A-AS1	RT^2 lncRNA qPCR Assay for Human EPB41L4A-AS1	LPH04387A

ISBN 978-82-326-6873-1 (printed ver.)
ISBN 978-82-326-6905-9 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology