Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

Ashish Goyal, MSc.

born in: Jaipur, India

Oral-examination: 5th May 2017

CRISPR/Cas9-mediated interrogation of a sense-antisense pair involved in the DNA damage response

Referees: Prof. Dr. Christine Clayton

Prof. Dr. Sven Diederichs

Summary

Previously, the Diederichs lab identified the long non-coding RNA *NOP14-AS1* to be induced in A549 and HepG2 cells treated with the DNA-damaging drugs Etoposide, Cisplatin, and Bleomycin. Preliminary data indicated that *NOP14-AS1* was inversely co-regulated with its antisense protein-coding gene *NOP14*. In this study, *NOP14-AS1* regulation in DNA damage was further characterized. *NOP14-AS1* was consistently induced in several cell lines upon treatment with various DNA damaging agents. *NOP14-AS1* induction was a p53-dependent transcriptional response. The DNA damage-induced inverse co-regulation between *NOP14* and *NOP14-AS1* was further confirmed in multiple cell lines using time-course expression analysis upon treatment with multiple DNA-damaging agents. *NOP14* repression upon DNA damage was p53-dependent and preliminary experiments performed indicated its involvement in the p53 pathway and a role in regulating cell proliferation.

Antisense lncRNAs often regulate the expression of their overlapping sense protein-coding genes through transcript- as well as transcription-dependent mechanisms. Therefore, it was proposed that *NOP14-AS1* induction could lead to *NOP14* repression *in cis*. The transcript-dependent effects of *NOP14-AS1* could be uncovered using an RNA interference-/ antisense oligo-based loss-of-function approach. However to uncover any transcription-dependent effects, a CRISPR/Cas9-based knockdown (CRISPRi) and overexpression (CRISPRa) system for *NOP14-AS1* was established. While establishing this system, a major limitation of the CRISPR/Cas9 system for lncRNA knockdown was discovered and experimentally demonstrated for multiple examples. CRISPR/Cas9 targeting of a lncRNA arising from a complex locus can unintentionally affect the expression of its neighboring genes advocating for caution while using these systems for knockdown. However, these systems were determined to be safe to study the impact of *NOP14-AS1* on *NOP14* expression.

NOP14 knockdown using siRNAs had no impact on *NOP14-AS1* expression. On the other hand, *NOP14* transcriptional repression using CRISPRi resulted in a *NOP14-AS1* induction which could not be reversed by an ectopic rescue of *NOP14* expression, indicating that this was due to a reduced transcriptional interference. However, the *NOP14-AS1* induction observed upon CRISPRi knockdown of *NOP14* could not account for the much stronger *NOP14-AS1* induction observed upon DNA damage, indicating that an independent mechanism was responsible for the latter. *NOP14-AS1* knockdown using antisense oligos / CRISPRi did not have any impact on the *NOP14* expression. The DNA damage-induced *NOP14* repression could not be reversed upon *NOP14-AS1* knockdown. Also, a CRISPRa-mediated *NOP14-AS1* induction did not affect *NOP14* expression. Together these data indicated that neither the *NOP14-AS1* transcript nor its transcription could regulate *NOP14* expression. In summary, this study concludes that *NOP14-AS1* and *NOP14* are independently regulated upon DNA damage. However, the role of *NOP14-AS1* in DNA damage remains to be established.

Zusammenfassung

Die lange nicht-kodierende RNA (lncRNA) *NOP14-AS1*, wurde vom Diederichs lab als eine durch Etoposid, Cisplatin und Bleomycin (DNA schädigende Mittel) induzierbare lncRNA in A549 und HepG2 Zellen identifiziert. Vorläufige Daten zeigten, dass *NOP14-AS1* invers ko-reguliert wird mit seinem Antisense-Protein-kodierenden Gen *NOP14*. In dieser Studie wurde die *NOP14-AS1*-Regulation bei induzierter DNA-Schädigung weiter charakterisiert. *NOP14-AS1* konnte in mehreren Zelllinien bei Behandlung mit verschiedenen DNA-schädigenden Mitteln reproduzierbar induziert werden. Die Induktion von *NOP14-AS1* wurde von p53 vermittelt. Die DNA-schädigungsinduzierte inverse Ko-Regulation zwischen *NOP14* und *NOP14-AS1* wurde in mehreren Zelllinien unter Verwendung von Zeitverlaufsexpressionsanalysen bei gleichzeitiger Behandlung mit verschiedenen DNA-schädigenden Mitteln weiter bestätigt. Die transkriptionelle Repression von *NOP14* nach DNA-Schädigung war p53-abhängig, und vorläufige Experimente zeigten die Beteiligung von *NOP14-AS1* an der p53-Signalkaskade und eine Rolle bei der Regulierung der Zellproliferation.

Antisense-lncRNAs regulieren häufig die Expression ihrer überlappenden sense-Protein-kodierenden Gene durch Transkript- sowie Transkriptions-abhängige Mechanismen. Daher wurde vermutet, dass die *NOP14-AS1*-Induktion zu einer *NOP14*-Repression *in cis* führen könnte. Die Transkript-abhängigen Effekte von *NOP14-AS1* konnten unter Verwendung eines RNA-Interferenz- / Antisense-Oligo-basierten Funktionsverlust-Ansatzes aufgedeckt werden. Um die transkriptionsabhängigen Effekte untersuchen zu können, wurde ein CRISPR / Cas9-basiertes Knockdown- (CRISPRi) und Überexpressions- (CRISPRa) -System für *NOP14-AS1* etabliert. Während der Etablierung von CRISPRi wurde festgestellt, dass die Beeinflussung der Expression einer lncRNA in einem komplexen Locus unbeabsichtigte Effekte auslösen kann. So wurde an mehreren Beispielen gezeigt, dass die Expression der benachbarten Gene durch CRISPRi ebenfalls beeinflusst wurde. Daher sollte man bei der Nutzung von CRISPR / Cas9 zur Modulation der Expression von lncRNAs prinzipiell vorsichtig sein, insbesondere bei CRISPRi. In dieser Studie wurde die Verwendung von CRISPRi/a jedoch als geeignet befunden, um die Auswirkungen von *NOP14-AS1* auf die *NOP14*-Expression zu untersuchen.

Der Knockdown von *NOP14* mittels siRNA hatte keinen Einfluss auf die *NOP14-AS1*-Expression. Bei der Verwendung von CRISPRi zum Knockdwon von *NOP14*, zeigte sich jedoch eine *NOP14-AS1*-Induktion. Die Induktion von *NOP14-AS1* konnte auch mittels ektopischer *NOP14*-Expression nicht rückgängig gemacht werden, was als Ursache einen transkriptionellen Interferenz Mechanismus nahe legt. Allerdings konnte die *NOP14-AS1*-Induktion, die beim CRISPRi-Knockdown von *NOP14* beobachtet wurde, nicht für die viel stärkere *NOP14-AS1*-Induktion verantwortlich sein, die bei der Benutzung von DNA-schädigenden Mitteln auftrat. Der Knockdown von *NOP14-AS1* mittels Antisense-Oligos / CRISPRi, hatte keine Auswirkung auf die *NOP14*-Expression. Auch die durch DNA-schädigende Mittel verusachte *NOP14*–Repression konnte durch einen *NOP14-AS1*-Knockdown nicht rückgängig gemacht werden. Weiterhin hatte die CRISPRa-vermittelte *NOP14-AS1*-Induktion keinen Einfluss auf die *NOP14*-Expression. Die Gesamtschau der Daten zeigt, dass weder das *NOP14-AS1*-Transkript noch seine Transkription die *NOP14*-Expression regulieren können. Zusammenfassend kommt diese Studie zu dem Schluss, dass *NOP14-AS1* und *NOP14* im Falle von auftretender DNA-Schädigung unabhängig voneinander reguliert sind. Die Rolle von *NOP14-AS1* bei der zellulären Antwort auf DNA-Schäden muss noch weiter geklärt werden.

Table of contents

List of Figures

List of Tables

Abbreviations

°C	degree Celsius
µg	microgram
µl	microliter
ActD	Actinomycin D
ANRIL	Antisense non-coding RNA in the INK4 locus
ARF	Alternative Reading Frame
AS	Antisense
BLEO	Bleomycin
bp	base pair(s)
BRCA1	Breast Cancer 1
BSA	Bovine Serum Albumin
CAGE	Cap Analysis of Gene Expression
CARBO	Carboplatinum
Cas	CRISPR-associated
Cas9	CRISPR-associated protein 9
CBP	CREB-binding protein
CCND1	Cyclin D1
CDKN1A	Cyclin-Dependent Kinase Inhibitor 1a
cDNA	complementary DNA
CDS	coding sequence
ceRNA	competing endogenous RNA
chip-seq	Chromatin Immunoprecipitation sequencing
CIS	Cisplatin
CRIPRa	CRISPR activation
CRIPRi	CRISPR interference
CRISPR	Clustered Regularly Interspaced Palindromic Repeats
dCas9	dead Cas9
DDR	DNA damage response
DDSR1	DNA damage-sensitive RNA1
DINO	Damage Induced Noncoding
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DOXO	Doxorubicin
dsDNA	double-stranded DNA
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGFP	Enhanced Green Fluorescent Protein
ERIC	E2F1-Regulated Inhibitor of Cell death
ETO	Etoposide
FBS	Fetal Bovine Serum
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
h	hour(s)
H3K27	Histone 3 Lysine 27
HBA1	Hemoglobin α1

HDR	Homology-Directed Repair
HF	High Fidelity
hnRNP	heterogeneous nuclear ribonucleoprotein
HOTAIR	HOX antisense intergenic RNA
HOXC	Homeobox C cluster
HOXD	Homeobox D cluster
HRP	horseradish peroxidase
indel	Insertions or deletions
INK4	Inhibitors of CDK4
iv	improved version
kb	kilobases
KRAB	Krüppel-associated box
LB	Luria Bertani
lincRNA	Large intergenic non-coding RNA
LINP1	lncRNA in NHEJ pathway 1
LNA	Locked Nucleic Acid
lncRNA	long non-coding RNA
LUC7L	Putative RNA-binding protein Luc7-like 1
MALAT1	Metastasis Associated Lung Adenocarcinoma Transcript 1
MFSD10	Major Facilitator Superfamily Domain-Containing Protein 10
min(s)	minute(s)
miRNA	microRNA
MNX1	Motor Neuron And Pancreas Homeobox 1
mRNA	messenger RNA
NCI	National Cancer Institute
ncRNA	Non-coding RNA
NGS	Next-generation sequencing
NHEJ	Non-Homologous End Joining Repair
NOP14	Nucleolar Protein 14
NORAD	noncoding RNA activated by DNA damage
nt	nucleotides
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PAM	Protospacer Adjacent Motif
PANDAR	p21 associated ncRNA DNA damage activated RNA
PARP1	Poly (ADP-ribose) polymerase 1
PARTICLE	Promoter of MAT2A-antisense radiation-induced circulating lncRNA
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PINT	p53 induced noncoding transcript
Pol II	Polymerase II
PRC2	Polycomb Repressive Complex 2
RACE	Rapid amplification of cDNA ends
RB1	RB Transcriptional Corepressor 1
RIPA	Radioimmunioprecipitation buffer assay
RME2	Regulator of Meiosis
RNA	Ribonucleic Acid
RNAi	RNA-interference

ROR	Regulator Of Reprogramming
RPMI	Roswell Park Memorial Institute Medium
rRNA	ribosomal RNA
RT	room temperature
RT-qPCR	Reverse Transcription - quantitative PCR
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecylsulfate-PAGE
sec	second(s)
SEM	Standard Error of Mean
sgRNA	single guide RNA
SID	mSin interacting domain
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
ss	single-stranded
TBS	tris-buffered saline
TBS-T	tris-buffered saline and Tween 20
TERC	Telomerase RNA component
TP53	tumor protein p53
tRNA	transfer RNA
TSS	Transcriptional start site
WB	Western Blot
WRAP53	WD Repeat-Containing Antisense To TP53
XIST	X-Inactive Specific Transcript
Zeb2	Zinc finger E-box-binding homeobox 2
Zeb2NAT	Zeb2 Natural Antisense Transcript
ZFN	Zinc Finger Nuclease

1. Introduction

1.1. Long Non-Coding RNAs

1.1.1. Discovery and Definition

In 1956, in an early draft of his later published article "Ideas on protein synthesis", Francis Crick proposed that genetic information flows from DNA to RNA which in turn serves as a template for protein synthesis[1]. This was later called the central dogma of molecular biology. Five years later, in 1961, Jacob & Monod showed that protein synthesis from DNA indeed required an intermediate RNA molecule which they termed as messenger RNA or simply mRNA. Although central to the dogma, RNA was proposed to be merely a messenger for transferring the genetic information stored in the DNA to proteins – which were thought to be the major drivers of all the cellular functions.

At this point, it was also becoming clear that not all RNA molecules code for proteins. Early work done in Zamecnik lab indicated that an RNA molecule was involved in protein synthesis as an intermediate carrier of amino acids. Soon after this discovery, several other such molecules were discovered and it became apparent that these molecules constitute an entire class of RNAs and were named as transfer RNAs (tRNAs)[2]. These were the RNAs with non-protein-coding function. Around the same time, a link between rRNAs and ribosomes had been established[3]. Following the discovery of these structural RNAs, several other small RNAs were discovered. Small nuclear RNAs (snRNAs) were found to be involved in RNA splicing[4, 5]. Small nucleolar RNAs (snoRNAs) were found to be involved in rRNA processing[6]. With the discovery of ribozymes, it became apparent that RNA had additional roles outside protein synthesis[7, 8]. Thereafter, advancements in sequencing methods revealed several small (e.g. microRNAs) as well as long regulatory RNAs (e.g. *XIST*) with no protein-coding potential. miRNAs were found to be 20-22 nucleotide (nt) long RNA molecules which could bind their complementary mRNAs and negatively regulate their expression. *XIST* was identified to be a 15 kilobase (kb) long RNA arising from the X chromosome inactivation center and to coordinate X chromosome inactivation[9]. Next-generation sequencing (NGS) technology revealed that the majority of the genome gave rise to non-protein-coding or simply non-coding RNAs (ncRNAs)[10]. These were arbitrarily divided into two groups based on their lengths: small ncRNAs (less than 200 nt) and the rest were called long ncRNAs (lncRNAs).

1.1.2. Classification

LncRNAs can be classified on the basis of several attributes, for example, their length, their genomic location, and context, their association with genomic elements of known functions, their mechanism of function and so forth[11, 12]. The most common method of lncRNA classification is based on their genomic location in context to previously annotated protein-coding genes. lncRNA loci are widely distributed throughout the mammalian genome and they can be transcribed either from intergenic regions (such transcripts are called long intergenic

RNAs or lincRNAs) or from intragenic regions overlapping with other protein-coding genes (**Figure 1**). Nearly 60% of all lncRNAs fall into the category of long Intergenic RNAs[13]. lincRNAs can arise from independent promoters as well as divergently from bidirectional promoters which they share with other protein-coding genes (**Figure 1A**). Intragenic lncRNAs can be further classified as sense or antisense, depending on the orientation of the lncRNA with respect to its neighboring gene with which they overlap. They can arise either from external promoters that lie outside the gene that they overlap or from internal promoters that lie inside the gene that they overlap (**Figure 1B**). Antisense lncRNAs represent the second largest class and account for 30% of the total annotated lncRNAs[13].

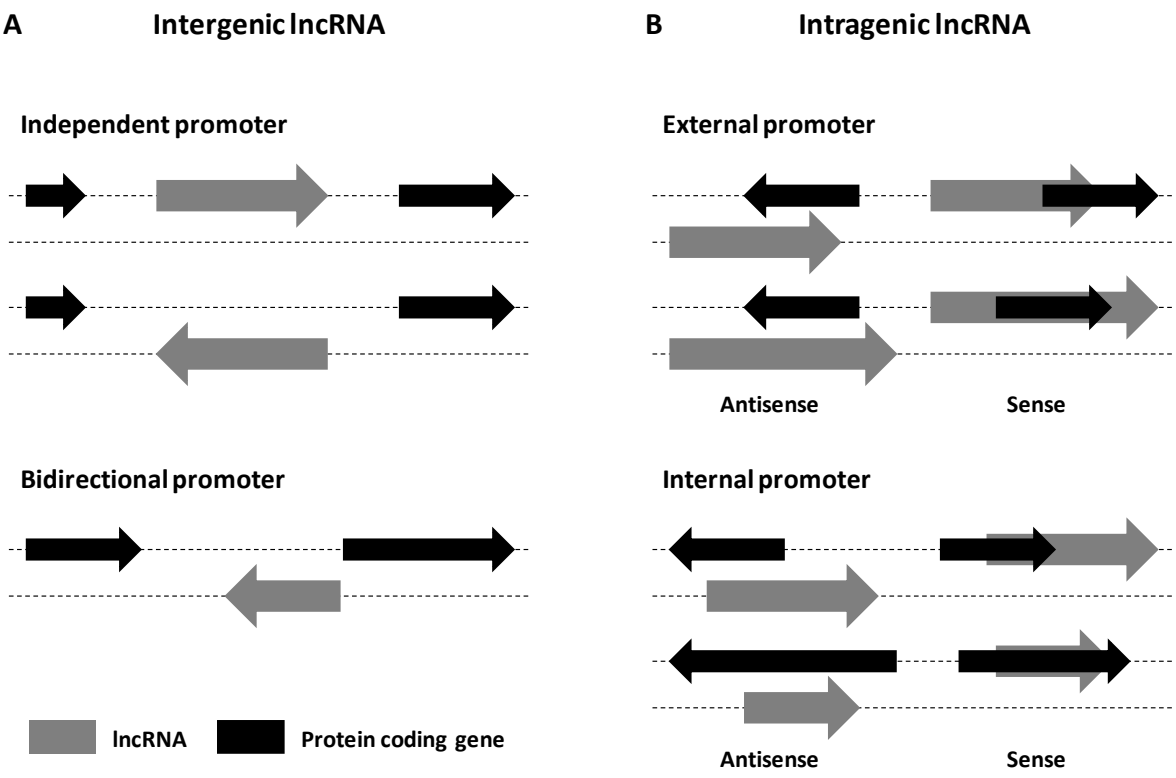


Figure 1: lncRNA classification based on genomic location
Schematic representation of lncRNA loci (in gray) distribution across the human genome and their classification based on their location in context to previously annotated protein-coding genes (in black). (A) Intergenic lncRNAs do not overlap with any protein-coding gene and can be transcribed from either independent promoters or bidirectional promoters. (B) Intragenic lncRNAs can be further divided into Antisense or Sense lncRNAs depending on their orientation with respect to their overlapping protein-coding genes and can be transcribed from either an external or an internal promoter.

1.1.3. Function and mechanism of action

LncRNAs are a functionally diverse group of molecules and play important roles in several cellular processes such as cell cycle regulation, cell migration, differentiation etc. and thus contribute to normal development and diseases, such as cancer[14-17]. Through their ability to interact with DNA, RNA, and protein, they can regulate gene expression at transcriptional as well as post-transcriptional levels.

LncRNAs can regulate the transcription of their target genes either by binding and recruiting chromatin modifying complexes or by the simple act of their transcription alone[18-20]. Moreover, transcriptional regulation by lncRNAs can be either *in cis* or *in trans*. A *cis*-acting lncRNA regulates the transcription of a gene residing on the same chromosome from which they are transcribed. On the other hand, a *trans*-acting lncRNA regulates the transcription of the gene(s) located on another chromosome(s)[20]. An example of a *cis*-acting lncRNA is *ANRIL* (also known as *p15AS* or *CDKN2B-AS*), which is overexpressed in cancer. *ANRIL* is transcribed antisense to the tumor suppressor locus of *INK4a/ARF/INK4b* (*p16/p14/p15*) where it binds and recruits the PRC2 complex to induce histone H3 lysine 27 (H3K27) methylation, thereby repressing transcription from this locus[21-23](**Figure 2A**). Another example of *in cis*-transcriptional regulation by a lncRNA is that of the hemoglobin α1 gene (*HBA1*) repression by *LUC7L*. In a class of α-thalassemia patients, an aberrant read-through transcription of *LUC7L* across its antisense protein-coding gene *HBA1* results in methylation of its promoter CpG island, thereby silencing *HBA1* expression and causing the disease[24] (**Figure 2B**). The metastasis-associated lung adenocarcinoma transcript 1 (*MALAT1*) is a marker for lung cancer metastasis and a regulator cell migration and cell cycle[25-27]. It is localized in the nuclear speckles where it binds the unmethylated Polycomb 2 (Pc2), a component of the PRC1 complex and regulates the transcription of Pc2-bound genes following serum restimulation. Apart from recruiting histone modifying complexes or inducing DNA methylation, the act of lncRNA transcription can also regulate the expression of their overlapping gene. Such an effect is called transcriptional interference and is mediated by direct collision of RNA polymerases (**Figure 2C**). An example of transcriptional interference is the regulation of the *IME4* gene by its antisense transcript *RME2* (Regulator of Meiosis). The expression of *RME2* in haploid *Saccharomyces cerevisiae* cells prevents *IME4* expression[28, 29]. In diploid cells, however, the transcription of *RME2* is repressed, thereby allowing IME4 to be induced during meiosis. *HOTAIR* (HOX transcript antisense RNA) is an example of a *trans*-acting lncRNA, which is an antisense transcript to the *HOXC* locus. *HOTAIR* silences the *HOXD* locus *in trans* through its interaction with the PRC2 complex[30] (**Figure 2D**).

Post-transcriptionally, lncRNAs can regulate gene expression in several ways[31]. LncRNA *BACE1-AS* plays an important role in Alzheimer's disease where it base pairs with its antisense mRNA of the β-secretase-1 (*BACE1*), masking a microRNA (miRNA) binding site, thereby preventing it from degradation[32, 33] (**Figure 2E**). The increased β-secretase-1 levels contribute to the pathophysiology of Alzheimer's disease. The mouse *Uchl1AS* lncRNA base pairs with the 5' end of its antisense *Uchl1* mRNA and upregulates its translation through an embedded SINEB2 repeat element[34]. LncRNAs can also regulate gene expression by acting as miRNA sponges[35]. Such lncRNAs are called competing endogenous RNAs (ceRNAs). Linc-MD1 regulates the muscle differentiation by sponging miR-133 and miR-135 to regulate the expression of *MAML1* and *MEF2C*[36].LncRNAs can also bind proteins to regulate their stability.

LncRNAs can also play a role outside gene expression regulation. An example is the lncRNA Telomerase RNA component (TERC) which is involved in the maintenance of the telomere by serving as a scaffold for telomerase components and template for repeat addition[37].

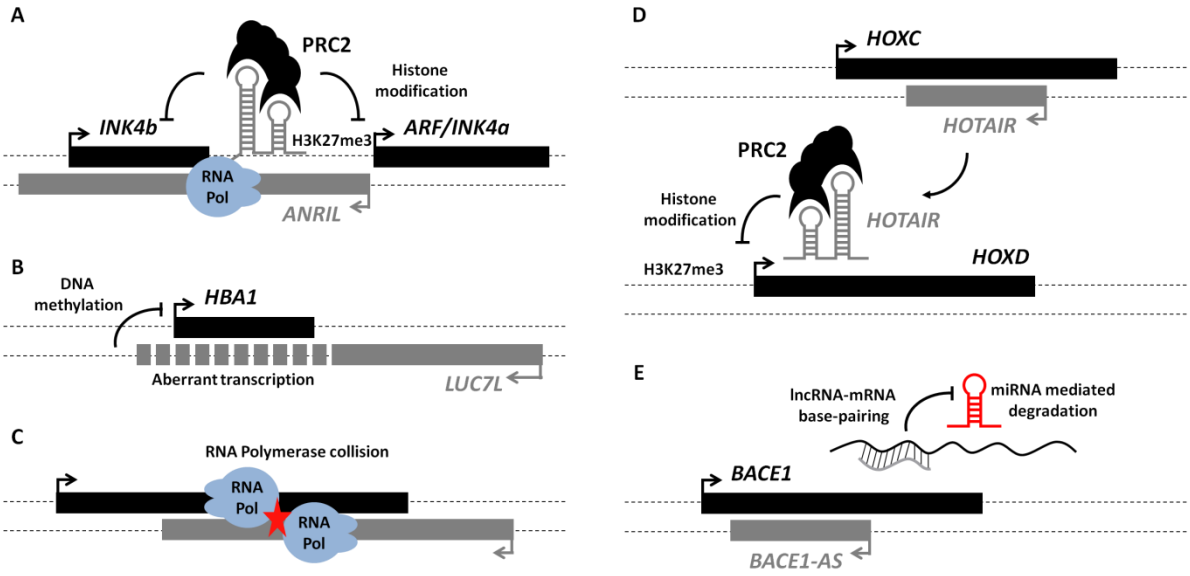


Figure 2: lncRNA mechanisms of action

A-C: Examples of LncRNAs regulating transcription *in cis* (A) *ANRIL* nascent transcript binds and recruits PRC2 to the *INK4* locus to induce H3K27 trimethylation, thereby silencing this locus. (B) Aberrant transcription of the *LUC7L* lncRNA over the *HBA1* promoter results in promoter DNA methylation, thereby silencing *HBA1* expression. (C) LncRNAs can regulate the expression of their overlapping genes by transcriptional interference mediated by a direct collision between the RNA polymerases of the two genes.

D: *HOTAIR* binds and recruits PRC2 to the HOXD locus to regulate its transcription in trans.

E: *BACE1-AS* lncRNA base pairs with its sense mRNA *BACE1* to mask the miR-485-5p binding site, thereby preventing miRNA-mediated degradation.

Figure adapted from Pelechano et al., Nature Reviews Genetics 2013[19]

1.1.4. LncRNAs in DNA Damage Response Pathway

The DNA damage response (DDR) pathway is a coordinated cellular response to genotoxic stress that could potentially lead to genomic instability, altered protein production or loss of genetic material. Mutations and strand breaks can activate DNA damage checkpoints to arrest the cell cycle to allow for the repair of DNA lesions. If the DNA repair fails or the DNA damage is too massive, apoptosis is triggered to ensure the removal of cells carrying damage and prevention of mutation accumulation[38].

In the past few years, several lncRNAs have emerged as major regulators of the DDR pathway[16, 39]. The first example of a lncRNA involved in the DDR pathway was that of the ncRNA-*CCND1*[40]. ncRNA-*CCND1* is induced from the promoter upstream region of *CCND1* upon DNA damage caused by ionizing radiation. It recruits the RNA binding protein TLS to the *CCND1* promoter, which in turn, inhibits the CREB-binding protein (CBP) and p300 histone acetyltransferase activities, thereby preventing *CCND1* expression[40, 41]. *ANRIL* is induced upon DNA damage via the ATM-E2F pathway and suppresses the expression of its overlapping protein-coding gene *INK4b* (p15) to regulate the cell cycle[42]. A non-coding isoform of *WRAP53* is transcribed from the first exon of *TP53* in antisense manner and regulates p53 mRNA stability as well as induction upon DNA damage by binding to the 5' untranslated region of the p53 mRNA[43]. *PARTICLE* (promoter of *MAT2A*-antisense radiation-induced

circulating lncRNA) is induced from the *MAT2A* promoter upon DNA damage. It forms a DNA-lncRNA triplex with the *MAT2A* promoter and recruits G9a and PRC2 to silence *MAT2A* transcription[44].

Not all lncRNAs involved in DNA damage are *cis*-acting. Many lncRNAs act *in trans* to modulate the DDR pathway. *DDSR1* is induced in an ATM-NF-κB pathway-dependent manner and interacts with BRCA1 to enhance DNA repair via homologous recombination[45]. *LINP1* serves as a scaffold for Ku80 and DNA-PKc to increase NHEJ-mediated repair efficiency[46]. *NORAD* serves as a molecular decoy for PUMILIO to regulate genome stability[47]. *ERIC* is an E2F-regulated lncRNA which controls DNA damage-induced apoptosis[48].

TP53 (p53) is a major transcription factor involved in the DDR. DNA damage stabilizes p53 which then activates the transcription of several protein-coding genes to coordinate the cellular response to DNA damage[49, 50]. A growing body of literature indicates that p53 also regulates expression of numerous lncRNAs and they can, in turn, regulate the p53 signaling[51-54]. *LincRNA-p21*, *PANDAR* (p21 associated ncRNA DNA damage activated RNA, also known as *PANDA*) and *DINO* (damage induced noncoding) are induced from promoter upstream regions of the *CDKN1A* (p21) gene upon DNA damage in a p53-dependent manner. While *lincRNA-p21* and *PANDAR* modulate apoptosis by regulating the expression of several p53 target genes[55, 56], *DINO* interacts with p53 through a stem loop motif and stabilizes it, thereby activating p53 target genes[54, 57]. *LincRNA-RoR* is induced upon DNA damage in a p53-dependent manner and binds hnRNP I to repress p53 mRNA translation, thereby constituting a negative feedback loop[58]. *PINT* is induced in a p53-dependent manner upon DNA damage and coordinates PRC2-mediated gene silencing of several p53 target genes[59] (**Figure 3**).

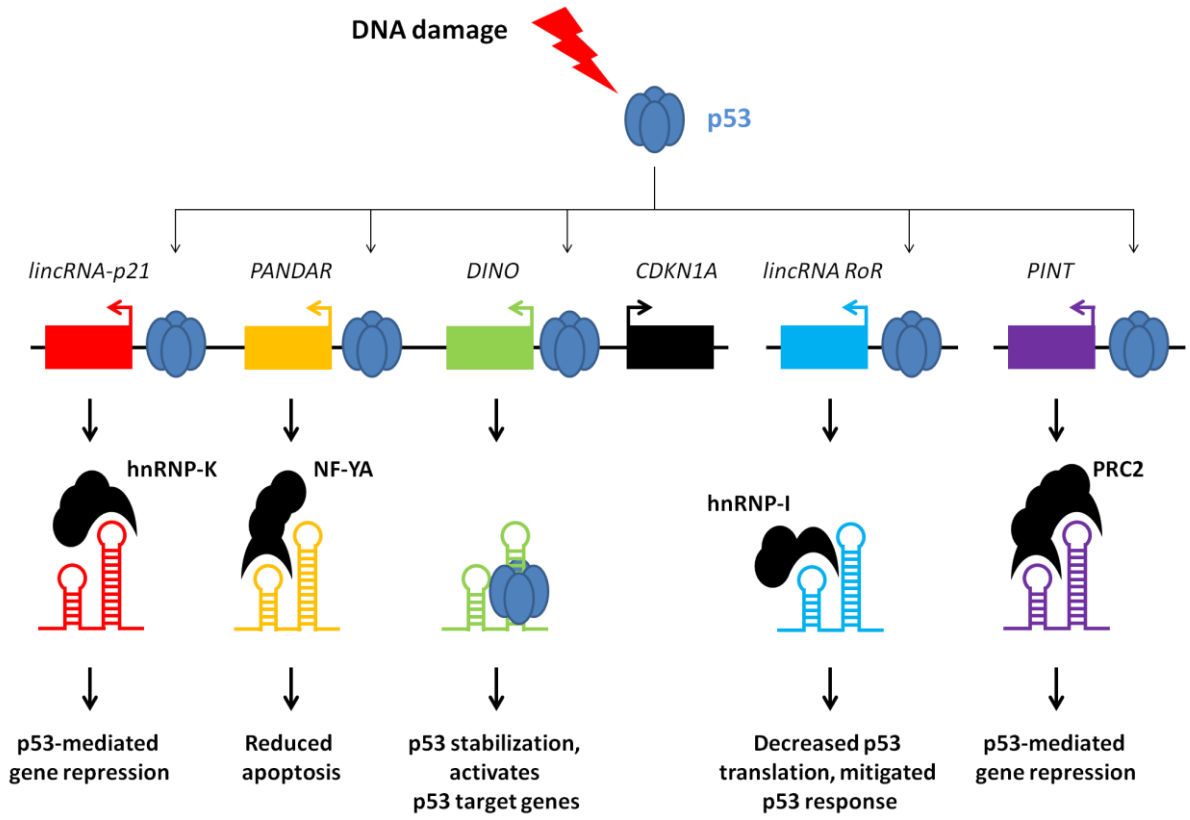


Figure 3: LncRNAs in the p53 regulatory network

p53 is stabilized upon DNA damage and activates transcription of numerous lncRNAs. *LncRNA-21*, *PANDAR*, and *DINO* are all transcribed from the promoter upstream region of the p53 target gene *CDKN1A*. *LincRNA-p21* binds to hnRNP-K and this complex represses the transcription of several p53 target genes. *PANDAR* binds to and sequesters the pro-apoptotic transcription factor NF-YA thereby resulting in reduced apoptosis. *DINO* binds and stabilizes p53, thereby leading to activation of the p53 target gene. *LncRNA-RoR* binds and sequesters hnRNP-I which otherwise binds to the p53 mRNA and enhances its translation, thereby leading to decreased p53 mRNA translation and mitigation of p53 response. *PINT* binds to PRC2 complex to silence the expression of several pro-apoptotic and anti-proliferative p53 target genes.

In summary, these and several other examples prove the importance of lncRNAs in the DDR pathway[44, 60-67].

1.2. CRISPRs

1.2.1. Introduction

CRISPRs (Clustered regularly interspaced palindromic repeats) were first discovered in archael[68] and later in bacterial[69, 70] genomes as arrays of near perfect palindromic repeats separated by spacer regions derived from invading phages and plasmids. These serve as a memory of previous infections and together with CRISPR-associated (Cas) proteins constitute a prokaryotic adaptive immune system[70-73]. The CRISPR/Cas system incorporates DNA fragments from invading plasmids or phages into the host CRISPR locus, which is then transcribed and processed into crRNAs, which in turn guide the Cas proteins to recognize and cleave the invading genome[73]. CRISPR systems can be classified into several types depending on their components and mechanisms of action[74].

The type II CRISPR system from *S. pyogenes* is the most widely studied and was the first system to be adapted for mammalian genome editing[75, 76]. In its simplest form, this system consists of two components: the Cas9 nuclease enzyme and a single guide RNA (sgRNA)[76, 77]. The first 20 nucleotides (nt) on the 5' end of the sgRNA can hybridize with complementary DNA sequence and thus determine targeting specificity of this system. The rest of the sgRNA sequence binds the Cas9 nuclease and thus help to recruit the Cas9-sgRNA complex to its target, where Cas9, using its two nuclease domains (HNH and RuVc), generates a double-stranded break (DSB) within the base paired region. Since the target specificity is determined by only a 20 nt RNA sequence, and the nuclease remains the same, this system can be readily programmed to target any sequence of choice provided it is followed by a 3 bp PAM sequence (NGG, where N is any of the four bases and G is Guanine)(**Figure 4A**). Cas9-induced DSB can be repaired by one of the two intrinsic cellular repair mechanisms, namely, non-homologous end joining repair (NHEJ)[78] or homology-directed repair (HDR)[79]. NHEJ is an error-prone mechanism and causes random insertions or deletions (indels) of base pairs around the cut site in the process of joining the break (**Figure 4B**). HDR, on the other hand, utilizes the homologous chromosome as a template for error-free repair of the break via homologous recombination (**Figure 4C**). This ability of Cas9 to bind and cleave double-stranded DNA (dsDNA) in an easily programmable and a sequence-specific manner makes it a very powerful tool for genome engineering.

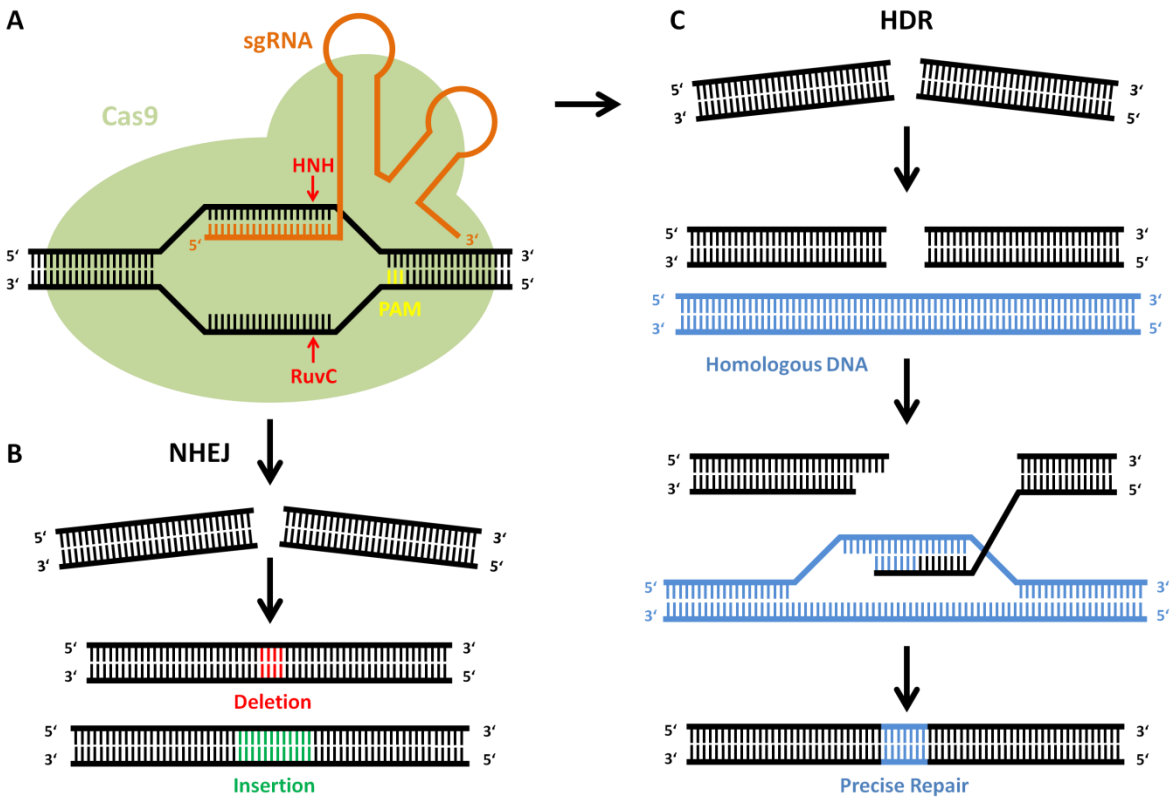


Figure 4: Cas9-mediated genome editing
(A) The Cas9-sgRNA complex recognizes target DNA sequences through a 20 nt region on the 5' end of the sgRNA, which base pairs with its complementary sequence in the target genome. If the PAM sequence (NGG) is present adjacent to the target sequence, Cas9 generates a DSB using its two nuclease domains (HNH and RuvC). Such a DSB can be repaired using either (B) the NHEJ pathway, which results in random insertions or deletions around the cut site or (C) the HDR pathway, which uses the homologous DNA as a template to carry out the precise repair.

1.2.2. Applications for lncRNA research

Cas9 is widely used to generate DSBs in the open reading frame (ORF) of a protein-coding gene, resulting in indel formation via the NHEJ repair pathway. This, in turn, induces frameshift mutations resulting in an effective knockout of the targeted protein-coding gene[76] (**Figure 5A**). This method is in general not applicable for knocking out lncRNAs as unlike their protein-coding counterparts, they lack an ORF, and a small indel mutation is unlikely to change its structure or expression (**Figure 5B**). It can be used to generate functional knockouts in the special case when the sequences of the lncRNA transcript responsible for carrying out transcript-dependent molecular functions are already known or well characterized. However, that is not the case for most of the lncRNAs as a vast majority of lncRNAs still remain uncharacterized. Moreover, predicting the active parts of a lncRNA is currently not possible as they can exert their molecular functions through diverse mechanisms. Also, for lncRNAs which exert their phenotype by the act of their transcription, a Cas9-induced indel mutation is highly unlikely to affect its transcription and thus its function, unless Cas9 is targeted to a genomic element controlling the lncRNA transcription. Given that lncRNAs often arise from regulatory elements which they share with other genes (e.g. lncRNAs arising from enhancer regions), this approach can potentially lead to the detection of false phenotypes.

Cas9 can also be used to induce small or large genomic deletions by simultaneously generating multiple DSBs[80, 81]. This approach can be used to generate lncRNA knockouts by either deleting their promoters[82-87] or the entire lncRNA genes[84, 88] (**Figure 5C**). In any case, genomic excisions, howsoever small, can lead to the deletion of regulatory DNA elements, which might affect transcription of other genes and give rise to phenotypes which are originally not attributable to the lncRNA[84, 89-91]. Since it is impossible to consider the position of all potential genomic regulatory elements, as they remain largely uncharacterized, any deletions should be preceded by a detailed examination of the genomic region of interest.

Alternatively, the HDR mechanism can be exploited to engineer precise changes in the genome by supplying a donor DNA template which harbors the desired mutation(s) and has homology to the region around the cut site of the target DNA. In this manner, the Cas9-induced DSBs can be used for gene corrections or to knockin DNA elements for gene overexpression, knockout as well as tagging[92, 93]. This approach can be used to knockdown lncRNA expression using Cas9 by homology-directed knockin of a transcriptional termination signal or RNA destabilizing elements immediately downstream of the TSS of the lncRNA gene[84, 90] (**Figure 5D**). This method was initially proposed and utilized by the Diederichs lab to knockdown *MALAT1*, using zinc finger nucleases (ZFNs) instead of Cas9[94]. Although superior to the Cas9 mediated deletion approach, homology-directed knockin can also potentially lead to disruption of any underlying regulatory elements controlling transcription of other genes (e.g. a transcription factor binding site). Moreover, HDR being a less efficient process as compared to NHEJ, this approach can be extremely labor intensive as it involves clonal expansion and screening for positive clones.

Mutation of both the nuclease domains of Cas9 results in a catalytically dead Cas9 (dCas9), which lacks the endonuclease activity but still possesses its sgRNA-dependent DNA-binding activity. dCas9 can be fused to effector domains and co-expressed with sgRNAs to generate custom transcription factors. dCas9, when recruited to the vicinity of the promoter of a gene, can interfere with either transcription initiation or elongation, thereby resulting in reduced transcription[95, 96]. dCas9 fused to the transcriptional repressor domains like KRAB (Krüppel-associated box domain of ZNF10)[97] or SID (mSin interacting domain of Mad1)[98, 99] results in an even more potent inhibitor of transcription (CRISPR interference or CRISPRi)[96, 100] (**Figure 5E**). On the other hand, dCas9, when fused to transcriptional activator domains like VP160, p65, or Rta, can induce target gene expression from the endogenous promoter of their target genes (CRISPR activation or CRISPRa)[101-103] (**Figure 5F**). CRISPRi/CRISPRa with dCas9-based transcription factors have been used to perform genome-wide loss-of-function as well as gain-of-function screens for protein-coding or lncRNA genes[104-108]. CRISPRi/CRISPRa systems have several advantages for lncRNA research: (1) Modulation of transcription from the endogenous promoter ensures that the *in cis* functions of the lncRNA can also be observed, which would otherwise be undetectable by RNA interference (RNAi)-mediated knockdown or plasmid-based overexpression. (2) lncRNA genes often give rise to several splice isoforms, and modulation of transcription from the endogenous promoter would lead to modulation of all splice variants in their natural ratios. (3) Unlike the Cas9-mediated deletion or insertion approaches, these systems do not perturb any underlying genomic regulatory elements. (4) Since no clonal selection is involved, a CRISPRi/CRISPRa mediated loss-of-function or

gain-of-function model can be generated in a much shorter time span as compared to Cas9-mediated deletion or insertion approaches.

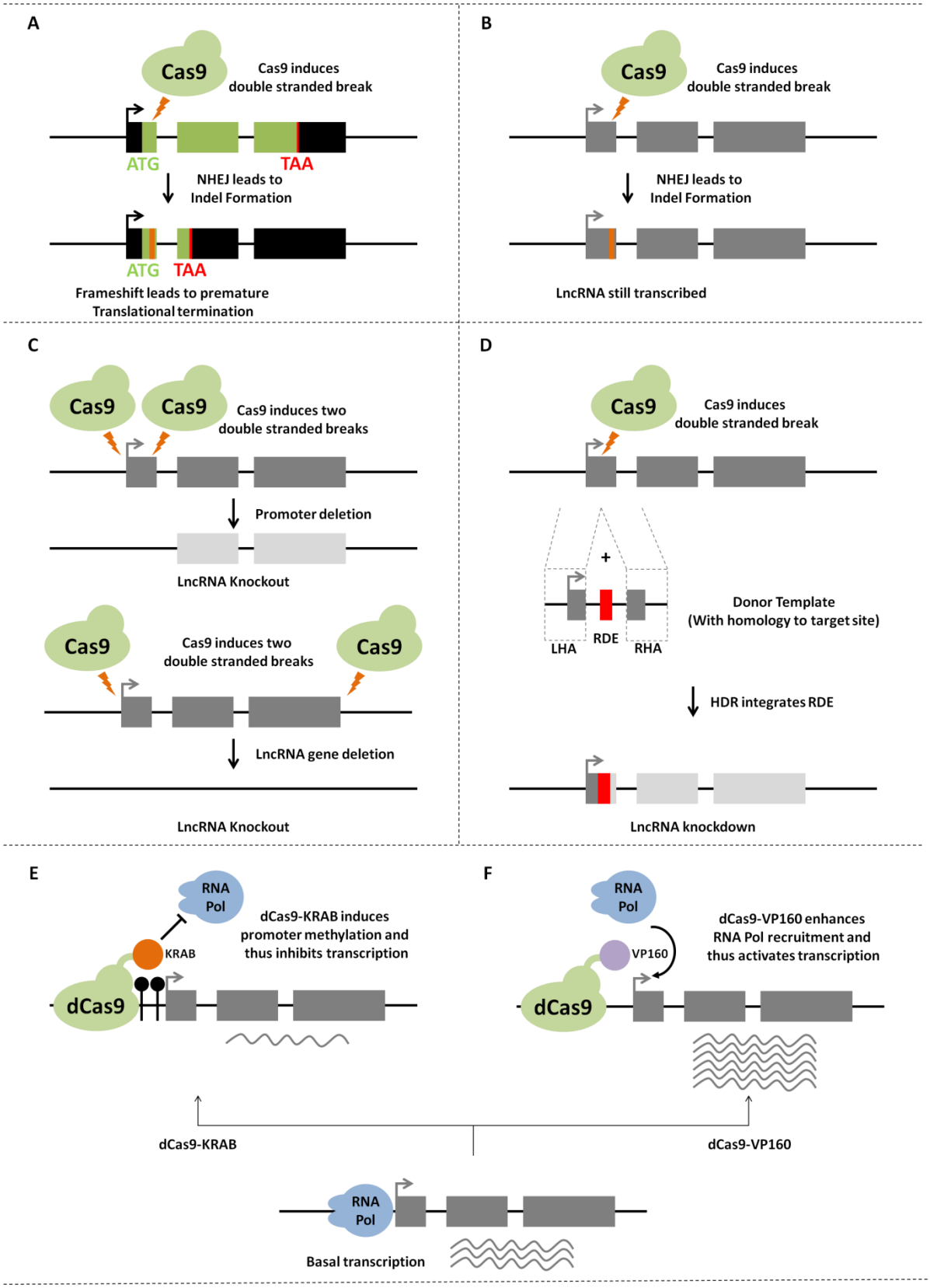


Figure 5: CRISPR/Cas9 applications for lncRNA research

A,B: A Cas9-induced DSB, when repaired using NHEJ, can lead to insertion or deletion of a few base pairs (indel mutations). Such mutations in the ORF of a (A) protein-coding gene often lead to a frameshift and premature translational termination, thereby leading to knockout of the full-length protein. However, for (B) lncRNAs, they do not affect their function or expression.
C,D: Simultaneous induction of two DSBs using Cas9 can lead to excision of the DNA fragment between them. This can be used to delete the (C) promoter or (D) the entire gene body of a lncRNA leading to its knockout.
E,F: dCas9 can be used to modulate transcription of a lncRNA from its endogenous promoter. dCas9 can be fused to either (E) a KRAB domain which leads to promoter methylation thereby inhibiting the transcription of the lncRNA, or (F) to a VP16 domain or a derivative, which interacts with multiple components of the transcription machinery and enhances RNA Pol-II recruitment, thereby activating transcription.

Overall, CRISPR/Cas9 systems offer a very powerful and easy to use lncRNA manipulation toolbox.

1.3. Previous results and basis of the dissertation

To identify novel lncRNAs involved in the DDR, Evgenij Fiškin treated A549 and HepG2 cells with the DNA-damaging agents Etoposide, Cisplatin, and Bleomycin or vehicle control DMSO and performed a microarray expression analysis together with Dr. Maria Polycarpou-Schwarz to identify differentially expressed lncRNAs. LncRNA *NOP14-AS1* was identified as the most promising candidate as it was induced in both cell lines upon treatment with all the three DNA-damaging drugs (**Figure 6A**). Julia Neugebauer found that inversely to *NOP14-AS1*, the expression of its neighboring antisense gene *NOP14* was downregulated upon DNA damage (**Figure 6B**).

The preliminary data on regulation of *NOP14-AS1* upon DNA damage and its negative correlation with *NOP14* indicated that *NOP14-AS1* lncRNA or its transcription could be functionally linked to *NOP14* mRNA expression or vice versa as it had been proposed for several other sense-antisense pairs of lncRNA and mRNA genes[19, 20, 109].

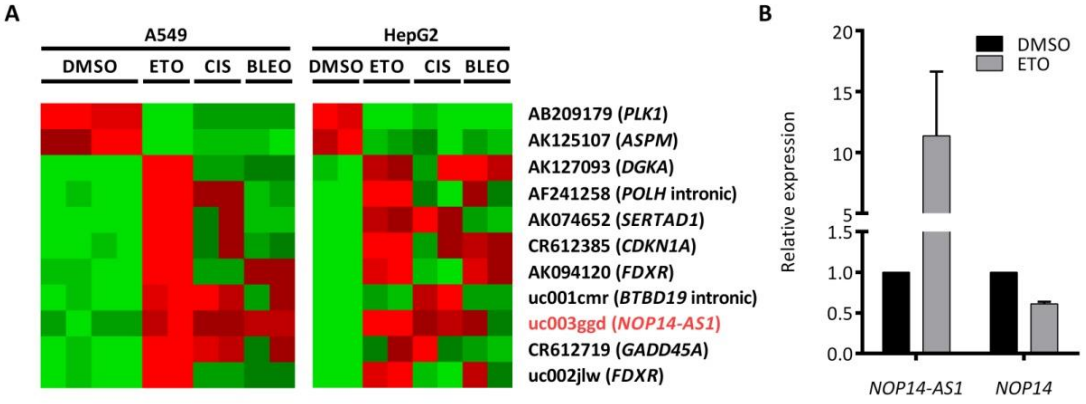


Figure 6: Identification of *NOP14-AS1* as a DNA damage-inducible lncRNA
A: Microarray analysis heat map of lncRNAs and mRNAs differentially expressed in A549 (left panel) and HepG2 (right panel) cells treated with 50 µM Etoposide (ETO) / 50 µM Cisplatin (CIS) / 20 µM Bleomycin (BLEO) or vehicle control DMSO for 8 hours.
Treatments performed by Evgenij Fiškin, microarray hybridization performed by Dr. Maria Polycarpou-Schwarz, microarray analysis performed by Prof. Dr. Sven Diederichs. Data also used for a manuscript under review for publication: Goyal et al.

B: HepG2 cells were treated with 50 µM Etoposide (ETO) or vehicle control DMSO for 8 hours. RT-qPCR results for *NOP14-AS1* and *NOP14* normalized to *Cyclophilin A* and DMSO controls. Error bars represent SD (n=2).
Experiment performed by Julia Neugebauer.

1.4. Aim of this study

In this study, the regulation of *NOP14-AS1*, as well as its neighboring genes *NOP14* and *MFSD10* in DNA damage should be further characterized. Particular aims included

1) the reproduction of the preliminary observation of an inverse co-regulation between *NOP14-AS1* and *NOP14*.

2) the analysis of the regulatory mechanism with a focus on TP53 (p53) since several DNA damage induced lncRNAs are p53-dependent.

3) establishing and testing loss-of-function and gain-of-function-models for *NOP14-AS1* and *NOP14*. Since antisense lncRNAs often regulate the expression of their neighboring gene *in cis*, these models should be capable of mimicking such a regulatory mechanism. These models serve to study the impact and mechanism of *NOP14-AS1* on the *NOP14* expression or vice versa.

2. Materials and Methods

2.1. Materials

2.1.1. Chemicals and enzymes

Table 1: List of chemicals and enzymes used in this study

Name	Sequence
30% Acrylamide/Bis Solution, 37.5:1	Carl Roth
Acetic Acid	Merck Biosciences
Acetone	Sigma-Aldrich
Agarose, LE	Biozym
Ammonium Persulfate Electrophoresis Reagent	Sigma-Aldrich
Ampicillin	Sigma-Aldrich
Beta-Mercaptoethanol	Sigma-Aldrich
Bicinchoninic Acid (BCA)	Sigma-Aldrich
Blasticidin	Invivogen
Bleomycin	Merck Millipore
BSA	Roche
Calcium Chloride	Carl Roth
Chloroform	Carl Roth
Cisplatin	Merck Millipore
Copper(II) sulfate solutions	Sigma-Aldrich
Dharmafect	GE Dharmacon
Dithiothreitol (DTT)	Roche
DMEM	Sigma-Aldrich
DMSO	AppliChem
DNase I	Roche
dNTP mix	Thermo Fisher Scientific
Doxorubicin	Merck Millipore
Dream Taq DNA-Polymerase	Thermo Fisher Scientific
EDTA	Gerbu
EDTA-free protease inhibitor mixture tablets	Roche
Ethanol, absolute	Sigma-Aldrich
Ethidium Bromide	Carl Roth
Etoposide	Cayman Chemical
Fast Digest AgeI	Thermo Fisher Scientific
Fast Digest BamHI	Thermo Fisher Scientific
Fast Digest BsmBI	Thermo Fisher Scientific
Fast Digest BsrGI	Thermo Fisher Scientific
Fast Digest KpnI	Thermo Fisher Scientific
Fast Digest NheI	Thermo Fisher Scientific
Fast Digest XbaI	Thermo Fisher Scientific
FastDigest enzymes	Thermo Fisher Scientific
FBS	Thermo Fisher Scientific
Formaldehyde	Thermo Fisher Scientific
GeneRuler 100bp Plus DNA Ladder	Thermo Fisher Scientific
GeneRuler 1kb DNA Ladder	Thermo Fisher Scientific
Isopropanol	Sigma-Aldrich
L-glutamine	Sigma-Aldrich
Lipofectamine 2000	Thermo Fisher Scientific
Lipofectamine RNAiMAX	Thermo Fisher Scientific
Methanol	Sigma-Aldrich

Table 1: List of chemicals and enzymes used in this study

Name	Sequence
NP-40	Sigma-Aldrich
Nutlin-3	Sigma-Aldrich
Opti-MEM	Thermo Fisher Scientific
PageRuler Prestained Protein Ladder	Thermo Fisher Scientific
Polybrene	Sigma-Aldrich
Powdered LB	Thermo Fisher Scientific
Powdered Skimmed milk	Fluka
Power SybrGreen Master Mix 2X	Thermo Fisher Scientific
Proteinase K	Sigma-Aldrich
Puromycin	Thermo Fisher Scientific
Q5 High-Fidelity DNA Polymerase	New England Biolabs
Random Hexamer Primers	Thermo Fisher Scientific
RevertAid Reverse Transcriptase	Thermo Fisher Scientific
RiboLock RNase Inhibitor	Thermo Fisher Scientific
Roti-Aqua-P/C/I for RNA extraction	Carl Roth
RPMI 1640	Sigma-Aldrich
SDS	Carl Roth
Sodium Acetate	Carl Roth
Sodium Chloride	Sigma-Aldrich
Sodium Deoxycholate	AppliChem
Super Signal West Femto Chemiluminescent Substrate	Thermo Fisher Scientific
Super Signal West Pico Chemiluminescent Substrate	Thermo Fisher Scientific
T4 DNA ligase	Thermo Scientific
TEMED	Carl Roth
TRI Reagent	Sigma-Aldrich
TritonX-100	Gerbu
Trizma base	Sigma-Aldrich
Trypsin-EDTA (0.05%, 0.25%)	Thermo Fisher Scientific
Tween-20	MP Biomedicals
UltraPure DNase/RNase-Free Distilled Water	Thermo Fisher Scientific

2.1.2. Kits and disposables

Table 2: List of kits and disposables used in this study

Name	Company
1.5 ml Reaction Tubes	Eppendorf
10 cm dishes, 96 x 20mm	TPP
12-well plates	TPP
15 cm dishes, 146x21 mm	TPP
15 ml Falcon Tubes	TPP
2 ml Cryotubes	Greiner Bio-One
2 ml Reaction Tubes	Eppendorf
24-well plates	TPP
50 ml Falcon Tubes	TPP
6-well plates	TPP
Amersham Hybond N+ nylon membrane	GE Healthcare
Axyprep Plasmid MiniPrep Kit	Axygen
Combitips advanced, 0.2, 0.5, 2.5, 5, 10, and 25 ml	Eppendorf
Filter paper Whatman 3MM	Whatman
Filter Tips 1000 µl	Nerbe Plus
Filter Tips 2, 20 and 200 µl	Neptune
GeneJET Gel Extraction Kit	Thermo Fisher Scientific

Table 2: List of kits and disposables used in this study

Name	Company
Millex-HV 0.45 µm	Merck Millipore
Multichannel pipette tips LiteTouch System LTS 20µl	Rainin
PCR SingleCap 8er-SoftStrips 0.2 ml	Biozym
Poly-L-Lysine coated 6-well plates	Greiner Bio-One
PureLink HiPure Plasmid Maxiprep Kit	Thermo Fisher Scientific
PureLink HiPure Plasmid Midiprep Kit	Thermo Fisher Scientific
QuikChange II Site-Directed Mutagenesis Kit	Agilent Genomics
Serological pipette PS 10 ml	Nerbe Plus
Serological pipette PS 25 ml	Nerbe Plus
Serological pipette PS 5 ml	Nerbe Plus
TOPO TA cloning kit	Thermo Fisher Scientific

2.1.3. Technical equipment

Table 3: Technical equipment used in this study

Name	Company
Applied Biosystems StepOnePlus cycler	Thermo Fisher Scientific
BioPhotometer Plus	Eppendorf
Cell incubator Labotect C200	Labotect
ChemoCam ECL Imager 3.2	Intas
Gilson Pipetman Pipette Set	Gilson
INFORS HT - Ecotron (Bacterial Shaker)	INFORS HAT
INTAS UV gel-imager	INTAS
Isotemp 202 (Waterbath)	Thermo Fisher Scientific
Leica DM IRB microscope	Leica
Metal-Block-Thermostat MBT-250	Kleinfeld
Mini Trans-Blot Electrophoretic Transfer Cell	Bio-Rad
Mini Protean Minigel System	Bio-Rad
Multistep pipette (Multipette stream)	Eppendorf
Multiwell pipette (pipet lite)	Rainin
NanoDrop 1000 Spectrophotometer	Thermo Fisher Scientific
Orbital Shaker DOS-10L	NeoLab
PCR thermal cycler PTC-200	Bio-Rad
PowerPac Basic, Power Supply	Bio-Rad
Sartorius pH Meter PB-11	Sartorius
SterilGARD III Advance cell culture hood	The Baker Company
Trans-Blot SD Semi-Dry Transfer Cell	Bio-Rad
Vortex Mixer	NeoLab

2.2. Methods

2.2.1. Molecular cloning

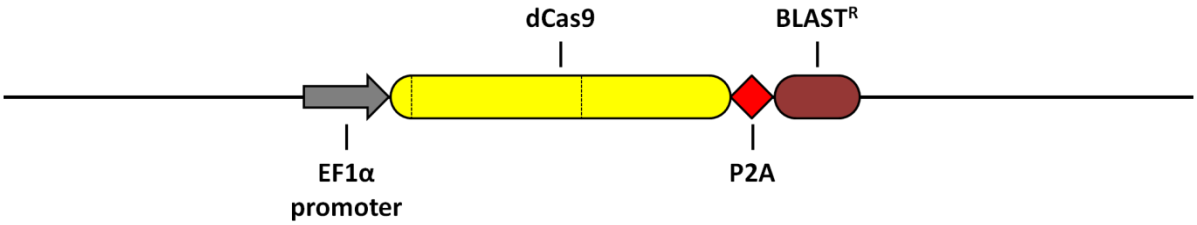
lentidCas9-Blast: The Cas9 coding sequence in the lentiCas9-Blast plasmid[110] (a gift from Prof. Feng Zhang, obtained through addgene, plasmid #52962) was mutated to generate the nuclease-deficient dCas9 (D10A and H841A)[95], using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies) according to manufacturers recommendation. In the first step the Aspartic acid at the 10th position of the Cas9 coding

sequence was mutated to Alanine (D10A), and in the second step, the Histidine at the 841st position of the Cas9 coding sequence was mutated to Alanine (D10A). For each step, a final reaction of 50 µl with 1x *PfuUltra* HF DNA polymerase buffer, 1 µl dNTPs, 125 ng forward and reverse primer each (**Table 4**), 10 ng parental plasmid and 1 µl (2.5 Units) of the *PfuUltra* HF DNA polymerase was prepared. The site-directed mutagenesis PCR was performed as follows:

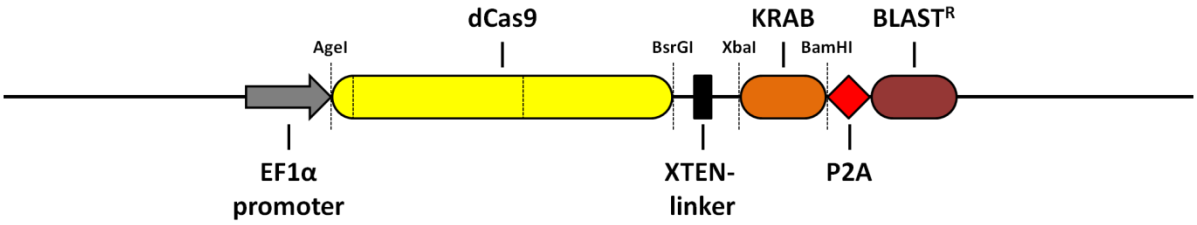
- Initial denaturation at 95°C for 2 min
  - 95°C for 30 sec
  - 60°C for 30 sec
  - 72°C for 15 min

x 18 cycles

The parental plasmid was subsequently digested by adding 1 µl of Dpn I to the reaction mix followed by incubation for 1 h at 37°C. 1 µl of this reaction was transformed into 50 µl chemically competent MachI cells. Following transformation, plasmid DNA was isolated using the Axyprep Plasmid MiniPrep Kit according to the manufacturer's recommendations. Plasmids thus obtained were screened for intended mutation using sanger sequencing of the region of interest. The final plasmid with D10A and H841A mutations was named lentidCas9-Blast.

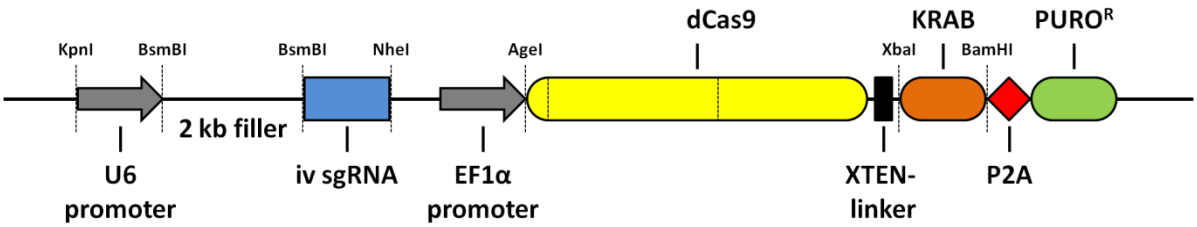


lentidCas9-KRAB-Blast: The cDNA for the repression domain of KRAB (amino acids 11-75)[99] fused with an XTEN-based linker[111] on its N-terminal end was synthesized (GeneArt/ThermoFisher Scientific). This was cloned in frame downstream of dCas9 in the BsrGI and BamHI restriction sites. The resulting plasmid was named lentidCas9-KRAB-Blast.

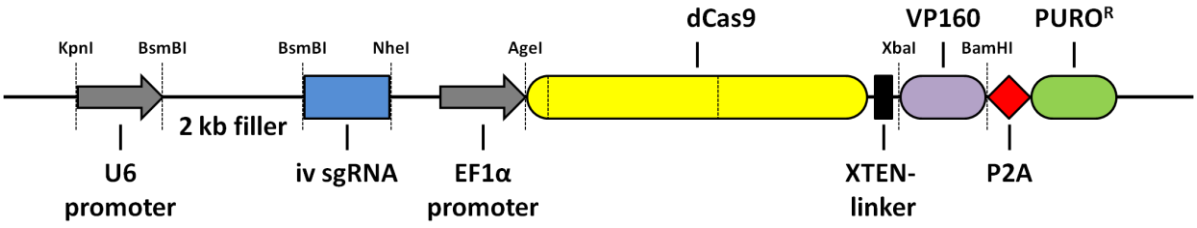


lentidCas9-KRAB-PURO iv sgRNA: The dCas9-KRAB coding sequence was cut out from the lentidCas9-KRAB-Blast vector using FastDigest AgeI and BamHI (ThermoFisher Scientific), cloned into the same sites in lentiCRISPR v2[110] (a gift from Prof. Feng Zhang, obtained through addgene, plasmid #52961) replacing Cas9. An

improved sgRNA scaffold[112] was synthesized (GeneArt) and cloned into the KpnI-NheI sites replacing the existing sgRNA. The resulting plasmid was named lentidCas9-KRAB-PURO iv sgRNA (iv = improved version).

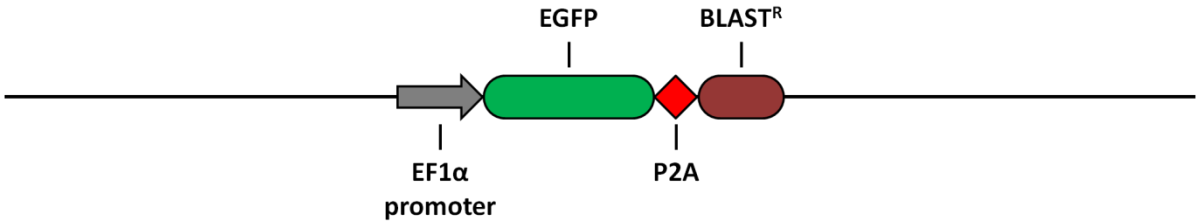


<u>lentidCas9-VP160-PURO iv sgRNA</u>: The activation domain VP160 (10 tandem repeats of VP16) was PCR amplified from pAC154-dual-dCas9VP160-sgExpression vector[113] (a gift from Prof. Rudolf Jaenisch, obtained through addgene, plasmid # 48240) using XbaI VP160 F and BamHI VP160 R primers (**Table 4**). This was cloned into XbaI-BamHI restriction sites of the lentidCas9-KRAB-PURO iv sgRNA to replace the KRAB domain in frame with dCas9. The resulting plasmid was named lentidCas9-VP160-PURO iv sgRNA.



PX458-2X-sgRNA: This plasmid was generated by Berta Duran Arqué under the supervision of Ashish Goyal. Briefly, individual sgRNAs (*NOP14-AS1* US sgRNA and *NOP14-AS1* DS sgRNA) for *NOP14-AS1* deletion were cloned individually into PX458[92] (a gift from Prof. Feng Zhang, obtained through addgene, plasmid #48138). The U6 – *NOP14-AS1* DS sgRNA expression cassette from the PX458 – *NOP14-AS1* DS sgRNA plasmid was PCR amplified and cloned into XbaI – KpnI restriction sites of the PX458 – *NOP14-AS1* US sgRNA plasmid. The resulting plasmid was named PX458-2X-sgRNA.

lenti EGFP-Blast: The *EGFP* open reading frame (ORF) was PCR amplified from PX458 using AgeI EGFP F and BamHI EGFP R primers (**Table 4**). This was cloned into AgeI-BamHI restriction sites of the lentiCas9-Blast [110] to replace the *Cas9* ORF in frame with the downstream *Blasticidin* ORF. The resulting plasmid was named lenti EGFP-Blast.



lenti *NOP14*-Blast: The *NOP14* ORF was PCR amplified from cDNA generated from NCI-H460 cells first using *NOP14* cDNA F and *NOP14* cDNA R primers and then using AgeI NOP14 F and BamHI *NOP14* R primers (**Table 4**). This was cloned into AgeI-BamHI restriction sites of the lentiCas9-Blast to replace the *Cas9* ORF in frame with the downstream *Blasticidin* ORF. The resulting plasmid was named lenti *NOP14*-Blast.

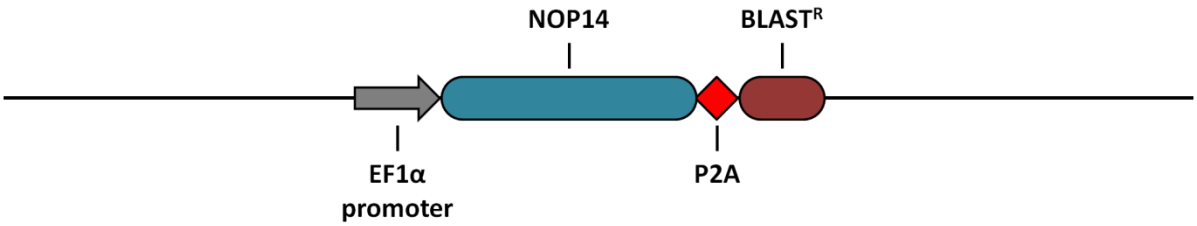


Table 4: List of primers used for molecular cloning

Name	Target	Sequence
Cas9D10A F	Cas9D10	5'-GAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGG-3'
Cas9D10A R	Cas9D10	5'-CCACAGAGTTGGTGCCGATGGCCAGGCCGATGCTGTACTTC-3'
Cas9H481A F	Cas9H481	5'-GTCCGACTACGATGTGGACGCCATCGTGCCTCAGAGCTTTC-3'
Cas9H481A R	Cas9H481	5'-GAAAGCTCTGAGGCACGATGGCGTCCACATCGTAGTCGGAC-3'
XbaI-sgRNA-F	sgRNA scaffold	5'-AAGTTCTAGAGGGCCTATTTCCCATGA-3'
KpnI-sgRNA-R	sgRNA scaffold	5'-ATATGGTACCTTGTCTGCAGAATTGGC-3'
XbaI VP160 F	VP160	5'-ATGCTCTAGAGACGCGCTGGACGATTTCG-3'
BamHI VP160 R	VP160	5'-ATATGGATCCCAACATATCCAAATCGAAGTCATCGAGC-3'
AgeI EGFP	EGFP	5'-ATACCGGTCCACCATGGTGAGCAAGGGCGAGGAGC-3'
BamHI EGFP	EGFP	5'-ATAGGATCCCTTGTACAGCTCGTCCATG-3'
NOP14 cDNA F	NOP14	5'-CCATGGCGAAGGCGAAGAAGGTCGGGG-3'
NOP14 cDNA R	NOP14	5'-TTATTTTTTGAACTTTTTCCTCTTCAGAGCCTTCC-3'
AgeI NOP14 F	NOP14	5'-CAGGACCGGTCCACCATGGCGAAGGCGAAGAAG-3'
BamHI NOP14 R	NOP14	5'-GCCGGATCCTTTTTTGAACTTTTTCCTCTTCAGAG-3'
5'-RACE GSP	NOP14-AS1	5'-GGGTGCTGGGGTTCTCCATTCAGGACA-3'
3'-RACE GSP	NOP14-AS1	5'-GTCTCCTCCCACCTCCGAGAGTGCAGTG-3'

2.2.2. Bacterial transformation

50 µl chemically competent Mach I cells were incubated with the indicated amount of plasmid/reaction mix to be transformed. Following 30 min incubation on ice, the transformation reaction was transferred to 42°C for 40 sec and then back on the ice for 2 mins. 150 µl sterile LB medium (without any antibiotics) was added to the transformation reaction and it was incubated for 1 h at 37°C under constant shaking. 100 µl of the transformation reaction was spread plated on LB Agar plates containing 100 µg/mL Ampicillin and incubated at 37°C overnight. Single colonies were picked and inoculated in 2 ml liquid LB medium containing 100 µg/mL Ampicillin.

2.2.3. Polymerase chain reaction (PCR)

All PCR reactions were performed using Q5® High-Fidelity DNA Polymerase (unless specified otherwise) according to manufacturer's recommendations. Briefly, a final reaction of 50 µl with 1x Q5® Reaction Buffer,

200 µM dNTP mix, 0.5 µM forward and reverse primer each, 10-500 ng template DNA and 1 µl of the Q5® High-Fidelity DNA Polymerase was prepared. The PCR was performed as follows:

- Initial denaturation at 98°C for 30 sec
  - 98°C for 10 sec
  - 50-72°C for 30 sec
  - 72°C for 30 sec per kb

  x 40 cycles
- Final elongation at 72°C for 10 min

2.2.4. sgRNA design and cloning

All sgRNAs against the *NOP14-AS1, MNX1-AS1, HOTAIR, LINC00441, HOXD1-AS1* and *TP53* loci were designed using the online tool available at www.crispr.mit.edu. To control for specificity, only those guides were chosen which had two or more mismatches to any of the predicted off-targets in the genome[114]. Sense and antisense oligonucleotides corresponding to the sgRNAs (**Table 5**) were annealed and cloned into PX458 / lentiGuide-Puro (a gift from Prof. Feng Zhang, obtained through addgene, plasmid #52963) / lentidCas9-KRAB-PURO iv sgRNA as described in Ran et al., Nature Protocols, 2013[92].

Table 5: Sequences of sgRNAs used in this study

Name	Target	Sequence
Control sgRNA	EGFP	5'-GGGCGAGGAGCTGTTCACCGGGG-3'
NOP14-AS1 sgRNA#1	NOP14-AS1	5'-GACGGACGGTCGCACAGACGCGG-3'
NOP14-AS1 sgRNA#2	NOP14-AS1	5'-AGAGATGTACGTCACTTCCGGGG-3'
NOP14-AS1 sgRNA#3	NOP14-AS1	5'-ACAGCCAAGCAGCGACCCGCAGG-3'
NOP14-AS1 sgRNA#4	NOP14-AS1	5'-GTCTCGGCCTCGGGGTTACGCGG-3'
NOP14-AS1 sgRNA#5	NOP14-AS1	5'-GCCCATGGGTCCGCTCCGCGGGG-3'
NOP14-AS1 sgRNA#6	NOP14-AS1	5'-GGTCGCACAGACGCGGAACAGGG-3'
NOP14-AS1 sgRNA#7	NOP14-AS1	5'-GGCTGCGCGCTCGGGACGGACGG-3'
NOP14-AS1 sgRNA#8	NOP14-AS1	5'-CAGACGCGGAACAGGGCACCAGG-3'
NOP14-AS1 sgRNA#9	NOP14-AS1	5'-GGGCACCAGGCACGCCGCCAGGG-3'
NOP14-AS1 sgRNA#10	NOP14-AS1	5'-ACCCCGCCACTGACTCCGGCCGG-3'
NOP14-AS1 sgRNA#11	NOP14-AS1	5'-GCAGCTCTTCCGCTCCGCTCAGG-3'
NOP14-AS1 sgRNA#12	NOP14-AS1	5'-CTCCGCGGGGCAGGCGTCGTGGG-3'
NOP14-AS1 US sgRNA	NOP14-AS1	5'-AGAGATGTACGTCACTTCCGGGG-3'
NOP14-AS1 DS sgRNA	NOP14-AS1	5'-AAGTAGGACAAGGCCAACTGTGG-3'
TP53 sgRNA#1	TP53	5'-GGGAAGCGTGTCACCGTCGTGG-3'
TP53 sgRNA#2	TP53	5'-GCTACCTGCTCCCTGGACGGTGG-3'
TP53 sgRNA#3	TP53	5'-GCCAGTCTTGAGCACATGGGAGG-3'
TP53 sgRNA#4	TP53	5'-CCTTTGCTTCCTCCGGCAGGCGG-3'
TP53 sgRNA#5	TP53	5'-TAGTATCTACGGCACCAGGTCGG-3'
HOXD-AS1 sgRNA#1	HOXD-AS1	5'-GGTCGCGACGGCTCTCCTCGGG-3'
HOXD-AS1 sgRNA#2	HOXD-AS1	5'-GCGAGGAGCGGCGCGCCGACCGG-3'
HOXD-AS1 sgRNA#3	HOXD-AS1	5'-GTGGCGCTGGCCGGCCAATGG-3'
LINC00441 sgRNA#1	LINC00441	5'-GCTGGTCGGTGCGCGGGCTGGG-3'
LINC00441 sgRNA#2	LINC00441	5'-GCTGGTCGGTGCGCGGGCTGGG-3'

Table 5: Sequences of sgRNAs used in this study

Name	Target	Sequence
LINC00441 sgRNA#3	LINC00441	5'-GTTCCCCACAGACGCCGGCGGG-3'
MNX1-AS1 sgRNA#1	MNX1-AS1	5'-GGGCGCACCTGTCACCGTCCCGG-3'
MNX1-AS1 sgRNA#2	MNX1-AS1	5'-GCTTAGGACCTCGCGGCGCGGG-3'
MNX1-AS1 sgRNA#3	MNX1-AS1	5'-GCCAGCGCCGCGCAACAGCCCGG-3'
MNX1-AS1 sgRNA#4	MNX1-AS1	5'-GGAGTATCCACTCCCCGTTGG-3'
MNX1-AS1 sgRNA#5	MNX1-AS1	5'-GGTTGCCAGTGCCCGCCGTCCGG-3'
HOTAIR sgRNA#1	HOTAIR	5'-GGGCCGCCCTCCTAGTGGTTCGG-3'
HOTAIR sgRNA#2	HOTAIR	5'-GAGGGGACGCACGTGTACCTGG-3'
HOTAIR sgRNA#3	HOTAIR	5'-GCGGCTCTCGCCTGAGAACTGG-3'
NOP14 sgRNA#1	NOP14	5'-GCGGCCCGGCACGTGTCTTA-3'
NOP14 sgRNA#2	NOP14	5'-CGTTCGAGGTGAAAGTTAAC-3'

2.2.5. Cell culture

NCI-H460 cells were cultured in RPMI medium (Sigma-Aldrich) supplemented with 10% FBS and 1% L-glutamine at 37°C and 5% CO_2 in a humidified chamber. HCT116 TP53(+/+) and HCT116 TP53(-/-) [115] cells were cultured in McCoy's 5A Modified Medium (Sigma-Aldrich) supplemented with 10% FBS and 1% L-glutamine at 37°C and 5% CO_2 in a humidified chamber. A549, HEK293T, MCF7 and HepG2 cells were cultured in Dulbecco's Modified Eagles Medium (DMEM, Sigma-Aldrich) supplemented with 10% FBS and 1% L-glutamine at 37°C and 5% CO_2 in a humidified chamber. Cells were regularly split when they reached 80-90% confluency. Briefly, the medium was aspirated and cells were washed with pre-warmed 1X PBS followed by incubation with 0.05% or 0.25% Trypsin-EDTA solution for 5 min at 37°C and 5% CO_2 in a humidified chamber. The cells were resuspended in appropriate medium and seeded for further experiments or maintenance.

2.2.6. Drug treatments

Etoposide (Topoisomerase II inhibitor, induces DSBs in genomic DNA) (33419-42-0, Cayman Chemical), Cisplatin (forms intrastrand crosslinks with purine bases in genomic DNA) (CAS 15663-27-1, Merck Millipore), Bleomycin (catalyses single-strand breaks as well as DSBs in genomic DNA) (CAS 9041-93-4, Merck Millipore), Doxorubicin (Topoisomerase II inhibitor, induces DSBs in genomic DNA) CAS 25316-40-9, Merck Millipore), Carboplatin (forms intrastrand crosslinks with purine bases in genomic DNA) (CAS 41575-94-4, Merck Millipore), Nutlin-3 (MDM2 antagonist, stabilizes p53) (CAS 548472-68-0, Sigma-Aldrich) and Actinomycin D (Intercalates with genomic DNA, inhibits transcription) (CAS 50-76-0, Sigma-Aldrich) were dissolved in DMSO (CAS 67-68-5, AppliChem GmbH) to prepare stock solutions. These were diluted in cell culture media to achieve the indicated final drug concentrations. At the indicated time point post drug treatments, cells were lysed in TRI reagent for RNA extraction or RIPA buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate supplemented with protease and phosphatase inhibitors) for protein extraction.

2.2.7. Virus production and transduction

Virus production was performed as described earlier[116]. Briefly, one day prior to transfection, HEK293T cells ($4x10^5$ cells / well, in a total of 2 ml media) were seeded in Poly-L-Lysine coated 6-well plates (Greiner Bio-One). Cells were co-transfected with 1.2 µg lentidCas9-Blast / lentidCas9-KRAB-Blast / lenti EGFP Blast / lenti NOP14 Blast / lentidCas9-KRAB-PURO iv sgRNA / lentidCas9-VP160-PURO iv sgRNA / lentiGuide-PURO (containing the indicated sgRNA), 0.9 µg psPAX2 (packaging plasmid) and 0.3 µg pMD2.G (envelope plasmid) using 6 µl Lipofectamine 2000 (ThermoFisher Scientific) in 200 µl Opti-MEM (ThermoFisher Scientific). 48 hours post transfection, the lentivirus-containing medium was filtered using a low protein binding 0.45 µm syringe filter (Millipore). 0.5 - 1 ml of the filtered virus along with Polybrene (Final concentration: 8 µg / ml) was then added to the indicated cells which were seeded in a 6 well plate one day prior to transduction ($1x10^5$ cells / well, in a total of 2 ml medium). 24 hours post-transduction, the medium was replaced with 2 µg/ml Puromycin (ThermoFisher Scientific) and/or 6 µg/ml Blasticidin (InvivoGen) containing media. For Puromycin selection, cells were incubated in Puromycin-containing medium for three days. For Blasticidin selection, cells were incubated in Blasticidin-containing medium for six days. Untransduced cells were used to monitor for a complete selection. The stable cell lines obtained were used for further experiments.

2.2.8. siPOOL / siRNA / ASO transfection

RNA interference (RNAi) was performed as described earlier with some minor modifications[117]. For *HOTAIR* knockdown using siPOOLs, Hela cells ($2x10^5$ cells per well) were reverse transfected with 10 nM (final concentration) of siPOOL control or siPOOL *HOTAIR* (siTOOLs Biotech) using 6 µl RNAiMAX reagent in 6-well plates. 48 hours post transfection, cells were lysed in TRI reagent for RNA extraction. For *TP53* knockdown using siPOOLs, NCI-H460 cells ($2x10^5$ cells per well) were reverse transfected with 10 nM (final concentration) of siPOOL control or siPOOL *TP53* (siTOOLs Biotech) using 6 µl RNAiMAX reagent in 6-well plates. 48 hours post transfection, cells were treated with 1 µM Doxorubicin (final concentration) or vehicle control DMSO. 24 hours post-treatment, cells were lysed in TRI reagent for RNA extraction or 1XRIPA buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1.0% NP-40 and 0.5% sodium deoxycholate) supplemented with Protease inhibitor (Roche) and Phosphatase inhibitor (Roche) cocktails for protein extraction.

For *NOP14* knockdown using siRNAs, NCI-H460 cells ($2x10^5$ cells per well) were reverse transfected with 40 nM (final concentration) of the control siRNA or two independent siRNAs against *NOP14* (Eurofins Genomics) using 2 µl RNAiMax reagent in 12-well plates. 24 hours post transfection, cells were treated with DMSO or 1 µM Doxorubicin (final concentration). 24 hours post-treatment, cells were lysed in TRI reagent for RNA extraction or RIPA buffer for protein extraction.

For *NOP14-AS1* knockdown using Antisense LNA GapmeRs (Exiqon), NCI-H460 cells ($2x10^5$ cells per well) were reverse transfected with 50 nM (final concentration) of the control GapmeR or two independent GapmeRs against *NOP14-AS1* using 2 µl Dharmafect1 reagent (Dharmacon GE Life Sciences) in 12 well plates. 24 hours post transfection, cells were lysed in 1 ml TRI reagent. For *MNX1-AS1* knockdown using ASOs (IDT DNA

Technologies), NCI-H460 cells ($4x10^5$ cells per well) were reverse transfected with 30 nM (final concentration) of the control ASO or two independent ASOs against *MNX1-AS1* using 4 µl Dharmafect1 reagent in 6-well plates.

For *NOP14-AS1* knockdown using Antisense LNA GapmeRs, followed by Doxorubicin treatment experiments, NCI-H460 cells ($4x10^5$ cells per well) were reverse transfected with 25 nM (final concentration) of the control or two independent LNA GapmeRs (Exiqon) against *NOP14-AS1* using 6 µl RNAiMAX reagent (ThermoFisher Scientific) in 6-well plates. 24 hours post transfection, cells were treated with 1 µM Doxorubicin (final concentration) or vehicle control DMSO. 24 hours post-treatment, cells were lysed in TRI reagent for RNA extraction or RIPA buffer for protein extraction.

The sequences of all the siRNAs / antisense oligos can be found in **Table 6**, whereas the sequences of all the individual siRNAs in the siPOOL HOTAIR, as well as siPOOL TP53, can be found in **Table 7**.

Table 6: List of siRNAs / antisense oligos used in this study

Name	Target	Sequence
Control LNA Gapmer	Control	5'-AACACGTCTATACGC -3'
NOP14-AS1 LNA Gapmer#1	NOP14-AS1	5'-AATTCACACCATCTTT -3'
NOP14-AS1 LNA Gapmer#2	NOP14-AS1	5'-GAGATCCGCAAGCATT -3'
Control ASO	Control	5'-GCGTATTATAGCCGATTAAC -3'
MNX1-AS1 ASO#1	MNX1-AS1	5'-TGGCCCGCAGGCTAGTGTCT -3'
MNX1-AS1 ASO#2	MNX1-AS1	5'-TCACGTAGCACTGTGGCCCG -3'
siNOP14(1)	NOP14	5'-CCAATCCGTTCGAGGTGAAAGTTAA-3'
siNOP14(2)	NOP14	5'-GGAAAGAGCTGATTGAAGA-3'

Table 7: Sequences of individual siRNAs of the siPOOLs used in this study

siRNA#	siPOOL TP53	siPOOL HOTAIR
1	5'-CACACCCTGGAGGATTTCA-3'	5'-GGGAGTACAGAGAGAATAA-3'
2	5'-GCTGTCCCTCACTGTTGAA-3'	5'-GCGCCAGACGAAGGTGAAA-3'
3	5'-GAAGGGTCAACATCTTTTA-3'	5'-CGGCAGTTCCCGGAACAAA-3'
4	5'-GTCTACCTCCCGCCATAAA-3'	5'-GAGTGCACTGTCTCTCAAA-3'
5	5'-CTTTTGCATGTTTTGTAGA-3'	5'-GACATAGGAGAACACTTAA-3'
6	5'-CTGGATGGAGAATATTTCA-3'	5'-GACTTGAGCTGCTCCGGAA-3'
7	5'-GCCTTGAAACCACCTTTTA-3'	5'-GGAATCCACCTGCCTGTTA-3'
8	5'-GCATTGTGAGGGTTAATGA-3'	5'-CCTAGACTTAAGATTCAAA-3'
9	5'-CAAGGCCCATATCTGTGAA-3'	5'-GGGTCCTAGCTCGCCACAT-3'
10	5'-GCCCACTTCACCGTACTAA-3'	5'-GCCTTTGGAAGCTCTTGAA-3'
11	5'-CATACCAGCTTAGATTTTA-3'	5'-GCTTGTTAACAAGACCAGA-3'
12	5'-GGATGGGGAGTAGGACATA-3'	5'-GAGACACATGGGTAACCTA-3'
13	5'-GTGTGGAGTATTTGGATGA-3'	5'-GCAACCACGAAGCTAGAGA-3'
14	5'-CAACAAGATGTTTTGCCAA-3'	5'-GCCAAGCACCTCTATCTCA-3'
15	5'-GATCTCTTATTTTACAATA-3'	5'-GCTGACAGGGTCTGGGACA-3'
16	5'-GAGGATTTCATCTCTTGTA-3'	5'-GCTCCGCTTCGCAGTGGAA-3'
17	5'-CACTGTTGAATTTTCTCTA-3'	5'-GCACATTCTGCCCTGATTT-3'
18	5'-CCCATCCTCACCATCATCA-3'	5'-GGGCCTAAGCCAGTACCGA-3'
19	5'-GAGGTTGGCTCTGACTGTA-3'	5'-GTCCGTTCAGTGTCAGAAA-3'

Table 7: Sequences of individual siRNAs of the siPOOLs used in this study

siRNA#	siPOOL TP53	siPOOL HOTAIR
20	5'-CTGACAACCTCTTGGTGAA-3'	5'-CGGGACTTTGCACTCTAAA-3'
21	5'-GCCAAACCCTGTCTGACAA-3'	5'-GGCGCCTTCCTTATAAGTA-3'
22	5'-CATGGGGTCTAGAACTTGA-3'	5'-GCCCAGAGAACGCTGGAAA-3'
23	5'-CTCACAGAGTGCATTGTGA-3'	5'-GGAGGCGCTAATTAATTGA-3'
24	5'-GTTAAGGGTTAGTTTACAA-3'	5'-GCTAAATAGACTCAGGACT-3'
25	5'-GGGATGTTTGGGAGATGTA-3'	5'-GCAGATGGAGATTACCATT-3'
26	5'-CACTACAACTACATGTGTA-3'	5'-GGCCTGGGAGTTCCACAGA-3'
27	5'-CATGAGCGCTGCTCAGATA-3'	5'-CCTTTGCTTCGTGCTGATT-3'
28	5'-GCTGTGGGTTGATTCCACA-3'	5'-GTAGACCCAGCCCAATTTA-3'
29	5'-GGTGAACCTTAGTACCTAA-3'	5'-CGGAACCCATGGACTCATA-3'
30	5'-GTACCACCATCCACTACAA-3'	5'-GAGAAGTGCTGCAACCTAA-3'

2.2.9. RNA extraction

RNA was isolated using TRI reagent (Sigma-Aldrich) according to the manufacturer's recommendations with some minor modifications. Cells were lysed using 1 ml TRI reagent per well of a 6 well plate. 200 µl Chloroform (Carl Roth) was added per ml of lysate, vortexed briefly and centrifuged for 15 minutes at 4°C and 13000g. Following phase separation, 500 µl of the aqueous layer was transferred to a new tube and 500 µl Isopropanol (Sigma-Aldrich) was added to it. This was gently mixed, incubated at room temperature for 10 min and centrifuged for 10 minutes at 4°C and 13000g. The supernatant was discarded and the precipitate was washed once with 70% Ethanol (Sigma-Aldrich) followed by centrifugation at 13,000 g for 5 min at 4°C. The supernatant was discarded and the precipitate was resuspended in a 50 µl solution containing 1x DNase I reaction buffer (Roche) and 1 µl of DNase I (10U/µl; Roche) followed by an incubation for 30 min at 37°C, to ensure removal of genomic DNA. 150 µl of nuclease free water was added to the DNAse I treated RNA. Thereafter, the RNA was subjected to Phenol: Chloroform extraction to remove any protein contaminations. To each tube, 200 µl of Roti-Aqua-P/C/I (Carl Roth) was added, vortexed briefly and centrifuged for 5 minutes at 4°C and 13000g. Following phase separation, 200 µl of the aqueous layer was transferred to a new tube and 200 µl Chloroform (Carl Roth) was added to it. This was again vortexed briefly and centrifuged for 5 minutes at 4°C and 13000g. 200 µl of the aqueous layer was transferred to a fresh tube and mixed with 20 µl of 3 M sodium acetate pH 5.2 (Carl Roth). 500 µl of ice-cold ethanol (Sigma-Aldrich) was added, mixed gently by inverting the tube and incubated overnight at -80°C. Following overnight incubation, the RNA solution was centrifuged for 10 minutes at 4°C and 13000g. The supernatant was discarded and the precipitate was washed once with 70% Ethanol (Sigma-Aldrich) followed by centrifugation at 13,000 g for 5 min at 4°C. The supernatant was discarded and the precipitate was resuspended in 20-50 µl nuclease free water.

2.2.10. Reverse Transcription - quantitative Polymerase Chain Reaction (RT-qPCR)

A total of 1-2 µg RNA was reverse transcribed to generate cDNA using random hexamer primers and RevertAid Reverse Transcriptase (ThermoFisher Scientific) as per the manufacturer's recommendations. Briefly, a mixture

of RNA, 2 µl of dNTP mix (Thermo Fisher Scientific) and 1 µl of Random Hexamer Primer (Thermo Fisher Scientific) was prepared in a final volume of 11.5 µl and incubated for 5 min at 65°C. The reaction was transferred on ice and 4 µl Reverse Transcriptase buffer, 1 µl RevertAid Reverse Transcriptase (Thermo Fisher Scientific). And 0.5 µl RiboLock RNase Inhibitor (Thermo Fisher Scientific) was added. Thereafter, the reaction was incubated at 25°C for 10 min, 42°C for 1 h, and 72°C for 10 min. To further control for any residual genomic DNA contamination, a minus-RT reaction was also performed where reverse transcriptase was replaced by water. The cDNA generated was diluted by a factor of 20 or 40 and RT-qPCR was performed using PowerSYBRGreen PCR Master Mix (ThermoFisher Scientific) in a StepOnePlus qPCR instrument (Life Technologies). Quantification was performed using the ΔΔCt method and the housekeeping gene *PPIA* (*Cyclophilin A*) was used for normalization. A final reaction volume of 15 µl containing 7.5 µl Power SYBR Green Master Mix, 0.45 µl forward and reverse primer each (10 µM), and 5 µl of pre-diluted cDNA and 1.6 µl water was prepared. The qPCR was performed as follows:

- Initial denaturation at 95°C for 10 min
  - 95°C for 15 sec
  - 60°C for 30 sec

  x 40 cycles
- Final elongation at 72°C for 10 min

The sequences of all RT-qPCR primers can be found in **Table 8**.

Table 8: List of primers used for RT-qPCR (F = Froward, R = Reverse)

Name	Target	Sequence
NOP14-AS1 amplicon#1 F	NOP14-AS1	5'-CCATGCCCTCCTTGTTTACT-3'
NOP14-AS1 amplicon#1 R	NOP14-AS1	5'-GGGAAAGGGCTGTTATCATCTT-3'
NOP14-AS1 amplicon#2 F	NOP14-AS1	5'-CCAGAGGTGCATTTCAGGAT-3'
NOP14-AS1 amplicon#2 R	NOP14-AS1	5'-AAGGCAGGAAGATTGCTTCA-3'
MFSD10 F	MFSD10	5'-GTCTACTTCCTCTACCTCTTCCT-3'
MFSD10 R	MFSD10	5'-CTGCTGTAGGCTACTGAACTG-3'
TP53 F	TP53	5'-GTGACACGCTTCCCTGGATT-3'
TP53 R	TP53	5'-TGTTTCCTGACTCAGAGGGG-3'
WRAP53α F	WRAP53α	5'-CGGAGCCCAGCAGCTACC-3'
WRAP53α R	WRAP53α	5'-TTGTGCCAGGAGCCTCGCA-3'
HOXD-AS1 F	HOXD-AS1	5'-TTCATCTGGGAGTTCTTGGCA-3'
HOXD-AS1 R	HOXD-AS1	5'-GCTGGACCTGGCCTGAAAAT-3'
HOXD1 F	HOXD1	5'-CGCACGAATTTCAGCACCAA-3'
HOXD1 R	HOXD1	5'-AGGTGCAAGCAGTTGGCTAT-3'
LINC00441 F	LINC00441	5'-GGACGTGCTTCTACCCAGAAC-3'
LINC00441 R	LINC00441	5'-TCCTTCTCAGTTGACGAGTTCA-3'
RB1 F	RB1	5'-GAGGACCTGCCTCTCGTCAG-3'
RB1 R	RB1	5'-TCCCAAGTTAACCAAGCTCTCT-3'
MNX1-AS1 F	MNX1-AS1	5'-CCAAAGCTCTGCAGGTCGAA-3'
MNX1-AS1 R	MNX1-AS1	5'-GCTGCAGCATTCTGGGAAAAG-3'
MNX1 F	MNX1	5'-GTTGAGCTTGAACTGGTGCTC-3'
MNX1 R	MNX1	5'-GATCCTGCCTAAGATGCCCG-3'
HOTAIRα F	HOTAIR	5'-GACAGAAGGAAAGCCCTCCAG-3'

Table 8: List of primers used for RT-qPCR (F = Froward, R = Reverse)

Name	Target	Sequence
HOTAIRα R	HOTAIR	5'-GGTTCCGGAAATCAGGGCAG-3'
HOTAIRβ F	HOTAIR	5'-GGGGTGTTGGTCTGTGGAAC-3'
HOTAIRβ R	HOTAIR	5'-CCTGTGCTCTGGAGCTTGAT-3'
HOXC11 F	HOXC11	5'-GGCTGAGGAGGAGAACACAAA-3'
HOXC11 R	HOXC11	5'-GTTCCCGGATCTGGAATTTCG-3'
NOP14 F	NOP14	5'-GAGGTTTGCTCTGGAACAGC-3'
NOP14 R	NOP14	5'-TCGATGTCTGCCAAAGACTG-3'
CDKN1A F	CDKN1A	5'-CGAAGTCAGTTCCTTGTGGAG-3'
CDKN1A R	CDKN1A	5'-CATGGGTTCTGACGGACAT-3'
CyclophilinA F	Cyclophilin A	5'-GTCAACCCCACCGTGTTCTT-3'
CyclophilinA R	Cyclophilin A	5'-CTGCTGTCTTTGGGACCTTGT-3'

2.2.11. Western Blot

Protein lysates were prepared as described earlier with some minor modifications[118]. Briefly, cells were rinsed with ice-cold phosphate-buffered saline (PBS, Sigma-Aldrich) and incubated for 30 min on ice in 200 µl RIPA buffer per well of a 6-well plate. Following lysis, the crude lysate was subjected to centrifugation for 30 min at 4°C and 17000g to precipitate cellular debris. The protein concentration was determined using the BCA assay and equal amounts (10 µg / 20 µg) of lysates were separated on self-cast 10% SDS-PAGE gels and transferred to Nitrocellulose membranes. The membranes were blocked using 5% milk in TBST (50 mM Tris, 150 mM NaCl, 0.05% Tween 20) and then incubated with primary antibodies according to the manufacturer's recommendations. The following primary antibodies were used: NOP14 (#HPA039596, Sigma-Aldrich), TP53 (#554293, BD Pharmingen) RB1 (#554136, BD Pharmingen), TP53 (#554293, BD Pharmingen / sc-126, Santa Cruz Biotechnology), PARP (#9542, Cell Signaling Technology) and GAPDH (#MAB374, EMD Millipore). Following incubation with the primary antibody, the membranes were washed with TBST and incubated with secondary antibodies which were either HRP-coupled (Dianova) or NIR-fluorescent reagent (LI-COR IRDye® 800CW) conjugated. Chemiluminescence detection was performed using Super Signal West Pico chemiluminescent substrate (ThermoFisher Scientific) or Super Signal™ West Femto chemiluminescent substrates on a ChemoCam Imager (Intas). Fluorescence detection was performed using LI-COR Biosciences Odyssey®. Image Studio™ Lite was used for image analysis and band quantification.

2.2.12. 5'- and 3'-RACE

RNA was extracted from HepG2 cells treated with 50 µM Etoposide for 16h and reverse transcribed using the SMARTer RACE cDNA Amplification Kit (Clontech) according to manufacturer's instructions. 5'- and 3'- RACE PCR products were generated using Advantage 2 PCR Enzyme System (Clontech) according to the manufacturer's instructions. The gene-specific primers used for 5'-and 3'-RACE can be found in **Table 4**. The PCR products obtained were gel eluted and cloned into the pCR2.1-TOPO vector using the TOPO TA cloning kit (ThermoFisher Scientific) and sequenced using the M13 reverse primer. Alignment of the RACE sequences to the UCSC genome browser was performed using BLAT tool[119].

3. Results

3.1. CRISPRi as a tool for knockdown of lncRNA genes

Loss-of-function models are indispensable for dissecting gene function. Prior to the advent of RNAi, generation of loss-of-function models required painstaking work and often relied on chance mutations in the genes of interest. Over the past few decades, RNAi has greatly facilitated generation of such models. However, when it comes to lncRNA research, RNAi has several limitations: (1) RNAi machinery is mainly cytoplasmic, making it difficult to target the nuclear fraction of lncRNAs[120, 121], (2) siRNAs are often inefficient in targeting RNAs with strong secondary structures, which is often the case for lncRNAs[122], (3) barring a few exceptions[123], the RNAi machinery acts post-transcriptionally and thus cannot uncover phenotypes arising as a result of the act of lncRNA transcription.

Antisense oligos (ASOs) offer an orthogonal technique to knockdown lncRNA expression[124]. ASOs do not rely on the RNAi machinery and instead guide the endogenous RNase H to their target in a sequence-specific manner. Unlike RNAi, ASOs can efficiently target nuclear RNAs and can also target nascent transcripts, making them suitable to study transcript-dependent *in cis* effects[121, 125, 126]. However, they cannot inhibit transcription and their short-lived knockdown efficiencies as well as the associated toxicity, make them an inferior tool for the generation of loss-of-function model.

CRISPR/Cas9 has recently emerged as a powerful tool for genome editing. As explained earlier, Cas9 can be used for lncRNA manipulation in several ways: Cas9 can be used to either delete promoter or the entire body of a lncRNA gene. It can also be used to knockin RNA destabilizing elements to knockdown lncRNA expression. Lastly, dCas9-based custom transcription factors can be used to modulate lncRNA transcription without altering its genomic sequence. All these approaches can overcome the limitations imposed by RNAi/ASO-based lncRNA knockdown. However, Cas9 based deletion approaches for lncRNA knockout can lead to simultaneous deletion of underlying genomic elements and thus assign false phenotypes to the lncRNA gene targeted. Also, these approaches involve time-consuming clonal selection. CRISPRi-mediated lncRNA knockdown promises to overcome these problems as dCas9 ordCas9-KRAB do not alter the genomic sequence, but inhibit the transcription of the lncRNA.

3.1.1. *NOP14-AS1* knockdown using CRISPRi affects *MFSD10* expression

To study the *in cis*-regulatory effects of *NOP14-AS1* lncRNA on the expression of *NOP14* mRNA, a loss-of-function model was required. Given the limitations of RNAi- and ASO-based approaches for lncRNAs, it was proposed to utilize CRISPR/Cas9 instead. A CRISPRi-based knockdown of *NOP14-AS1* from its endogenous promoter would be able to reverse the *NOP14-AS1* transcript- or its transcription-dependent effects on the expression of *NOP14* mRNA.

Five different sgRNAs spanning the *NOP14-AS1* promoter were designed (**Figure 7A**). sgRNA#1 and sgRNA#2 were designed upstream of the *NOP14-AS1* TSS whereas sgRNA#3 to sgRNA#5 were designed downstream of the *NOP14-AS1* TSS with the aim of identifying at least two sgRNAs that could repress *NOP14-AS1* expression. These sgRNAs were introduced into NCI-H460 cells expressing dCas9 (without an inhibitory domain) and expression of *NOP14-AS1* and *MFSD10* were monitored. Since *NOP14-AS1* arises from the bidirectional promoter of *MFSD10*, it was necessary to monitor the *MFSD10* expression as it was possible that dCas9 could also target its expression which would be indistinguishable from effects of *NOP14-AS1* knockdown on *MFSD10* expression via earlier described *in cis*-regulatory mechanisms.

Expression of these sgRNAs in NCI-H460 cells expressing dCas9 (without an inhibitory domain), did not yield in an efficient knockdown of *NOP14-AS1* (**Figure 7B,C**). sgRNA#4 and sgRNA#5 showed a minor repression of *NOP14-AS1* (as detected using amplicon #1 only), but did not affect *MFSD10* expression (**Figure 7D**). sgRNA#1, which bound a region immediately downstream of the *MFSD10* TSS, strongly repressed its expression without affecting *NOP14-AS1* indicating that this sgRNA targeted the core promoter of *MFSD10*. Several other sgRNAs targeting the promoter of *NOP14-AS1* were also designed and tested in combination with dCas9, but they failed to efficiently knockdown *NOP14-AS1* expression (**Supplementary Figure 33**). These results showed that dCas9 alone was not a potent inhibitor of transcription unless it was targeted to the core promoter.

On the other hand, introduction of these five sgRNAs into NCI-H460 cells expressing dCas9-KRAB resulted in strong repression of the intended target *NOP14-AS1* for three out of five of these sgRNAs (sgRNA#1, sgRNA#3 and sgRNA#5) (**Figure 7A,B**) indicating that dCas9-KRAB could be used to efficiently repress lncRNA transcription from their endogenous promoter. However, *MFSD10* expression was also affected by all sgRNAs which repressed *NOP14-AS1* expression (**Figure 7D**). To determine whether the simultaneous knockdown of *MFSD10* and *NOP14-AS1* using dCas9-KRAB was due to an endogenous cis-regulatory mechanism or was an artifact of this technique, antisense LNA GapmeRs were used to knockdown expression of the lncRNA *NOP14-AS1*. Knockdown of *NOP14-AS1* using two independent antisense LNA GapmeRs reduced the expression of *NOP14-AS1* (**Figure 7E**). In contrast to the results from a dCas9-KRAB-mediated knockdown, *MFSD10* expression was unchanged or even slightly induced upon *NOP14-AS1* knockdown using antisense LNA GapmeRs (**Figure 7E**) indicating that the observed reduction of *MFSD10* expression upon *NOP14-AS1* knockdown using dCas9-KRAB was not as a result of reduced *NOP14-AS1* lncRNA transcript. However, since ASOs cannot inhibit transcription, it could not be ruled out that dCas9-KRAB-mediated repression of *NOP14-AS1* transcription could affect *MFSD10* expression *in cis*.

To further clarify whether this dCas9-KRAB-mediated simultaneous knockdown of the neighboring gene from a proximal promoter along with the intended target was restricted to the *MFSD10*/*NOP14-AS1* locus, or was a widespread phenomenon, it was decided to test this system for knockdown of other similarly located lncRNA/mRNA gene pairs.

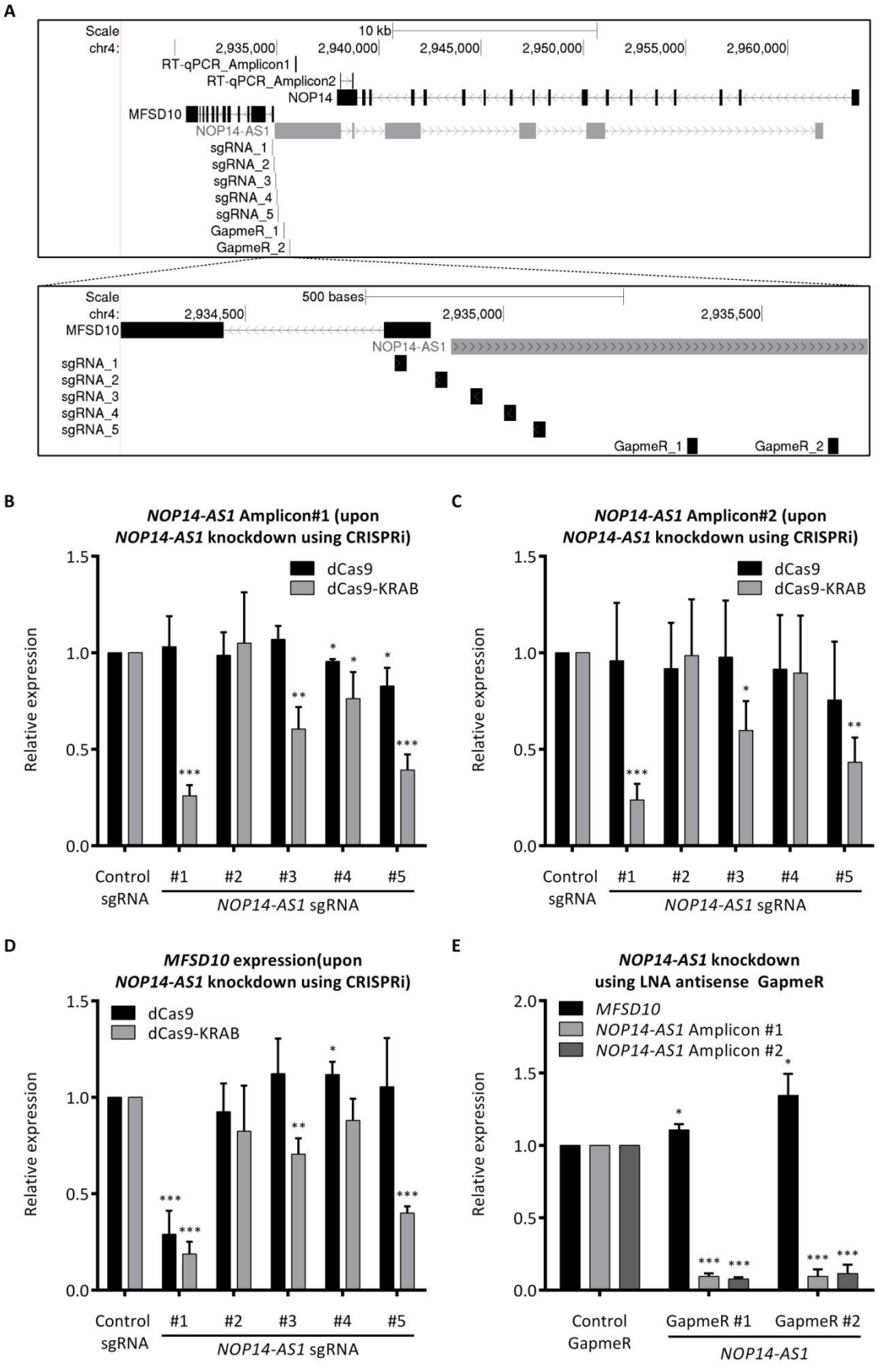


Figure 7: *NOP14-AS1* and *MFSD10* gene modulation using CRISPRi and antisense LNA GapmeRs

A: Schematic representation of the *MFSD10*/*NOP14-AS1* genomic locus depicting the sgRNAs and LNA antisense GapmeRs used to target *NOP14-AS1* as well as the RT-qPCR amplicons used to detect its expression.
B-D: NCI-H460 cells expressing either dCas9 or dCas9-KRAB were transduced with either a control sgRNA or one of the five indicated sgRNAs targeting the *NOP14-AS1* promoter. RT-qPCR results for (B) *NOP14-AS1* amplicon #1, (C) *NOP14-AS1* amplicon #2 and (D) *MFSD10* normalized to *Cyclophilin A* and control sgRNA. Error bars represent SD (n≥4). * p<0.05; ** p<0.01; *** p<0.001 compared to dCas9/dCas9-KRAB + control sgRNA, unpaired two-sided t-test.
E: NCI-H460 cells were transfected with either a control or two independent antisense LNA GapmeRs against NOP14-AS1. RT-qPCR results for *NOP14-AS1* amplicon#1, amplicon#2 and *MFSD10* normalized to *Cyclophilin A* and control antisense LNA GapmeR. Error bars represent SD (n=4). * p<0.05; ** p<0.01; *** p<0.001 compared to control GapmeR, unpaired two-sided t-test.
Data used here has been published as a part of the following manuscript: Goyal et al., Nucleic Acids Res, 2016[116].

3.1.2. *TP53* knockdown using CRISPRi affects *WRAP53α* expression

Similar to the *MFSD10*/*NOP14-AS1* locus, the well-known tumor suppressor gene *TP53* (p53) is also transcribed in a bidirectional fashion, partially overlapping with another protein-coding gene (*WRAP53*) in a head-to-head antisense orientation. *WRAP53* is expressed from three different TSSs, namely α, β, and γ. The α-TSS lies roughly 100 bp downstream of the *TP53* TSS overlapping with its first exon (**Figure 8A**). The transcript arising from this TSS (*WRAP53α*) does not code for a protein but is known to bind to *TP53* mRNA via sense-antisense base pairing and regulate *TP53* mRNA stability, expression and is also required for *TP53* induction upon DNA damage[43]. Since *TP53* is one of the most widely characterized genes and this locus is another example of a bidirectional promoter, it was decided to knockdown *TP53* expression using dCas9/dCas9-KRAB as well as an siPOOL (a pool of 30 defined siRNAs with non-identical seed regions to minimize any potential off-target effects[127]) to compare the effects on *WRAP53α* expression.

Similar to *NOP14-AS1*, five different sgRNAs distributed across the *TP53* promoter were designed to target *TP53* expression. sgRNA#1, sgRNA#2 and sgRNA#3 were designed downstream of the *TP53* TSS whereas sgRNA#4 and sgRNA#5 were designed upstream of the *TP53* TSS (**Figure 8A**). When these sgRNAs were introduced into dCas9 expressing NCI-H460 cells, sgRNA#1 and sgRNA#2 resulted in *TP53* mRNA as well as protein knockdown (**Figure 8B,C,D**). Both of these sgRNAs also resulted in *WRAP53α* knockdown (**Figure 8E**). When these sgRNAs were introduced into dCas9-KRAB expressing NCI-H460 cells, all sgRNAs led to a significant repression of *TP53* mRNA and protein expression to varying extents (**Figure 8B,C,D**) indicating once again that dCas9-KRAB was a more potent inhibitor of transcription as compared to dCas9 alone. More importantly, all sgRNAs strongly affecting *TP53* expression also had a strong impact on *WRAP53α* expression except for sgRNA#1 which displayed a non-significant trend (**Figure 8E**).

In contrast to the results from a dCas9-/dCas9-KRAB-mediated knockdown, RNAi-mediated silencing of *TP53* using an siPOOL resulted in a strong downregulation of *TP53* mRNA as well as protein but had no effect on *WRAP53α* expression (**Figure 9**). Thus, dCas9-/dCas9-KRAB-mediated targeting of bidirectional promoters again led to a repression of both genes in the locus.

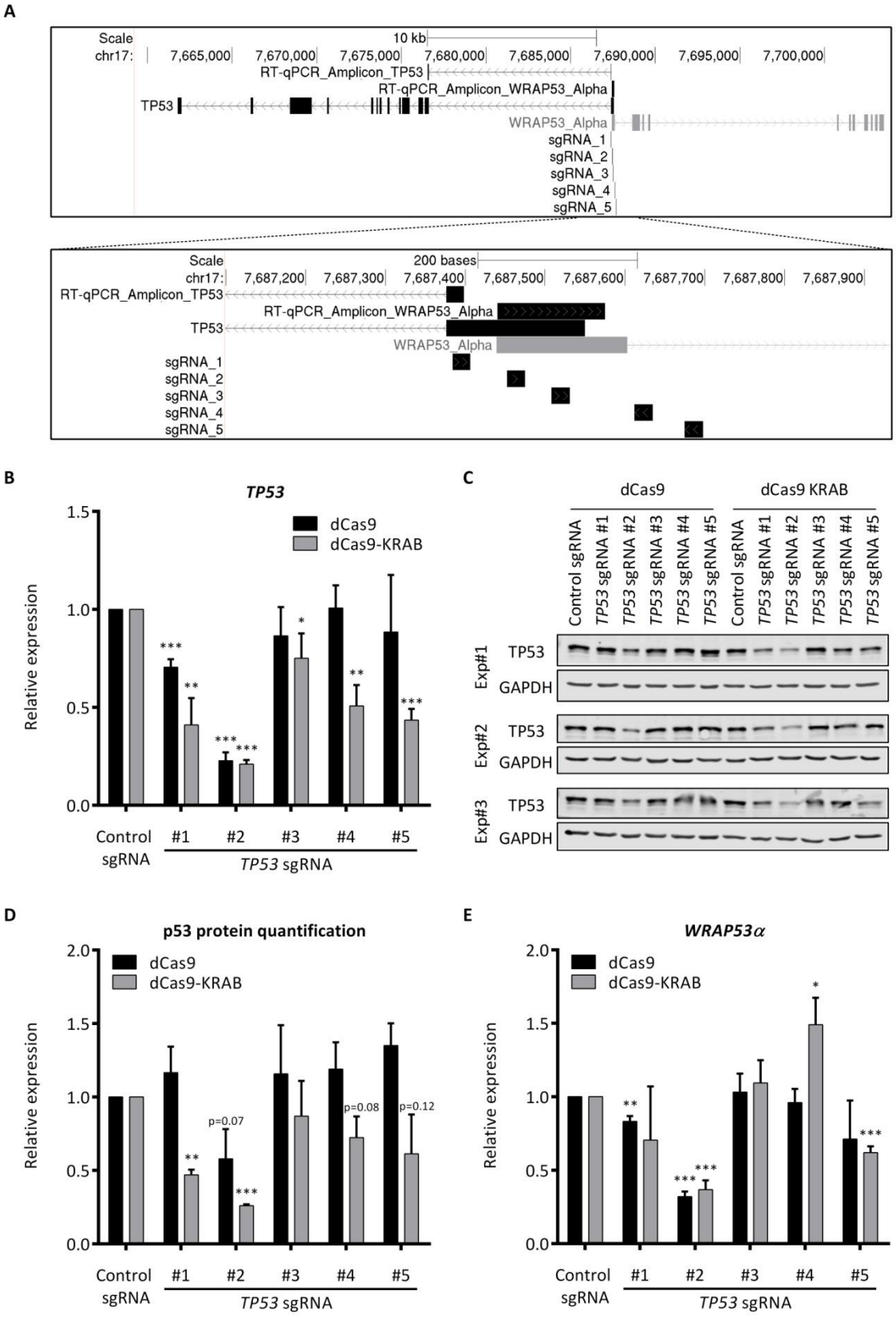


Figure 8: *TP53* knockdown using CRISPRi

A: Schematic representation of the *TP53 / WRAP53α* genomic locus depicting the sgRNAs used to target *TP53* as well as the RT-qPCR amplicons used to detect *TP53* and *WRAP53α* expression.

B-D: NCI-H460 cells expressing either dCas9 or dCas9-KRAB were transduced with either a control sgRNA or one of five indicated sgRNAs targeting *TP53*. (B) RT-qPCR results for *TP53* normalized to *Cyclophilin A* and control sgRNA. Error bars represent SD (n=4). * p<0.05; ** p<0.01; *** p<0.001 compared to dCas9/dCas9-KRAB + control sgRNA, unpaired two-sided t-test. (C) Western blot results for p53. GAPDH was used as loading control. (D) Quantification of the p53 expression normalized to GAPDH and control sgRNA. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to dCas9/dCas9-KRAB + control sgRNA, unpaired two-sided t-test. (E) RT-qPCR results for *WRAP53α* normalized to *Cyclophilin A* and control sgRNA. Error bars represent SD (n=4). * p<0.05; ** p<0.01; *** p<0.001 compared to dCas9/dCas9-KRAB + control sgRNA, unpaired two-sided t-test.
Data used here has been published as a part of the following manuscript: Goyal et al, NAR, 2016[116]

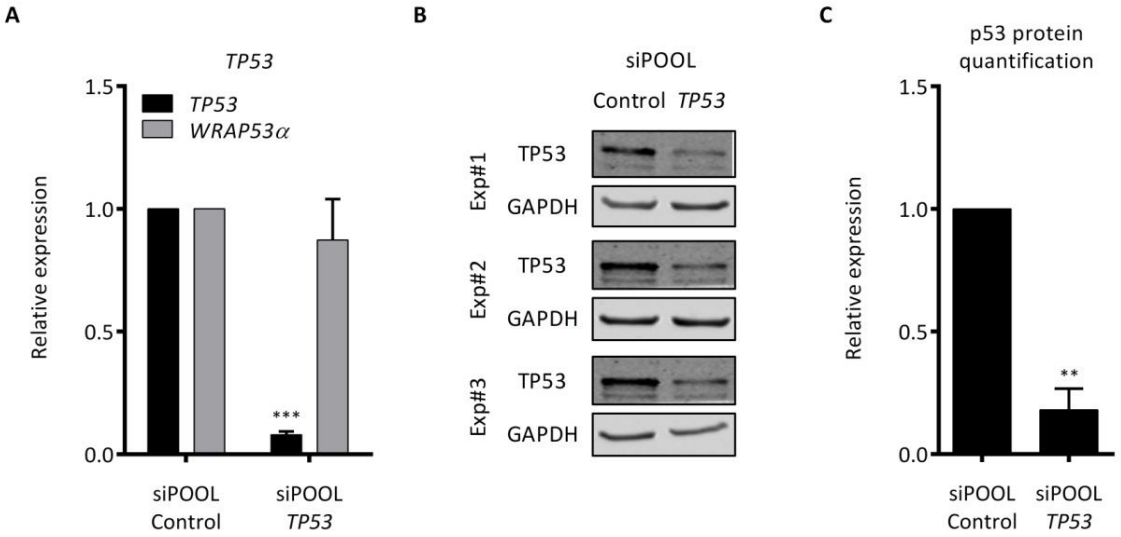


Figure 9: *TP53* knockdown using an siPOOL
NCI-H460 cells were transfected with either a control siPOOL or an siPOOL targeting *TP53*. (A) RT-qPCR results for *TP53* and *WRAP53α* normalized to *Cyclophilin A* and siPOOL control. Error bars represent SD (n=6). * p<0.05; ** p<0.01; *** p<0.001 compared to siPOOL control, unpaired two-sided t-test. (B) Western blot results for p53. GAPDH was used as loading control. (C) Quantification of the p53 expression normalized to GAPDH and control siPOOL. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control siPOOL, unpaired two-sided t-test.
Data used here has been published as a part of the following manuscript: Goyal et al., Nucleic Acids Res, 2016[116].

3.1.3. *HOXD-AS1* knockdown using dCas9-KRAB affects *HOXD1* expression

As another example of a lncRNA arising from a bidirectional promoter, *HOXD-AS1* was selected. *HOXD-AS1* is a lncRNA encoded in the *HOXD* locus and similarly to *NOP14-AS1*, *HOXD-AS1* also arises from the bidirectional promoter shared with a protein-coding gene, *HOXD1*. *HOXD-AS1* knockdown using siRNAs was shown to have no impact on *HOXD1* expression. Three different sgRNAs spanning the *HOXD-AS1* promoter were designed (**Figure 10A**) and expressed in A549 cells along with dCas9-KRAB. All three sgRNAs resulted in strong repression of *HOXD-AS1* as well as *HOXD1* (**Figure 10B,C**).

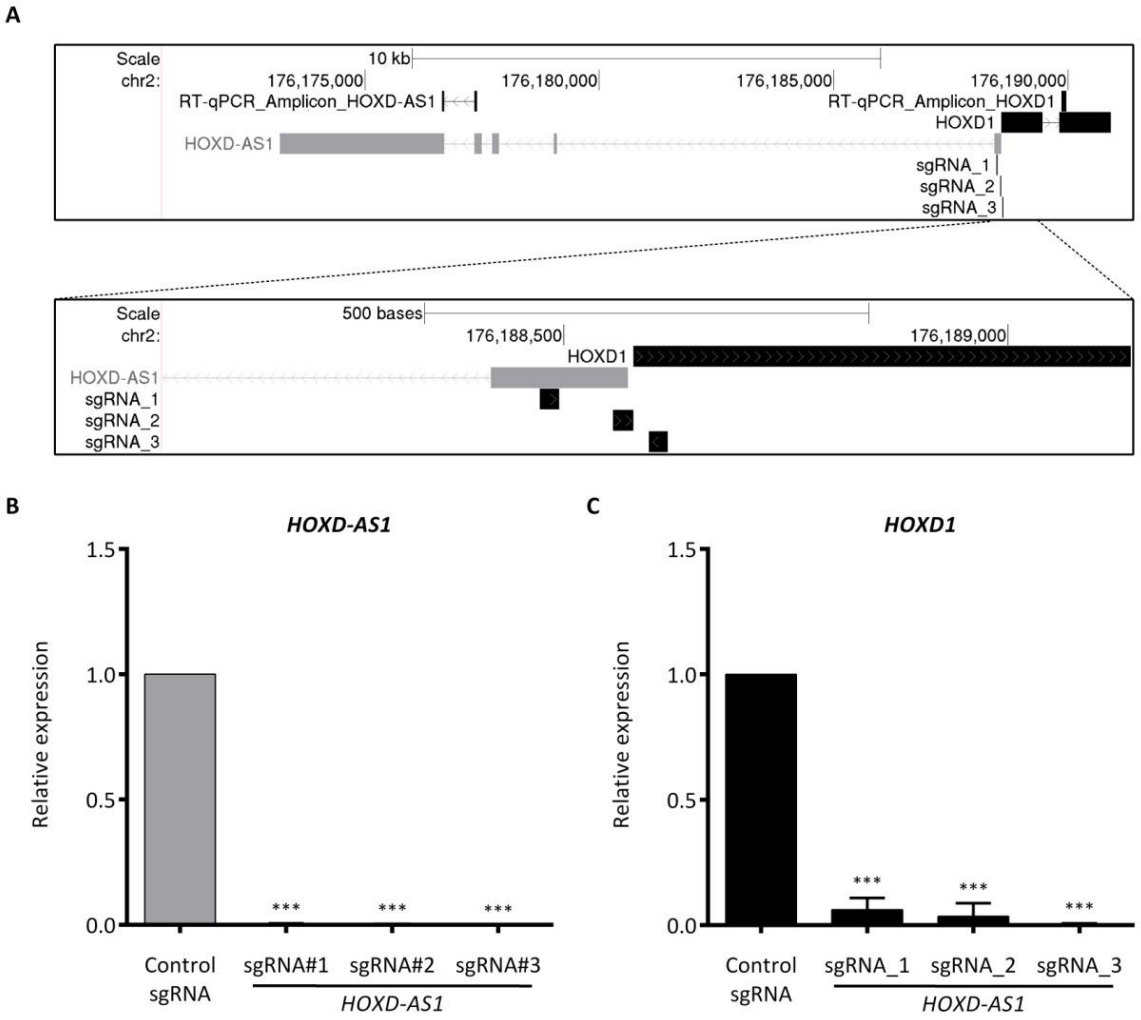


Figure 10: *HOXD-AS1* knockdown using dCas9-KRAB

A: Schematic representation of the *HOXD1* / *HOXD-AS1* genomic locus depicting the sgRNAs used to target *HOXD-AS1* as well as the RT-qPCR amplicons used to detect *HOXD1* and *HOXD-AS1* expression.

B,C: A549 cells were transduced with lentidCas9-KRAB-PURO iv sgRNA containing either control sgRNA or one of the three indicated sgRNAs targeting *HOXD-AS1*. RT-qPCR results for (B) *HOXD1-AS1* and (C) *HOXD* normalized to *Cyclophilin A* and control sgRNA. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test.

Data used here has been published as a part of the following manuscript: Goyal et al., Nucleic Acids Res, 2016[116].

3.1.4. *LINC00441* knockdown using dCas9-KRAB affects *RB1* expression

LINC00441 arises from the bidirectional promoter shared with the tumor suppressor gene *RB1* and its knockdown using siRNAs in A549 cells did not impact *RB1* expression[128]. However, dCas9-KRAB-mediated knockdown of *LINC00441* using three independent sgRNAs targeting its promoter strongly reduced *RB1* mRNA as well as protein expression in these cells (**Figure 11A-D**).

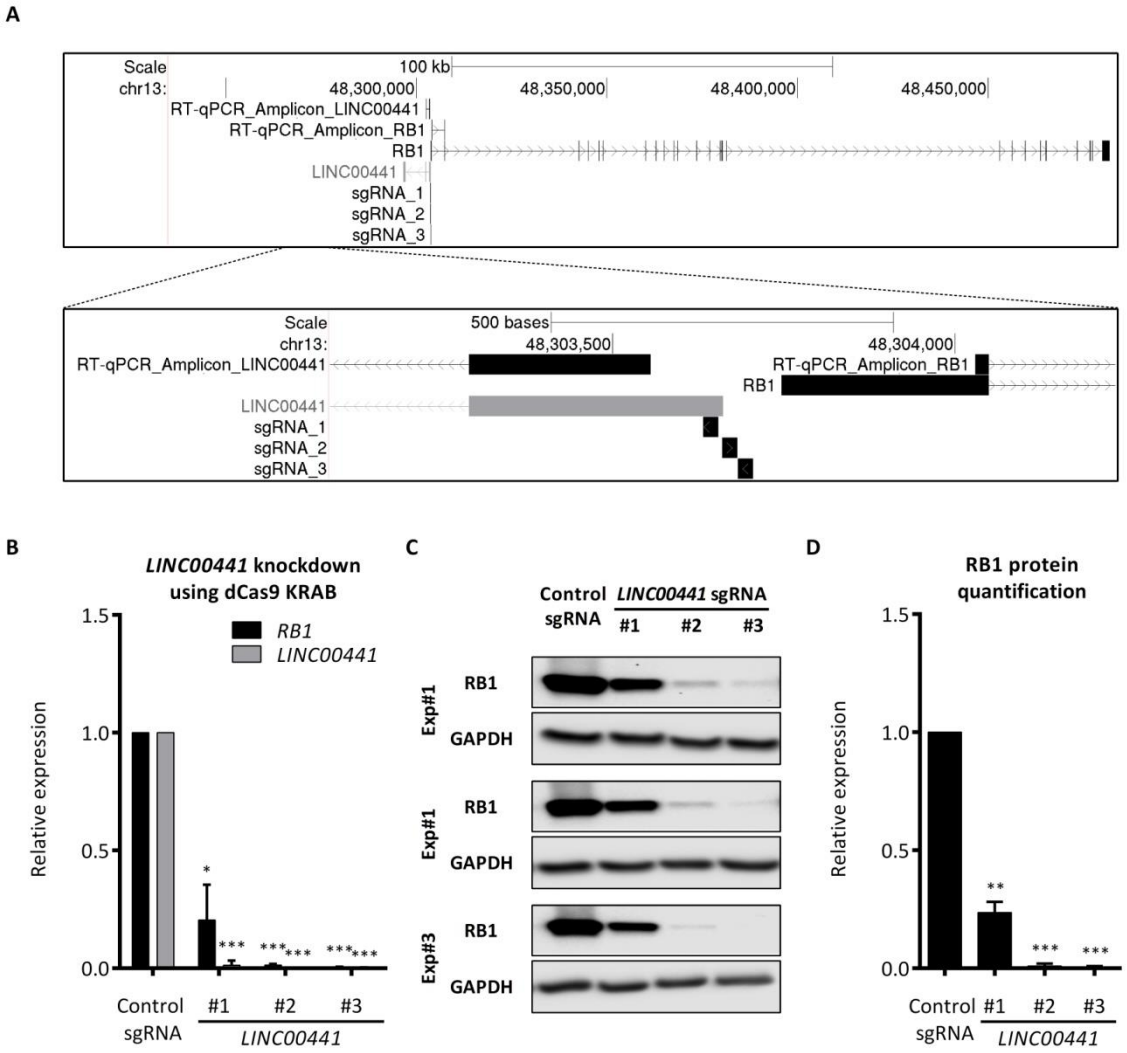


Figure 11: *LINC00441* knockdown using dCas9-KRAB

A: Schematic representation of the *RB1 / LINC00441* genomic locus depicting the sgRNAs used to target *LINC00441* and the RT-qPCR amplicons used to detect *RB1* and *LINC00441* expression.

B-D: A549 cells were transduced with lentidCas9-KRAB-PURO iv sgRNA containing either a control sgRNA or one of the three indicated sgRNAs targeting *LINC00441*. (B) RT-qPCR results for *RB1* and *LINC00441* normalized to *Cyclophilin A* and Control sgRNA. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test. (C) Western blot results for RB1. GAPDH was used as loading control. (D) Quantification of the RB1 protein expression normalized to GAPDH and control sgRNA. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test.

Data used here has been published as a part of the following manuscript: Goyal et al., Nucleic Acids Res, 2016[116].

3.1.5. *MNX1-AS1* knockdown using dCas9-KRAB affects *MNX1* expression

LncRNA *MNX1-AS1* arises from the bidirectional promoter of *MNX1*. A dCas9-KRAB-mediated knockdown of *MNX1-AS1* using five independent sgRNAs targeting its promoter led to knockdown of both *MNX1-AS1* as well as its neighboring mRNA *MNX1* (**Figure 12A,B**). On the other hand, *MNX1-AS1* knockdown using two independent ASOs reduced *MNX1-AS1* expression without affecting *MNX1* expression (**Figure 12C**).

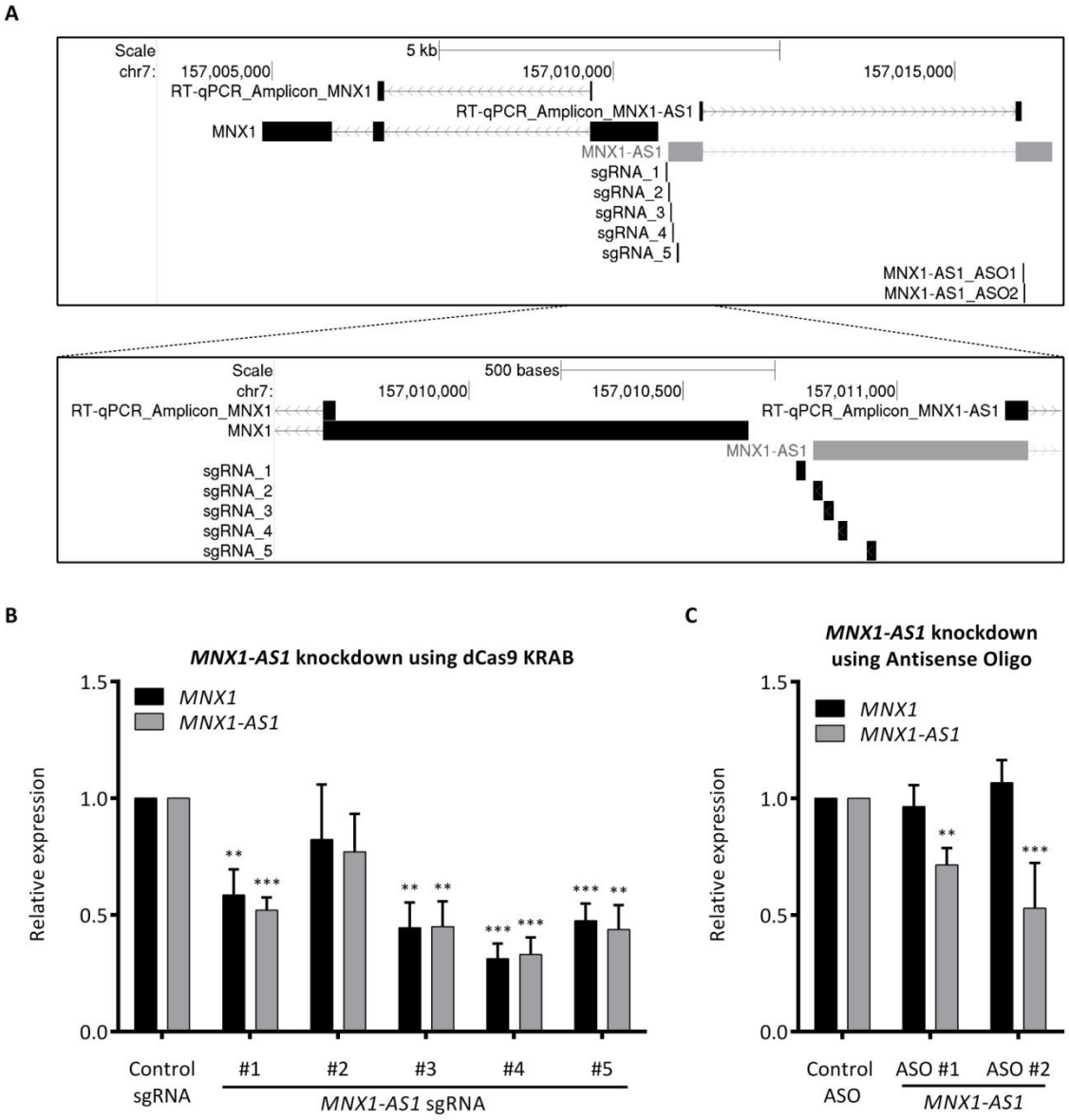


Figure 12: *MNX1-AS1* gene modulation using dCas9-KRAB and ASOs

A: Schematic representation of the *MNX1/MNX1-AS1* genomic locus depicting the sgRNAs and antisense oligos used to target *MNX1-AS1* as well as the RT-qPCR amplicons used to detect *MNX1* and *MNX1-AS1* expression.

B: NCI-H460 cells expressing dCas9-KRAB were transduced with either a control sgRNA or one of the five indicated sgRNAs targeting *MNX1-AS1*. RT-qPCR results for *MNX1* and *MNX1-AS1* normalized to *Cyclophilin* A and control sgRNA. Error bars represent SD (n=4). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test.

C: NCI-H460 cells were transfected with either a control ASO or two independent ASOs against *MNX1-AS1*. RT-qPCR results for *MNX1* and *MNX1-AS1* normalized to *Cyclophilin* A and control ASO. Error bars represent SD (n=4). * p<0.05; ** p<0.01; *** p<0.001 compared to control ASO, unpaired two-sided t-test.

Data used here has been published as a part of the following manuscript: Goyal et al., Nucleic Acids Res, 2016[116].

3.1.6. *HOTAIR* knockdown using dCas9-KRAB affects *HOXC11* expression

HOTAIR is a 2.2 kilobase pair (kb) long lncRNA that is transcribed in an antisense direction to the *HOXC* gene cluster and recruits the PRC2 complex to silence the expression of the *HOXD* locus *in trans*[30, 129]. *HOTAIR* is

expressed from two different TSSs. The α-TSS lies inside the intron of the *HOXC11*, separated from its TSS by roughly 1.8 kb. The β-TSS lies roughly 4 kb upstream of the *HOXC11* TSS and accounts for the majority of *HOTAIR* transcription (**Figure 13A**). *HOTAIR* deletion was previously shown to have no impact on *HOXC11* expression[130]. *HOTAIR* knockdown using an siPOOL targeting *HOTAIR* also had no impact on the *HOXC11* expression. Since *HOTAIR*β and *HOTAIR*α have common exons, they were both knocked down upon siPOOL treatment (**Figure 13B**). dCas9-KRAB-mediated targeting of the α-TSS using three independent sgRNAs led to a very strong knockdown of *HOTAIR*α expression (**Figure 13C**). However, no such repression was observed for *HOTAIR*β indicating that dCas9-KRAB specifically represses transcription from the α-TSS without affecting the β-TSS. Importantly, simultaneous to the *HOTAIR*α knockdown, *HOXC11* was also repressed in these cells (**Figure 13C**). Thus, the *HOXC11* TSS which was nearly 1.8 kb upstream from the *HOTAIR*α TSS was repressed while targeting the *HOTAIR*α TSS, whereas the *HOTAIR*β TSS, which was 4 kb upstream from the *HOTAIR*α TSS was unaffected.

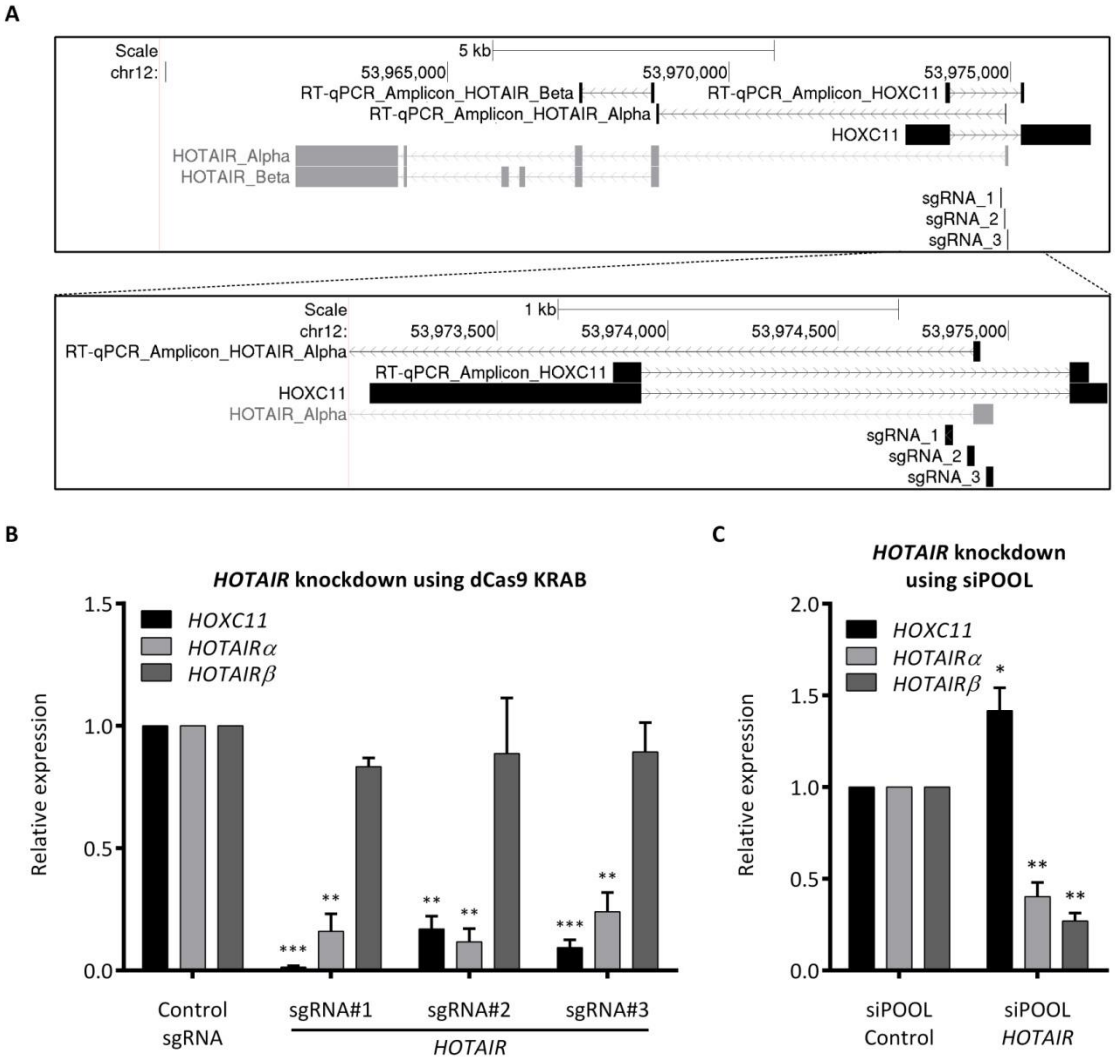


Figure 13: *HOTAIR* gene modulation using dCas9-KRAB and siPOOL

A: Schematic representation of the *HOXC11 / HOTAIR* genomic locus depicting the sgRNAs used to target *HOTAIR*α as well as the RT-qPCR amplicons used to detect *HOXC11* and *HOTAIR* expression.

B: Hela cells were transduced with lentidCas9-KRAB-PURO iv sgRNA containing either control sgRNA or one of the three indicated sgRNAs targeting *HOTAIR*α. RT-qPCR results for *HOXC11*, *HOTAIR*α, and *HOTAIR*β normalized to *Cyclophilin* A and control sgRNA. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test.
C: Hela cells were transfected with either siPOOL control or siPOOL *HOTAIR*. RT-qPCR results for *HOXC11*, *HOTAIR*α, and *HOTAIR*β normalized to *Cyclophilin* A and siPOOL control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to siPOOL control, unpaired two-sided t-test.
Data used here has been published as a part of the following manuscript: Goyal et al., Nucleic Acids Res, 2016[116].

In summary, these results indicated that dCas9-KRAB could simultaneously repress transcription from proximal promoter apart from its intended target advocating for caution while using CRISPRi to target lncRNAs as well as protein-coding genes. However, given that the *NOP14* promoter is more than 28 kb downstream of the *NOP14-AS1* promoter, this should not prohibit the use of dCas9-KRAB to knockdown *NOP14-AS1* expression and study its impact on *NOP14* expression.

3.2. *NOP14-AS1* and its impact on *NOP14* expression in DNA damage

3.2.1. Characterization of *NOP14-AS1* in DNA damage

Before the beginning of this dissertation project, lncRNA *NOP14-AS1* had already been identified to be induced in A549 as well as HepG2 cells upon treatment with DNA damaging drugs Etoposide, Cisplatin and Bleomycin using microarray expression analysis. In order to independently validate the results obtained from this microarray analysis, an RT-qPCR analysis was performed. *NOP14-AS1* along with the well-known p53 target gene *CDKN1A*[131] (*p21*) was induced in this RT-qPCR analysis further validating the results from the microarray analysis (**Figure 14**).

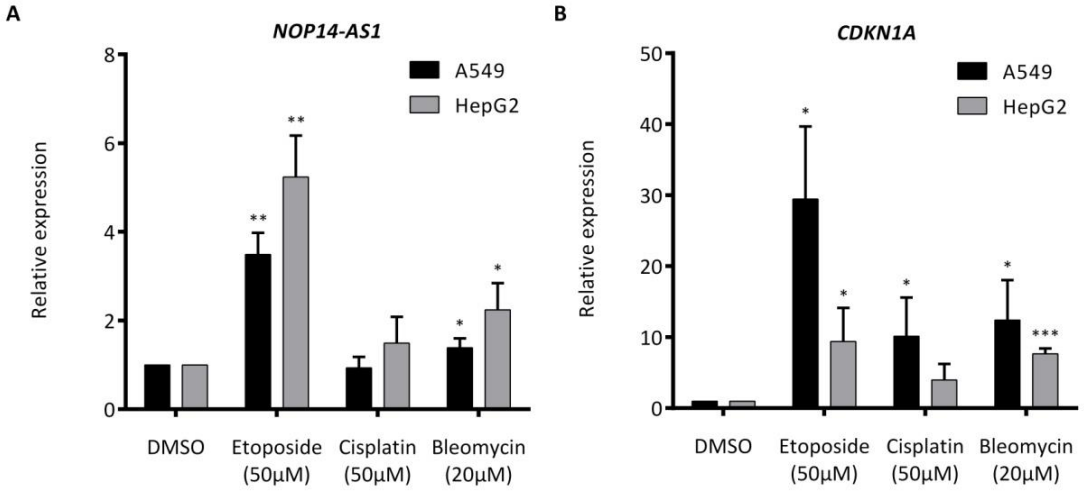


Figure 14: RT-qPCR validation of the microarray analysis
A549 and HepG2 cells were treated with 50 µM Etoposide (ETO) / 50 µM Cisplatin (CIS) / 20 µM Bleomycin (BLEO) or vehicle control DMSO for 8 hours. RT-qPCR results for (**A**) *NOP14-AS1* and (**B**) *CDKN1A* normalized to *Cyclophilin A* and DMSO controls. Error bars represent SD (n=4). * p<0.05; ** p<0.01; *** p<0.001 compared to DMSO control, unpaired two-sided t-test.
Data also used for a manuscript under review for publication: Goyal et al.

Analysis of the *NOP14-AS1* genomic loci in the UCSC genome browser[132] revealed that it could give rise to several splice isoforms (**Figure 15**, GENCODE transcripts). To establish the full-length sequence of *NOP14-AS1*, cDNA was extracted from HepG2 cells treated with Etoposide and was subjected to 5'- and 3'- Rapid Amplification of cDNA Ends (RACE). This analysis could detect most of the previously annotated *NOP14-AS1* splice isoforms as well as several novel variants (**Figure 15**, 5_RACE and 3_RACE). Moreover, these isoforms were 5'-m7G capped and 3'-polyadenylated as the applied RACE protocol could only detect such RNAs[133]. This was corroborated by publically available CAGE and polyA$^+$ RNA-Seq data for several cell lines (**Figure 15**)[134, 135]. Based on the results from RACE analysis, it was decided to use an RT-qPCR amplicon (Amplicon #2) which detected all the variants of *NOP14-AS1* for the subsequent experiments (**Figure 15**).

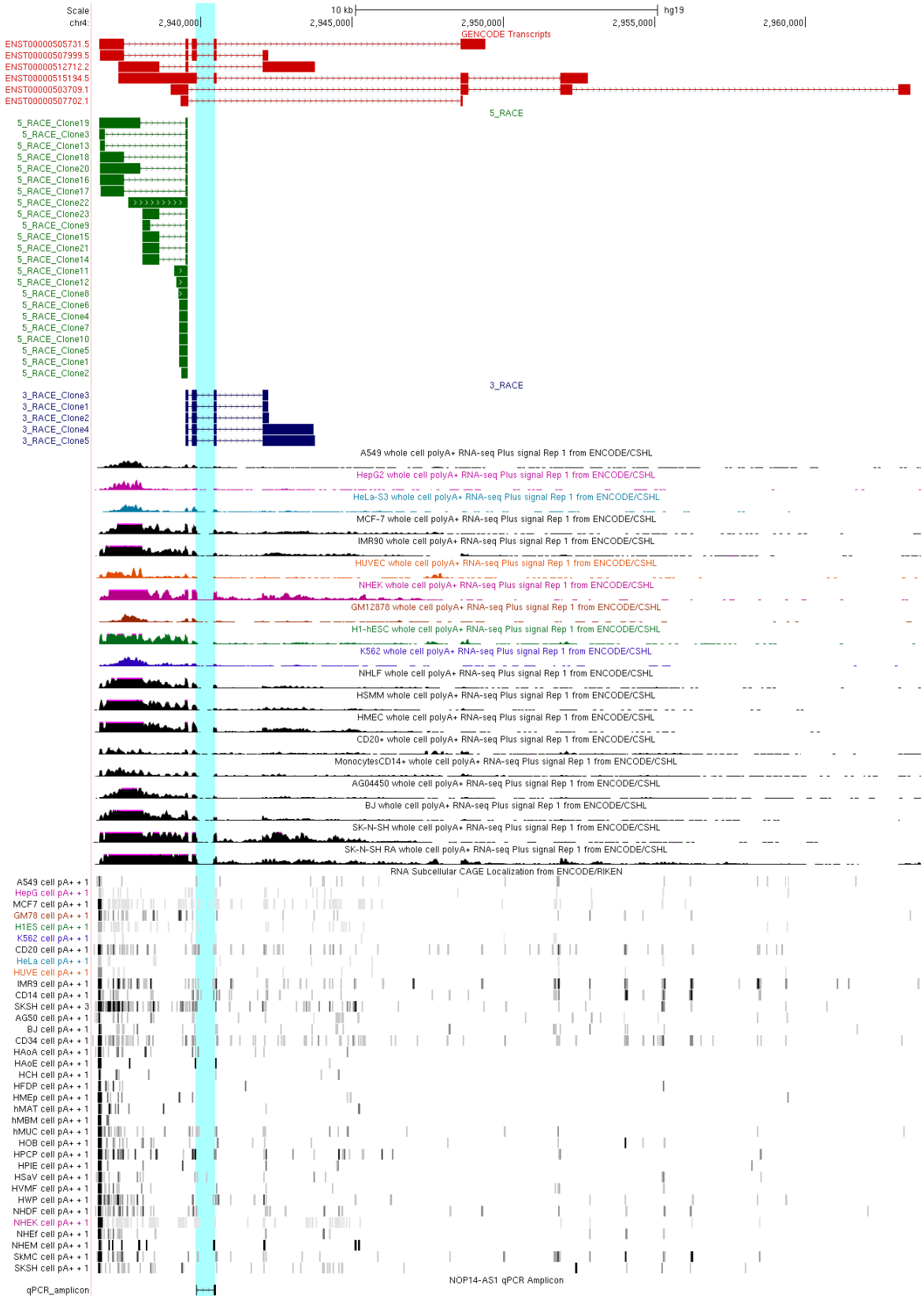


Figure 15: **Rapid amplification of cDNA ends identifies *NOP14-AS1* splice isoforms**
HepG2 cells were treated with 50 µM Etoposide for 16 hours. cDNA from these cells were subjected to 5' and 3' RACE for *NOP14-AS1*. Sanger sequencing results of the clones obtained from 5' RACE (in green) and 3' RACE (in blue) reactions aligned to the *NOP14-AS1* genomic locus in UCSC genome browser. Gencode-annotated transcripts are depicted in red. Strand-specific polyA$^+$ RNA seq as well CAGE data (only + strand data shown here) for the indicated cell lines was obtained from ENCODE data portal on UCSC genome browser (http://genome.ucsc.edu). RT-qPCR

amplicon used in this dissertation is also shown. All alignments were made to GRCh37/hg19 assembly and were performed using the UCSC Blat tool[119].
Data also used for a manuscript under review for publication: Goyal et al.

To identify additional cell line models and to investigate the drug-specific strengths of *NOP14-AS1* induction, four different cell lines were treated with a panel of five DNA damage-inducing reagents of clinical relevance. All cell lines showed DNA damage-specific induction of *NOP14-AS1*. Overall NCI-H460 (Lung Cancer cell line) showed the strongest induction of *NOP14-AS1*. Also, Etoposide and Doxorubicin (both inhibit Topoisomerase II and induce DSBs) were the strongest inducers of *NOP14-AS1* (**Figure 16A**). *CDKN1A* served as a positive control to check for functional induction of the DDR in these cells (**Figure 16B**). Based on these results, it was decided to use A549 and NCI-H460 (both lung cancer cell lines) in combination with Etoposide and Doxorubicin as the model system to study the role of *NOP14-AS1* in the DDR.

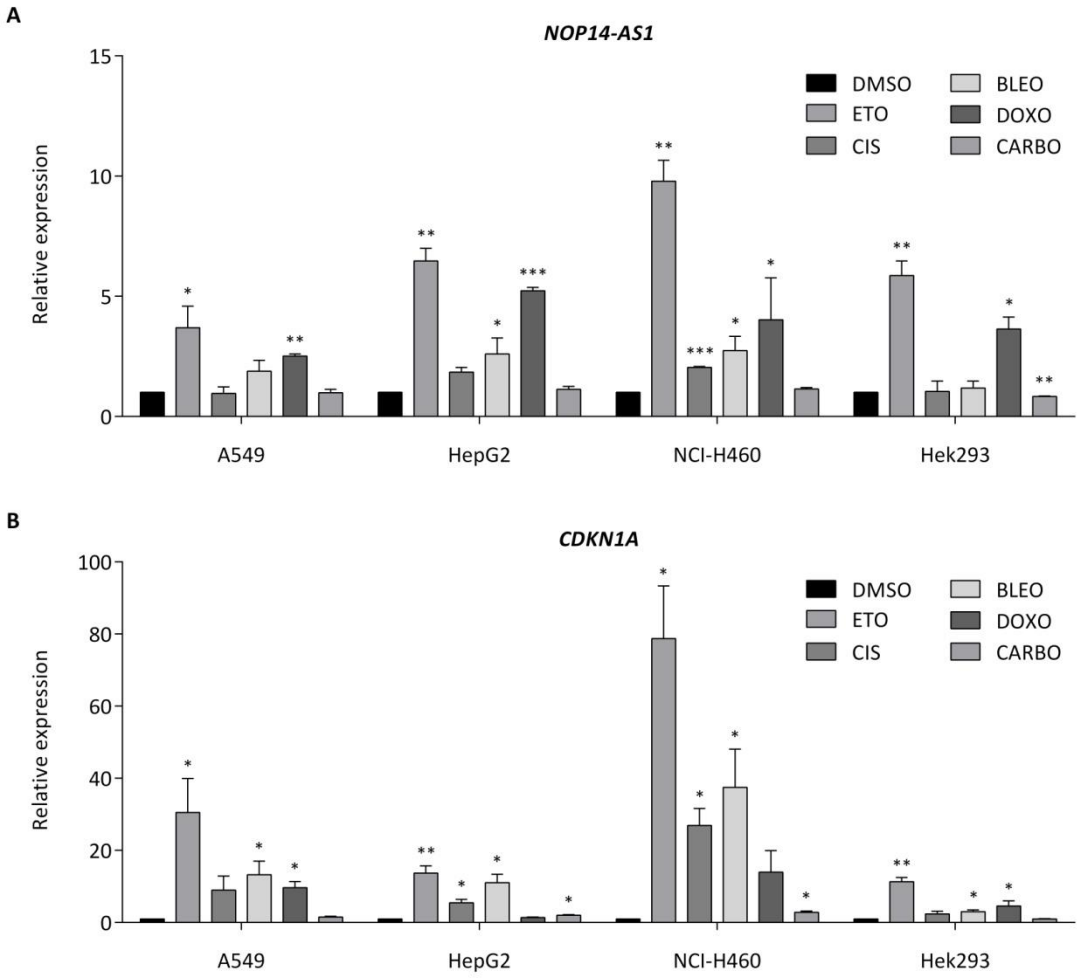


Figure 16: Etoposide and Doxorubicin are the strongest inducers of *NOP14-AS1*
A549 / HepG2 / NCI-H460 / Hek293 cells were treated with 50 µM Etoposide (ETO) / 50 µM Cisplatin (CIS) / 20 µM Bleomycin (BLEO) / 1 µM Doxorubicin (DOXO) / 20 µM Carboplatinum (CARBO) or vehicle control DMSO for 8 hours. RT-qPCR results for (A) *NOP14-AS1* and (B) *CDKN1A* normalized to *Cyclophilin A* and DMSO controls. Error bars represent SD (n≥3). * p<0.05; ** p<0.01; *** p<0.001 compared to DMSO control, unpaired two-sided t-test.
Data also used for a manuscript under review for publication: Goyal et al.

To uncover the mechanism of *NOP14-AS1* induction upon DNA damage, A549 cells were simultaneously treated with the transcriptional inhibitor Actinomycin D (ActD) as well as Doxorubicin. Transcriptional inhibition resulted in a loss of Doxorubicin-dependent *NOP14-AS1* induction, indicating that a transcriptional response gave rise to elevated *NOP14-AS1* levels upon DNA damage (**Figure 17A**). *CDKN1A*, which is known to be transcriptionally induced upon DNA damage in a p53-dependent manner[131], was used as a positive control for this experiment (**Figure 17B**).

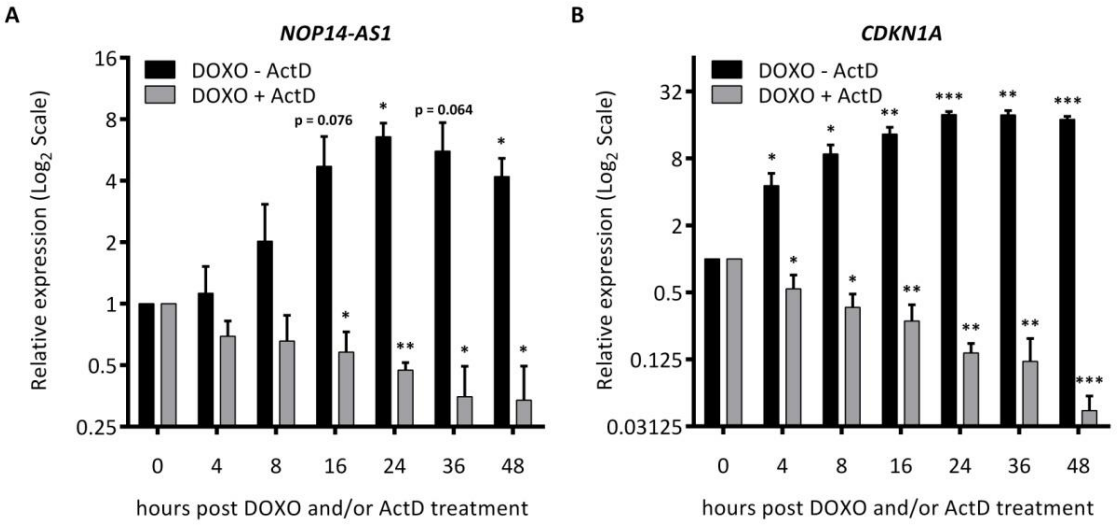


Figure 17: *NOP14-AS1* induction upon DNA damage is a transcriptional response
A549 cells were treated with 1 µM Doxorubicin (DOXO) and either 10 µg/ml ActD or vehicle control DMSO. RT-qPCR results for (**A**) *NOP14-AS1* and (**B**) *CDKN1A* normalized to *Cyclophilin A* and untreated controls. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to untreated controls, unpaired two-sided t-test.
Data also used for a manuscript under review for publication: Goyal et al.

Since the transcription factor p53 is responsible for activating several other lncRNAs involved in the DDR pathway[51-54], it was hypothesized that *NOP14-AS1* induction upon DNA damage could be p53-dependent. Nutlin-3 inhibits the interaction between MDM2 and p53 thereby stabilizing p53 and resulting in activation of p53 target genes[136]. Treatment of A549, NCI-H460, and HepG2 cells (p53 wild-type) with the MDM2 antagonist Nutlin-3 resulted in *NOP14-AS1* induction (**Figure 18A**). *CDKN1A* as a known p53 target gene was also induced upon Nutlin-3 treatment indicating the activation of p53 in these cells (**Figure 18B**). The *TP53* (-/-) HCT-116 colon cancer cell line is a frequently used p53 loss-of-function model with the parental *TP53* (+/+) HCT-116 cell line serving as the control. *NOP14-AS1* was upregulated in *TP53* (+/+) HCT-116 cells upon treatment with Etoposide, Doxorubicin or Nutlin-3. However, this induction was not observed in *TP53* (-/-) HCT-116 cells (**Figure 18C**). The p53 target gene *CDKN1A* served as a positive control (**Figure 18D**). Furthermore, *TP53* knockdown using dCas9-KRAB or siPOOL abrogated the *NOP14-AS1* induction upon Doxorubicin treatment in NCI-H460 cells (**Figure 22A,E**) suggesting that *NOP14-AS1* upregulation upon DNA damage induction was p53-dependent.

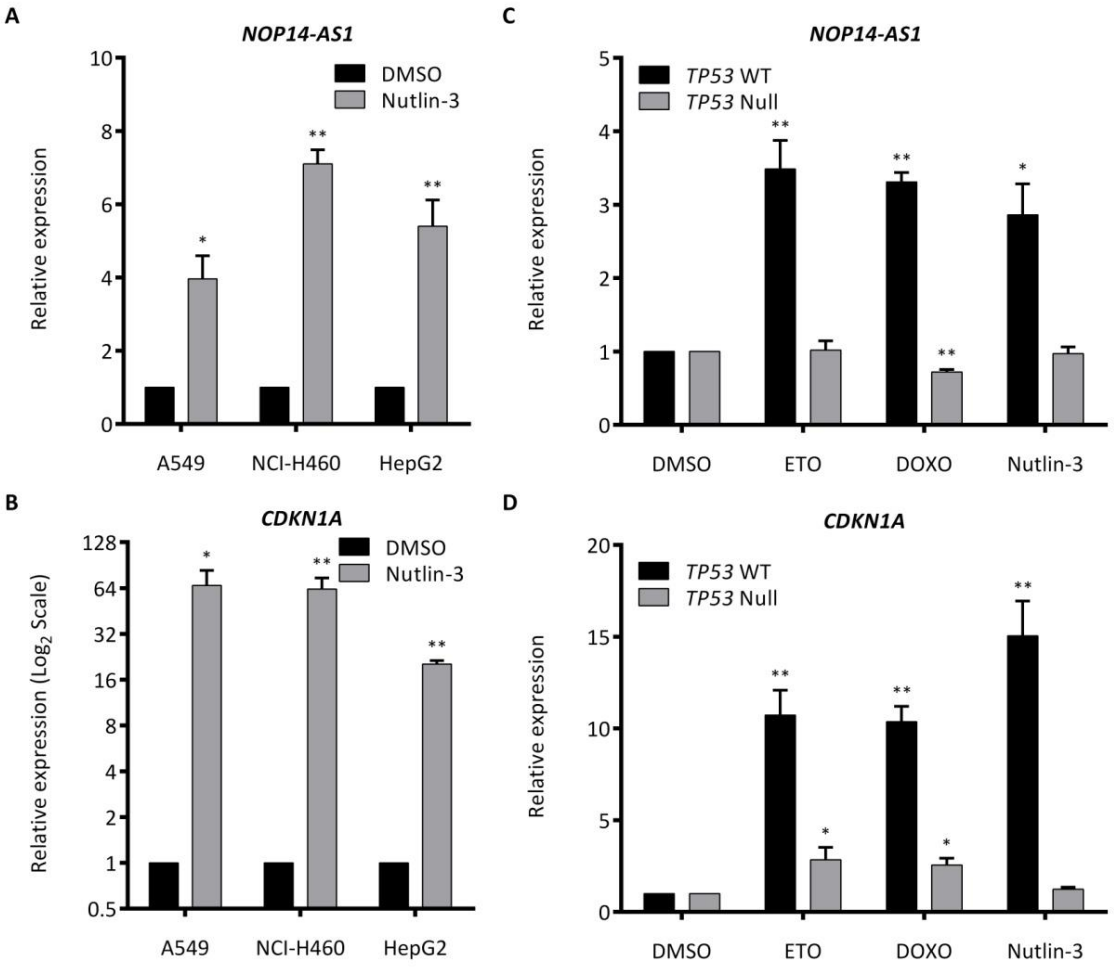


Figure 18: *NOP14-AS1* induction upon DNA damage is p53-dependent
A,B: A549 / NCI-H460 / HepG2 cells were treated with either 10 µM Nutlin-3 or vehicle control DMSO for 24 hours. RT-qPCR results for (A) *NOP14-AS1* and (C) *CDKN1A* normalized to *Cyclophilin A* and DMSO controls. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to DMSO controls, unpaired two-sided t-test.
C,D: HCT116 *TP53* WT / *TP53* Null cells were treated with 50 µM Etoposide (ETO) / 1 µM Doxorubicin (DOXO) / 10 µM Nutlin-3 or vehicle control DMSO for 12 hours. RT-qPCR results for (C) *NOP14-AS1* and (D) *CDKN1A* normalized to *Cyclophilin A* and DMSO controls. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to DMSO controls, unpaired two-sided t-test.
Data also used for a manuscript under review for publication: Goyal et al.

3.2.2. *NOP14-AS1* and *NOP14* are inversely regulated upon DNA damage

NOP14-AS1 is transcribed divergently from a bidirectional promoter of the protein-coding gene *MFSD10* and overlaps with its antisense protein-coding gene *NOP14* (**Figure 19A**). Divergent and antisense transcripts often mediate their function by regulating the expression of their neighboring genes *in cis*[19, 109, 137]. To check whether *MFSD10* or *NOP14* were regulated upon DNA damage induction, a time-course expression analysis of these genes in NCI-H460 cells treated with Doxorubicin was performed. *NOP14-AS1* expression gradually increased and peaked at 24 hours post-Doxorubicin treatment, thereafter gradually decreasing (**Figure 19B**). Inversely to *NOP14-AS1*, *NOP14* mRNA levels gradually decreased reaching a minimum at 16 hours, thereafter gradually increasing to reach basal levels again at 48 hours (**Figure 19B**). NOP14 protein expression was also reduced

reaching a minimum at 36 hours, thereafter showing a slight increase at 48 hours (**Figure 19C,D**). This analysis revealed a strong and statistically significant negative correlation between *NOP14-AS1* and *NOP14* mRNA (Pearson correlation = -0.848, p-value = 0.016) as well as protein expression (Pearson correlation = -0.851, p-value = 0.015) upon DNA damage induction. On the other hand, *MFSD10* expression was unchanged upon doxorubicin treatment and only a non-significant positive correlation was observed between *NOP14-AS1* and *MFSD10* mRNA expression (Pearson correlation = 0.495, p-value = 0.259) (**Figure 19E**). No inferences could be drawn for MFSD10 protein expression as it was undetectable using two independent antibodies.

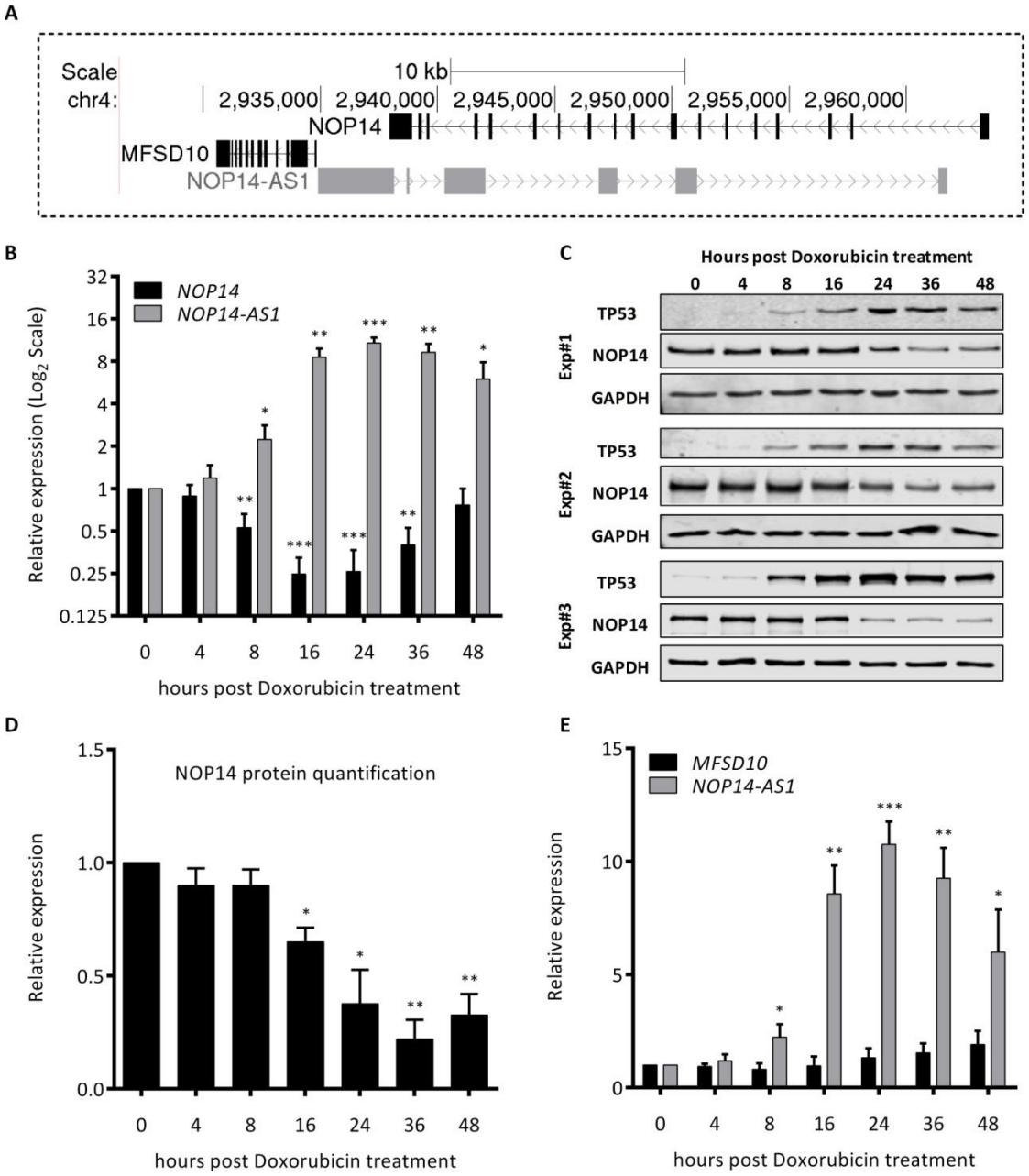


Figure 19: *NOP14-AS1* and *NOP14* are inversely co-regulated upon DNA damage in NCI H460 cells
A: *NOP14-AS1* genomic locus as depicted in UCSC genome browser. The *NOP14-AS1* gene is divergently expressed from its upstream neighbor, the *MFSD10* gene, and overlapping tail-to-tail with the antisense *NOP14* gene.

B-E: NCI-H460 cells were treated with 1 µM Doxorubicin for the indicated time points. (B) RT-qPCR results for *NOP14-AS1* and *NOP14* normalized to *Cyclophilin A* and untreated control. Error bars represent SD (n=4). * p<0.05; ** p<0.01; *** p<0.001 compared to untreated control, unpaired two-sided t-test. (C) Western blot results for p53 and NOP14. GAPDH was used as a loading control. (D) Quantification of the NOP14 protein expression normalized to GAPDH and untreated control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to untreated control, unpaired two-sided t-test. (E) RT-qPCR results for *NOP14-AS1* (same as Figure 19B, data given for comparison) and *MFSD10* normalized to *Cyclophilin A* and untreated control. Error bars represent SD (n≥3). * p<0.05; ** p<0.01; *** p<0.001 compared to untreated control, unpaired two-sided t-test.
Data also used for a manuscript under review for publication: Goyal et al.

To further corroborate the Doxorubicin-induced inverse co-regulation observed in NCI-H460 between *NOP14-AS1* and *NOP14*, time-course expression analyses of these genes in NCI-H460 cells treated with Etoposide and Cisplatin were performed. Etoposide induced similar changes in expression for *NOP14-AS1* as well as *NOP14* as were observed upon Doxorubicin treatment (Pearson correlation = -0.875, p-value = 0.01) (**Figure 20A**). Treatment with Cisplatin, on the other hand, resulted in a gradual increase of *NOP14-AS1* expression peaking at 48 hours post treatment, which was the end point of this analysis (**Figure 20B**). Strikingly, *NOP14* expression gradually decreased and reached a minimum at 48 hours post treatment (Pearson correlation = -0.647, p-value = 0.116).

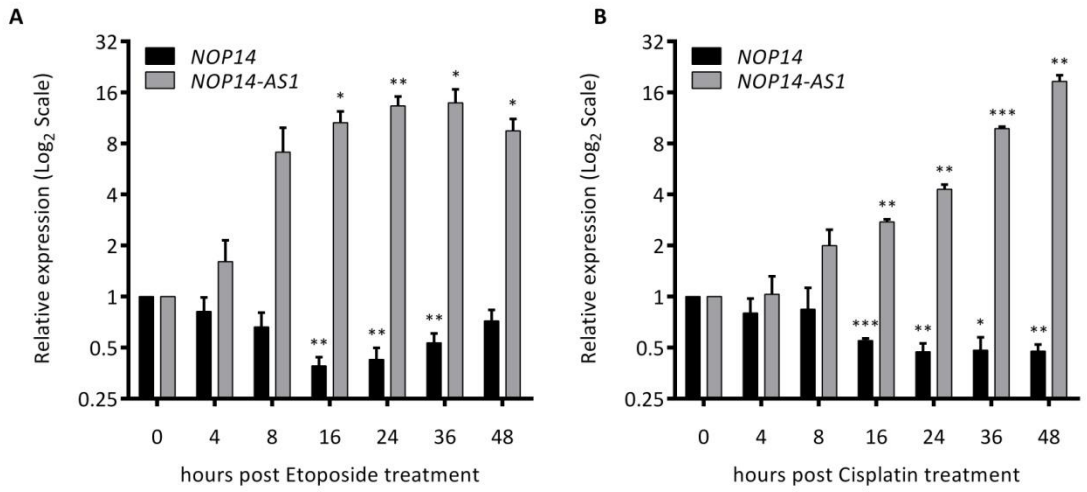


Figure 20: Etoposide- / Cisplatin-induced inverse co-regulation of *NOP14-AS1* and *NOP14* in NCI-H460 cells
A,B: NCI-H460 cells were treated with (A) 50 µM Etoposide or (B) 50 µM Cisplatin for the indicated time points. RT-qPCR results for *NOP14-AS1* and *NOP14* normalized to *Cyclophilin A* and untreated control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to untreated control, unpaired two-sided t-test.
Data also used for a manuscript under review for publication: Goyal et al.

The inverse co-regulation of the sense-antisense pair of *NOP14* and *NOP14-AS1* upon DNA damage induction was not restricted to NCI-H460 cells. A time-course expression analysis of these genes in A549 cells treated with Doxorubicin or Etoposide also revealed inverse co-regulation between them (Pearson correlation = -0.810, p-value = 0.027 for Doxorubicin time course; Pearson correlation = -0.616, p-value = 0.140 for Etoposide time course) (**Figure 21A,B**). A time-course expression analysis of these genes in HepG2 cells treated with Doxorubicin or Etoposide again revealed an inverse co-regulation between them (Pearson correlation = -0.740,

p-value = 0.092 for Doxorubicin time course; Pearson correlation = -0.538, p-value = 0.270 for Etoposide time course) (**Figure 21C,D**).

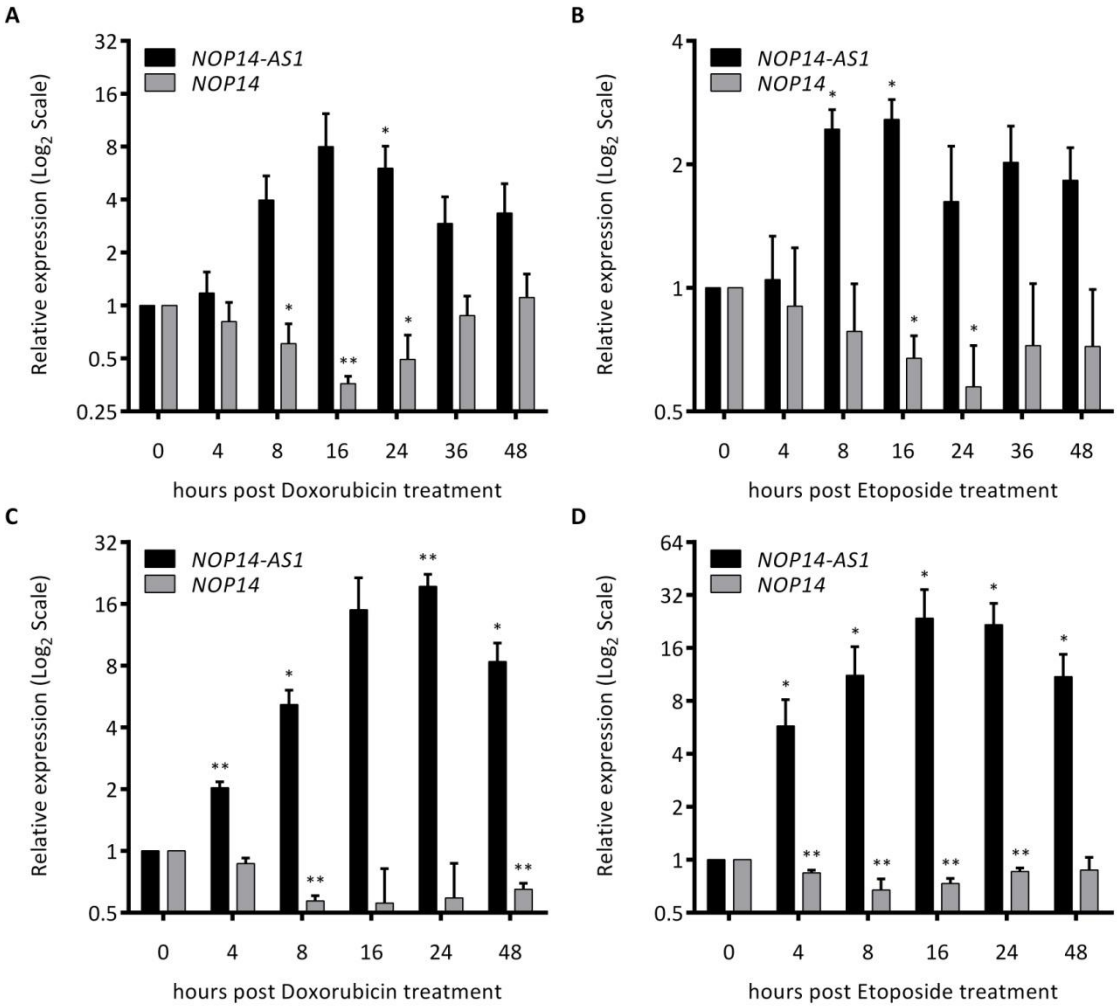


Figure 21: *NOP14-AS1* and *NOP14* are inversely co-regulated upon DNA damage in A549 and HepG2 cells
A,B: A549 cells were treated with (A) 1 µM Doxorubicin or (B) 50 µM Etoposide for the indicated time points. RT-qPCR results for *NOP14-AS1* and *NOP14* normalized to *Cyclophilin A* and untreated control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to untreated control, unpaired two-sided t-test.
C,D: HepG2 cells were treated with (C) 1 µM Doxorubicin or (D) 50 µM Etoposide for the indicated time points. RT-qPCR results for *NOP14-AS1* and *NOP14* normalized to *Cyclophilin A* and untreated control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to untreated control, unpaired two-sided t-test.
Data also used for a manuscript under review for publication: Goyal et al.

Since *NOP14-AS1* was induced upon DNA damage in a p53-dependent manner, it was analyzed whether the DNA damage induced *NOP14* repression was also p53-dependent. A p53 loss-of-function model in NCI-H460 was generated using the earlier described dCas9-KRAB method. sgRNA#2, which was the most efficient sgRNA at repressing *TP53* expression was used to generate a stable knockdown cell line in NCI-H460. *TP53* silencing resulted in a loss of *NOP14-AS1* induction as well as a partial rescue of *NOP14* mRNA and protein expression upon Doxorubicin treatment (**Figure 22A-D**). The p53 target gene *CDKN1A* served as a positive control (**Figure 22E**). Similar results were obtained in A549 cells (**Figure 22F,G**). As an alternative strategy to knockdown *TP53*,

NCI-H460 cells were transfected with an siPOOL against *TP53*. *TP53* knockdown using an siPOOL again resulted in the loss of *NOP14-AS1* induction as well as *NOP14* repression upon Doxorubicin treatment (**Figure 22H,I**).

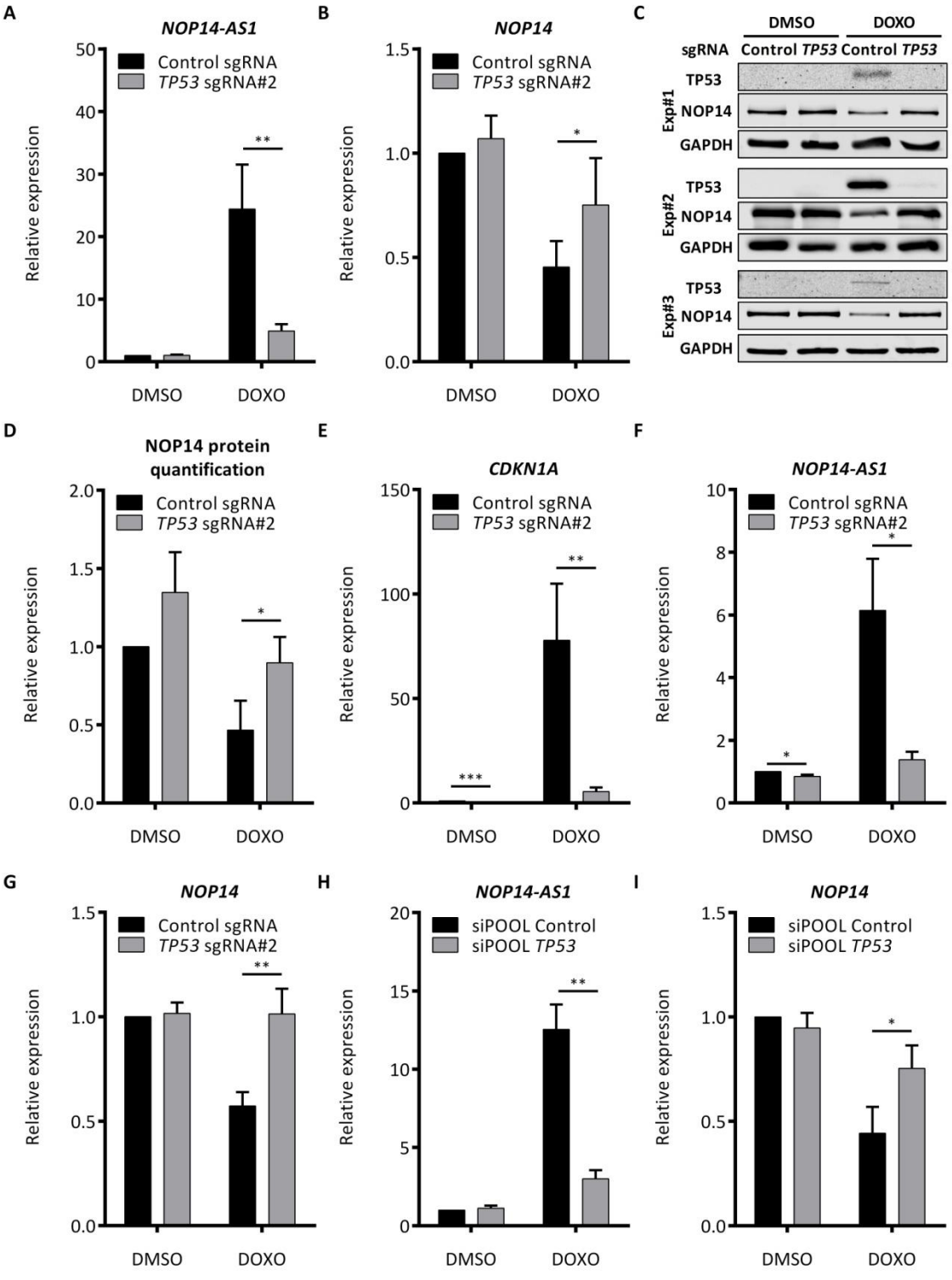


Figure 22: *NOP14-AS1* induction and *NOP14* repression upon DNA damage are p53-dependent

A-E: NCI-H460 cells were transduced with lentidCas9-KRAB-PURO iv sgRNA containing either a control sgRNA or a sgRNA targeting *TP53*. These were treated with either 1 µM Doxorubicin (DOXO) or vehicle control DMSO for 24 hours. (A) RT-qPCR results for *NOP14-AS1* and (B) *NOP14* normalized to *Cyclophilin A* and control sgRNA + DMSO control. Error bars represent SD (n=5). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-

sided t-test. (C) Western blot results for p53 and NOP14. GAPDH was used as a loading control. (D) Quantification of the NOP14 protein expression normalized to GAPDH and control sgRNA + DMSO control. (E) RT-qPCR results for *CDKN1A* normalized to *Cyclophilin A* and control sgRNA + DMSO control. Error bars represent SD (n=5). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test.
F-G: A549 cells were transduced with lentidCas9-KRAB-PURO iv sgRNA containing either a control sgRNA or a sgRNA targeting *TP53*. These were treated with either 1 µM Doxorubicin (DOXO) or vehicle control DMSO for 24 hours. RT-qPCR results for (F) *NOP14-AS1,* (H) *NOP14* normalized to *Cyclophilin A* and control sgRNA + DMSO control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test.
H,I: NCI-H460 cells were transfected with either siPOOL *TP53* or siPOOL control. These were treated with either 1 µM Doxorubicin (DOXO) or vehicle control DMSO for 24 hours. RT-qPCR results for (E) *NOP14-AS1* and (F) *NOP14* normalized to *Cyclophilin A* and siPOOL Control + DMSO control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to siPOOL Control, unpaired two-sided t-test.
Data also used for a manuscript under review for publication: Goyal et al.

Given the strong negative correlation between *NOP14-AS1* and *NOP14* expression in three different cell lines upon treatment with up to three different DNA damaging drugs, their co-dependence on p53 activation and their sense-antisense orientation, it was hypothesized that the *NOP14-AS1* lncRNA transcript or its transcription would regulate *NOP14* mRNA expression or vice versa as it had been proposed for several other examples of sense-antisense pairs [21, 22, 24, 32, 34, 138-144].

A loss-of-function and a gain-of-function model would be required to experimentally validate this hypothesis. To address any transcript-dependent effects, a standard siRNA or antisense oligonucleotide mediated knockdown of *NOP14-AS1* could serve as a loss-of-function model. On the hand a CRISPRi-based *NOP14-AS1* knockdown would have the additional advantage to shut down its transcription, thereby uncovering any transcription-dependent effects. Since the *NOP14-AS1* locus gives rise to several splice isoforms, overexpression of all the variants using plasmid-based approach would be inefficient. On the other hand, a CRISPRa-mediated activation of *NOP14-AS1* would induce all splice isoforms (at least from the targeted TSS).

3.2.3. *NOP14* does not regulate *NOP14-AS1* upon DNA damage

To analyze the impact of *NOP14* repression on *NOP14-AS1* expression, *NOP14* was knocked down in NCI-H460 cells using siRNAs. RNAi-induced loss of *NOP14* mRNA and protein did not result in any *NOP14-AS1* induction (**Figure 23A-D**). Treatment of these cells with Doxorubicin resulted in further repression of *NOP14* but no further increase in *NOP14-AS1* levels was observed (**Figure 23A-D**). This indicated that the increased *NOP14-AS1* expression upon doxorubicin treatment was not due to the reduced *NOP14* expression. However, it was observed that the *CDKN1A* mRNA and protein expression was elevated upon *NOP14* knockdown (**Figure 23B,E**). Similar results were obtained in A549 cells (**Figure 24**), except that *NOP14* knockdown using one of the two siRNAs resulted in decreased *NOP14-AS1* induction upon Doxorubicin treatment (**Figure 24D**). However, since both siRNAs were equally effective in knocking down *NOP14* expression (**Figure 24A-C**), this effect on *NOP14-AS1* was most likely an off-target effect of this particular siRNA.

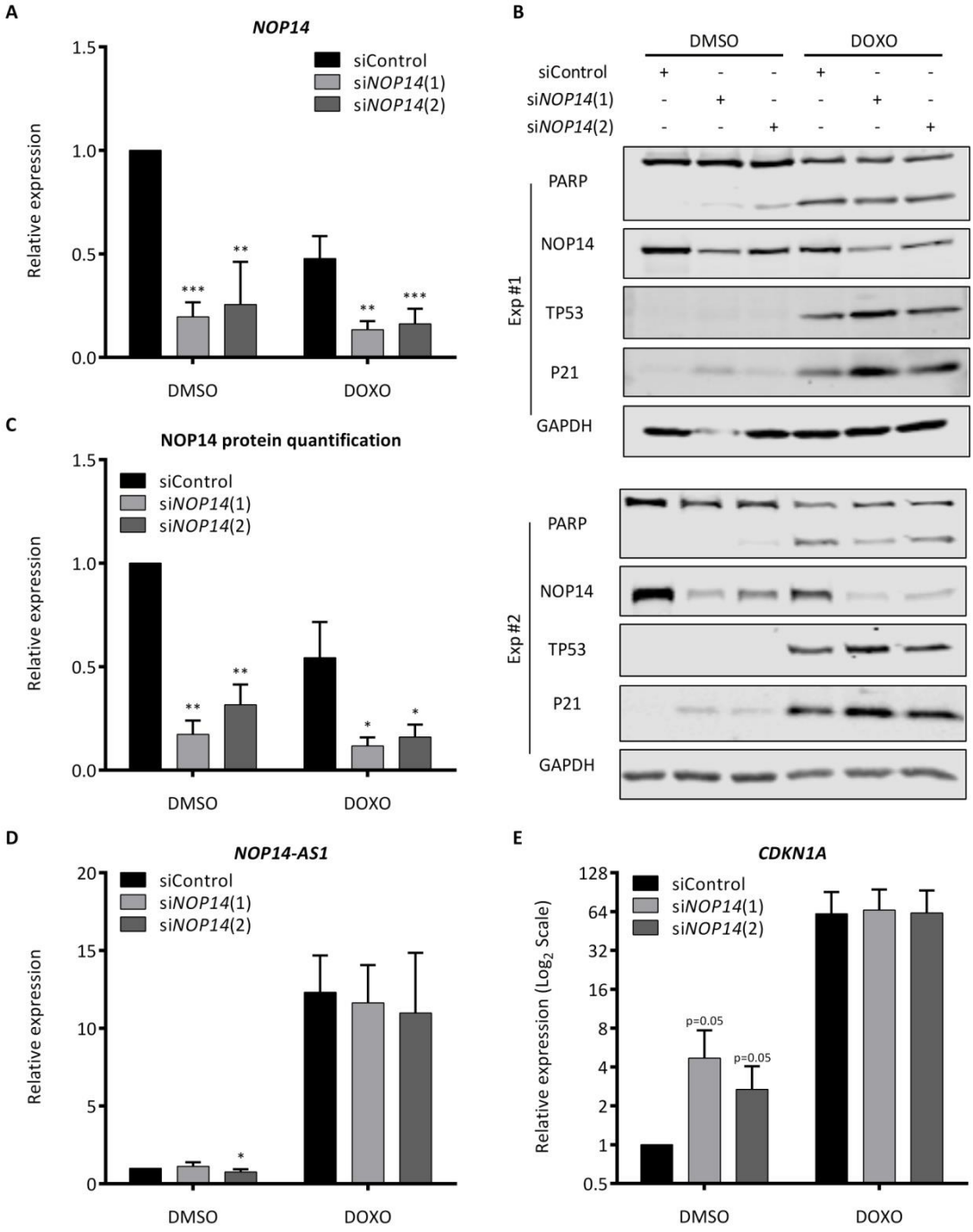


Figure 23: *NOP14* knockdown using siRNAs does not affect *NOP14-AS1* in NCI-H460 cells

NCI-H460 cells were transfected with either a control siRNA (siControl) or two independent siRNAs (si*NOP14*(1) and si*NOP14*(2)) targeting the *NOP14* mRNA. These were then treated with either 1 µM Doxorubicin (DOXO) or vehicle control DMSO for 24 hours. (A) RT-qPCR results for *NOP14* normalized to *Cyclophilin A* and siControl + DMSO control. Error bars represent SD (n=5). * p<0.05; ** p<0.01; *** p<0.001 compared to control siRNA, unpaired two-sided t-test. (B) Western blot results for PARP, NOP14, p53 and p21. GAPDH was used as a loading control. (C) Quantification of the NOP14 protein expression normalized to GAPDH and siControl + DMSO control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control siRNA, unpaired two-sided t-test. RT-qPCR results for (D) *NOP14-AS1* and (E) *CDKN1A* normalized to *Cyclophilin A* and siControl + DMSO control. Error bars represent SD (n=5). * p<0.05; ** p<0.01; *** p<0.001 compared to control siRNA, unpaired two-sided t-test.

Data also used for a manuscript under review for publication: Goyal et al.

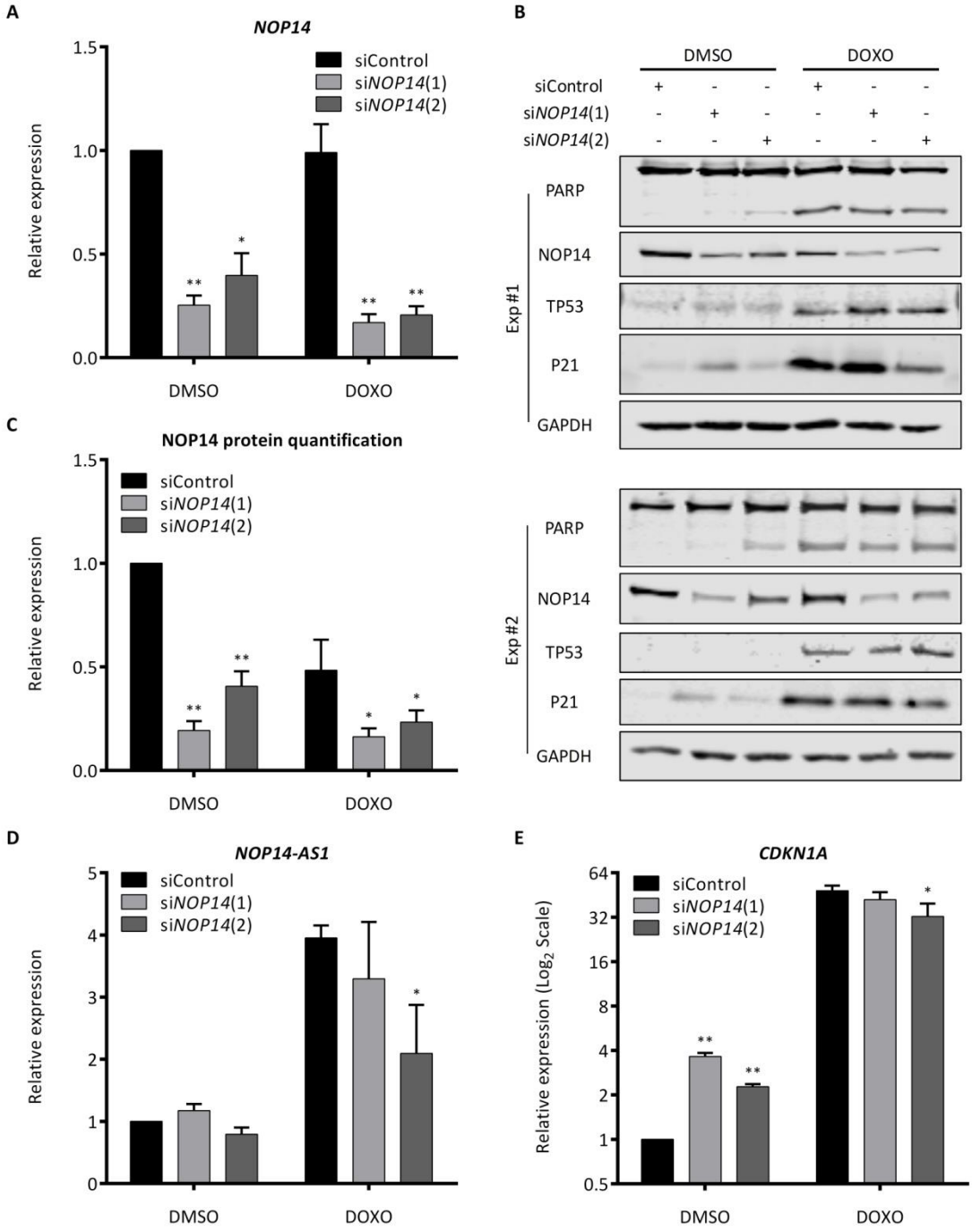


Figure 24: *NOP14* knockdown using siRNAs does not affect *NOP14-AS1* in A549 cells

A549 cells were transfected with either a control siRNA (siControl) or two independent siRNAs (si*NOP14*(1) and si*NOP14*(2)) targeting the *NOP14* mRNA. These were treated with either 1 µM Doxorubicin (DOXO) or vehicle control DMSO for 24 hours. (A) RT-qPCR results for *NOP14* normalized to *Cyclophilin A* and siControl + DMSO control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control siRNA, unpaired two-sided t-test. (B) Western blot results for PARP, NOP14, p53 and p21. GAPDH was used as a loading control. (C) Quantification of the NOP14 protein expression normalized to GAPDH and siControl + DMSO control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control siRNA, unpaired two-sided t-test. RT-qPCR results for (D) *NOP14-AS1* and (E) *CDKN1A* normalized to *Cyclophilin A* and siControl + DMSO control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control siRNA, unpaired two-sided t-test.

Data also used for a manuscript under review for publication: Goyal et al.

Since the RNAi-mediated knockdown of *NOP14* could only rule out the possibility of any *in trans*-regulatory effects of *NOP14* on *NOP14-AS1*, dCas9-KRAB was used to suppress the transcription of *NOP14* from its endogenous promoter to uncover any *in cis*-regulatory effects. Indeed a dCas9-KRAB-mediated repression of *NOP14* using two independent sgRNAs targeting its promoter resulted in a small but statistically significant induction (4-fold) of *NOP14-AS1* steady-state expression (**Figure 25A-D**).

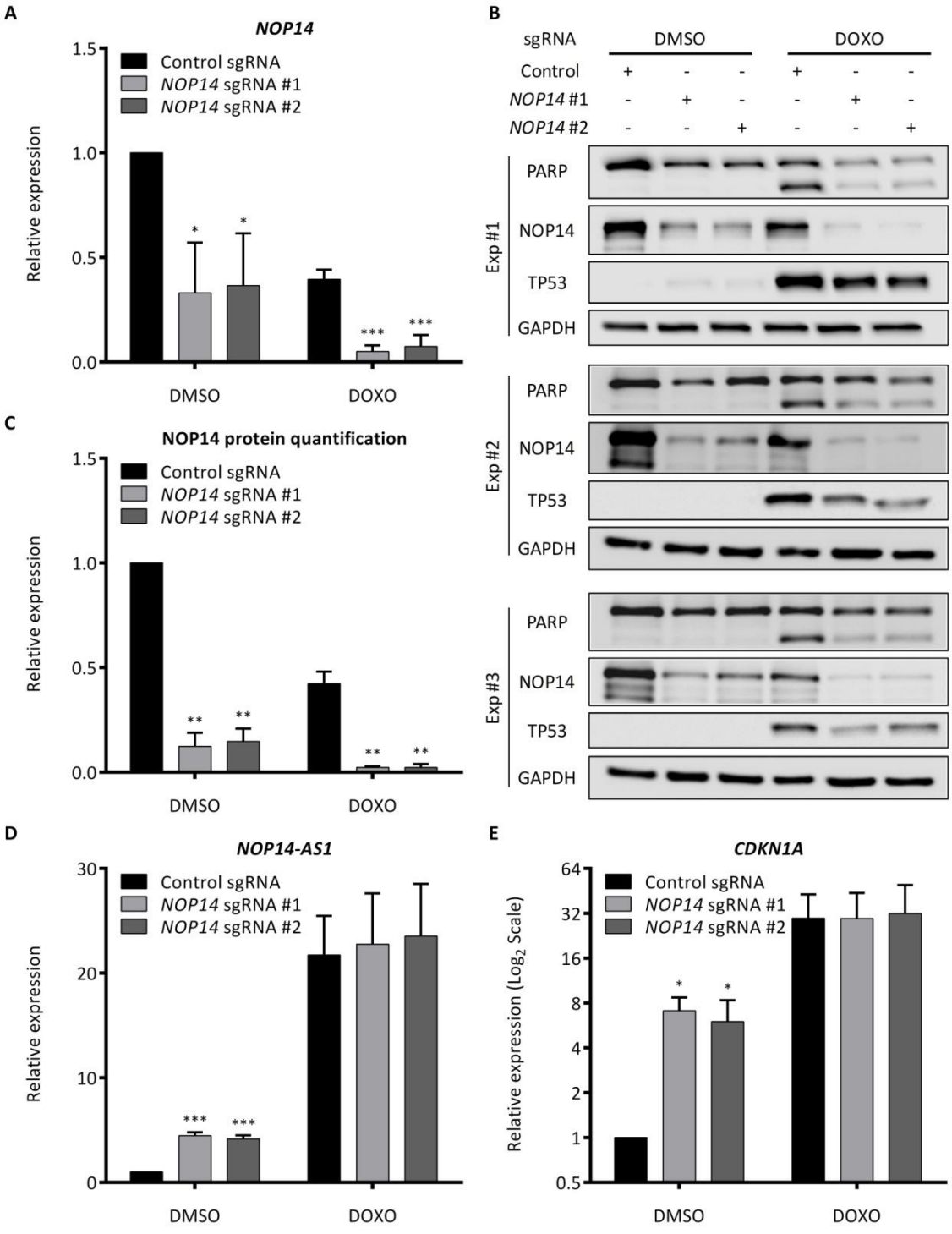


Figure 25: *NOP14* does not regulate *NOP14-AS1* upon DNA damage

NCI-H460 cells were transduced with lentidCas9-KRAB-PURO iv sgRNA containing either a control sgRNA or one of the two indicated sgRNAs targeting *NOP14*. These were then treated with either 1 µM Doxorubicin (DOXO) or vehicle control DMSO for 24 hours. (A) RT-qPCR results for *NOP14* normalized to *Cyclophilin A* and control sgRNA + DMSO control. Error bars represent SD (n=4). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test. (B) Western blot results for PARP, NOP14, and p53. GAPDH was used as a loading control. (C) Quantification of the NOP14 protein expression normalized to GAPDH and control sgRNA + DMSO control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test. RT-qPCR results for (D) *NOP14-AS1* and (E) *CDKN1A* normalized to *Cyclophilin A* and control sgRNA + DMSO control. Error bars represent SD (n=4). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test.
Data also used for a manuscript under review for publication: Goyal et al.

However, this induction was much weaker than the Doxorubicin-mediated increase of *NOP14-AS1* (22-fold) in these cells (**Figure 25D**). Also, the decrease of *NOP14* by CRISPRi and DNA damage led to comparable *NOP14* levels, but to vastly different *NOP14-AS1* levels (**Figure 25A-D**). Moreover, no significant further increase in *NOP14-AS1* levels upon *NOP14* knockdown was observed in the Doxorubicin-treated cells (**Figure 25D**), altogether indicating that an independent mechanism was responsible for *NOP14-AS1* induction upon DNA damage.

Interestingly, p53 levels were slightly elevated in the *NOP14* knockdown cells (**Figure 25B,** Exp#1) but upon Doxorubicin treatment in these cells, p53 expression was consistently less than what was observed in the control cells (**Figure 25B**) indicating a role in the p53 pathway. PARP cleavage, which served as an indicator for apoptosis induction was also reduced upon *NOP14* knockdown in Doxorubicin-treated cells (**Figure 25B**). But since the uncleaved PARP levels were also affected in the DMSO-treated cells, no conclusions could be drawn about the effect of *NOP14* on Doxorubicin-induced apoptosis.

In line with the results from the siRNA-mediated knockdown of *NOP14*, *CDKN1A* was also induced upon dCas9-KRAB-mediated *NOP14* knockdown (**Figure 25E**) which could be explained by the mild increase in p53 levels in *NOP14* knockdown cells (**Figure 25B,** Exp#1). Simultaneous knockdown of *TP53* in the *NOP14* CRISPRi knockdown cells resulted in a complete loss of *CDKN1A* induction, indicating that the *CDKN1A* induction observed upon dCas9-KRAB-mediated *NOP14* knockdown was indeed due to p53 stabilization (**Supplementary Figure 34B**). *TP53* knockdown, however, could not prevent the *NOP14-AS1* induction observed upon dCas9-KRAB-mediated *NOP14* knockdown, indicating that this was a p53-independent event (**Supplementary Figure 34C**).

It was also observed that *NOP14* knockdown resulted in severe proliferation defects and the stable *NOP14* knockdown cell lines stopped proliferating beyond second passage in cell culture (**Figure 26**) indicating a role of *NOP14* in cell proliferation. The reduced proliferation upon *NOP14* knockdown could be due to cell cycle arrest induced by increased *CDKN1A* expression in these cells.

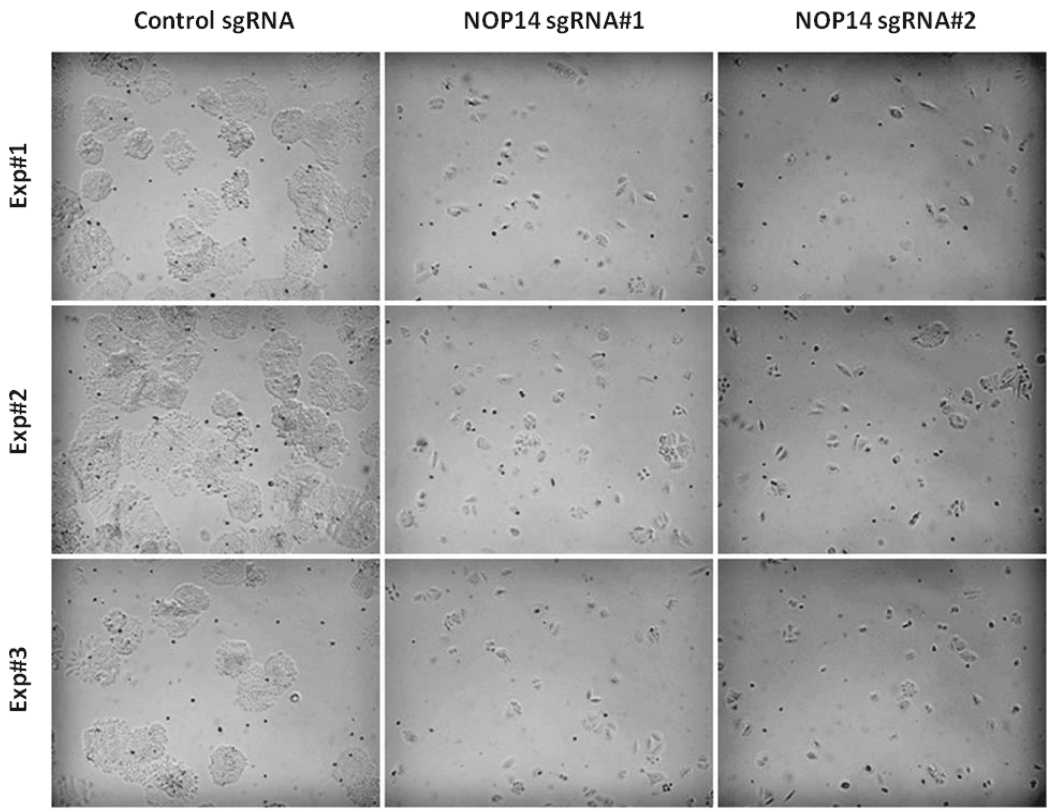


Figure 26: *NOP14* knockdown results in growth arrest
NCI-H460 cells were transduced with lentidCas9-KRAB-PURO iv sgRNA containing either a control sgRNA or one of the two indicated sgRNAs targeting *NOP14*. Equal numbers of cells for the each stable cell line were seeded. Brightfield images from three independent experiments 6 days post seeding.

To test whether the small induction of *NOP14-AS1* upon dCas9-KRAB-mediated *NOP14* knockdown was a specific and rescuable effect, *NOP14* was exogenously expressed from a lentiviral plasmid in the *NOP14* knockdown cells. *NOP14* cDNA from NCI-H460 cells was cloned into a lentiviral plasmid downstream of a constitutive promoter (EF1α) and transduced into *NOP14* knockdown cells. Since the sgRNA binding site was absent in the lentiviral plasmid containing the *NOP14* cDNA, it could rescue *NOP14* expression in these cells even though the transcription from the endogenous *NOP14* promoter continued to be repressed (**Figure 27A-C**).

Ectopic expression of *NOP14* expression resulted in only a partial reversal of the *NOP14-AS1* induction observed upon *NOP14* knockdown using dCas9-KRAB (**Figure 27D**). The *CDKN1A* induction observed upon *NOP14* knockdown could also be reversed completely (**Figure 27E**) indicating that *CDKN1A* was induced as a result of reduced *NOP14* mRNA or protein levels rather than its transcription. Similarly, the induction of p53 observed upon *NOP14* knockdown could also be reversed (**Figure 27B**). The reduced PARP levels were also restored further confirming a role of *NOP14* in regulating PARP levels inside these cells (**Figure 27B**). Taken together these data indicate that the small increase in *NOP14-AS1* expression observed with the CRISPRi approach was a result of *in cis*-regulation by transcriptional interference, but that the DNA damage-induced regulation of *NOP14-AS1* was independent of *NOP14* expression.

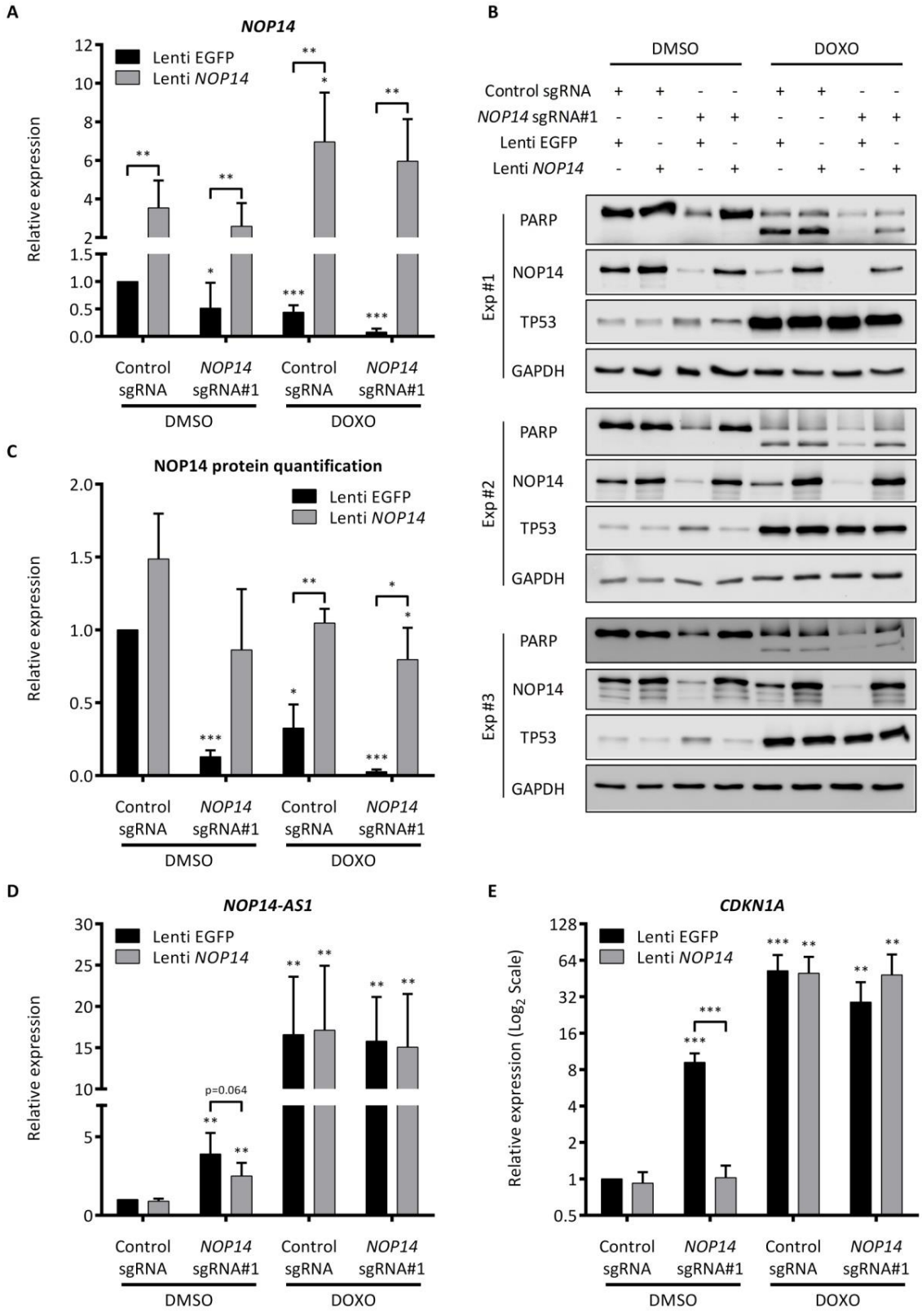


Figure 27: Ectopic rescue of *NOP14* expression does not reverse *NOP14-AS1* induction

A-D: NCI-H460 cells were transduced with lentidCas9-KRAB-PURO iv sgRNA containing either a control sgRNA or a sgRNA targeting *NOP14* in combination with either lenti EGFP Blast or lenti *NOP14* Blast. These were treated with either 1 µM Doxorubicin (DOXO) or vehicle control DMSO for 24 hours. (A) RT-qPCR results for *NOP14* normalized to *Cyclophilin A* and control sgRNA + DMSO control. Error bars represent SD (n=6). * p<0.05; ** p<0.01; *** p<0.001

compared to control sgRNA, unpaired two-sided t-test. (B) Western blot results for PARP, NOP14, and p53. GAPDH was used as a loading control. (C) Quantification of the NOP14 protein expression normalized to GAPDH and control sgRNA + DMSO control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test. RT-qPCR results for (D) *NOP14-AS1* and (E) *CDKN1A* normalized to *Cyclophilin A* and control sgRNA + DMSO control. Error bars represent SD (n=6). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test.
Data also used for a manuscript under review for publication: Goyal et al.

In turn, *NOP14-AS1* could regulate *NOP14* expression upon DNA damage thereby exerting its function in the DDR.

3.2.4. *NOP14-AS1* does not regulate *NOP14* upon DNA damage

To analyze the impact of the *NOP14-AS1* transcript on *NOP14* expression, *NOP14-AS1* was knocked down in NCI-H460 cells using two independent antisense LNA GapmeRs. The cells were treated with Doxorubicin or vehicle control (DMSO). Despite the reduced *NOP14-AS1* expression, no significant increase in *NOP14* mRNA or protein expression could be observed. Moreover, *NOP14-AS1* knockdown could not rescue the Doxorubicin-induced *NOP14* repression (**Figure 28A-D**) indicating that increased *NOP14-AS1* lncRNA transcript levels was not responsible for the *NOP14* regulation upon Doxorubicin treatment. Also, no changes in the PARP (cleaved or uncleaved) and p53 levels, as well as the CDKN1A expression, were observed upon *NOP14-AS1* knockdown (**Figure 28C,E**).

To establish whether *NOP14-AS1* can regulate *NOP14* expression *in cis* via transcription-dependent mechanisms, *NOP14-AS1* transcription was modulated by using the earlier established method of CRISPRi using dCas9-KRAB. From the initial panel of five sgRNAs used to repress *NOP14-AS1* (**Figure 7A**), sgRNA#1 and sgRNA#3 were chosen for dCas9-KRAB-mediated knockdown of *NOP14-AS1*. Additionally to the dCas9-KRAB-mediated loss-of-function model, a dCas9-VP160 mediated CRISPRa gain-of-function model was also established. To identify sgRNAs that can efficiently activate *NOP14-AS1* transcription, an initial screening using a panel of eight sgRNAs was performed. All sgRNAs (including sgRNA#3) that bound a region downstream of the *NOP14-AS1* TSS, failed to activate its transcription. The upstream sgRNA#1 and sgRNA#6 were the most efficient in activating *NOP14-AS1* transcription and were therefore used for the subsequent experiments (**Supplementary Figure 35**).

Repression or induction of *NOP14-AS1* transcription from its endogenous promoter using two independent sgRNAs was effective in *NOP14-AS1* regulation (**Figure 29A**), but did not result in any change in *NOP14* mRNA or protein expression. Moreover, *NOP14-AS1* transcriptional modulation using CRISPRi/CRISPRa had no impact on Doxorobucin induced *NOP14* repression (**Figure 29B-D**), indicating that the *NOP14* repression upon DNA damage was independent of *NOP14-AS1* lncRNA transcript as well as transcription. Also, no changes were observed in the PARP (cleaved or uncleaved), p53 and *CDKN1A* mRNA as well as protein expression upon *NOP14-AS1* knockdown. Similar results were obtained in A549 cells (**Figure 30**) further confirming that *NOP14-AS1* had no impact on *NOP14* mRNA or protein expression.

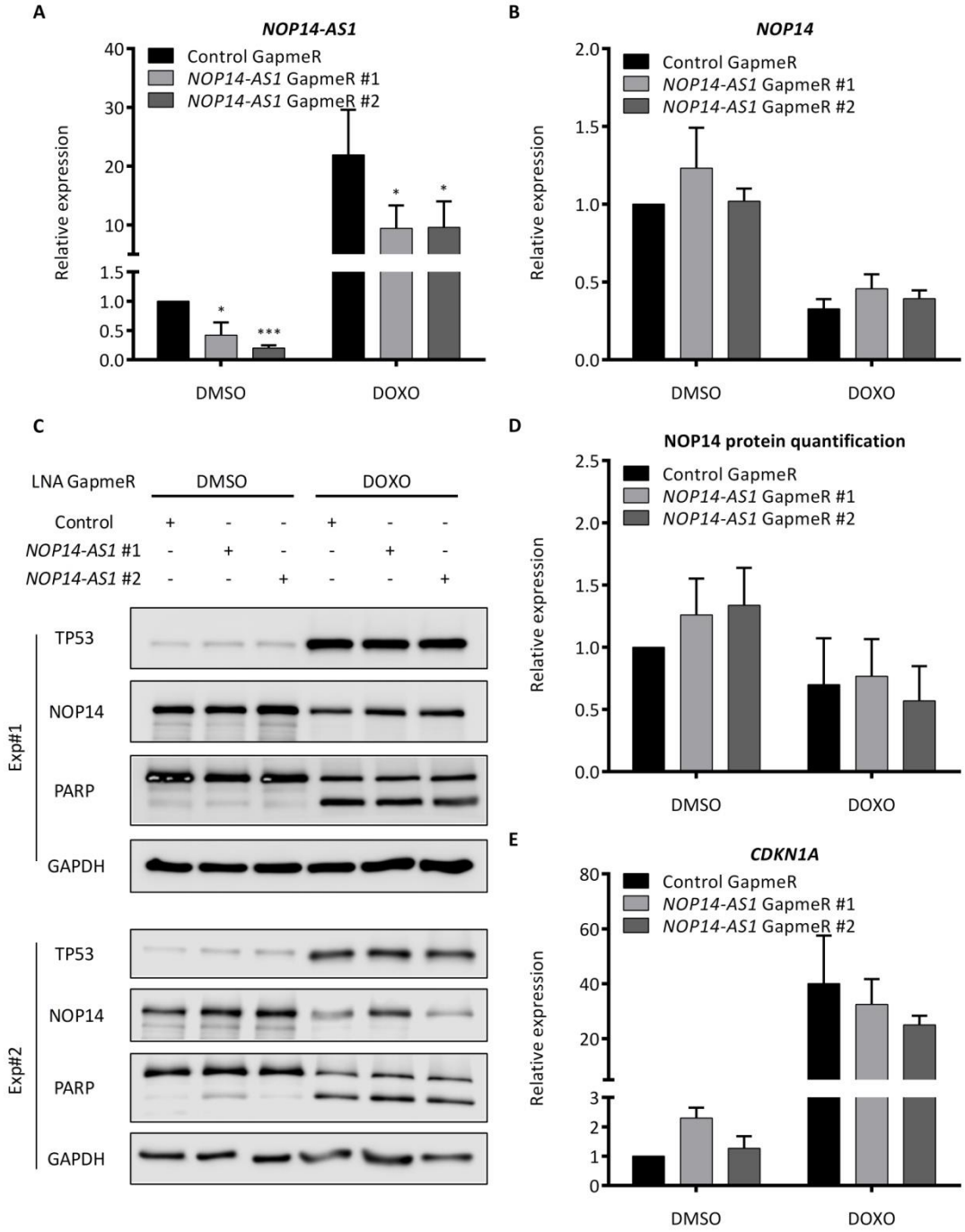


Figure 28: *NOP14-AS1* knockdown using LNA antisense GapmeRs

A-D: NCI-H460 cells were transfected with either a control GapmeR or one of the two indicated GapmeRs (GapmeR#1 and GapmeR#2) targeting the *NOP14-AS1* lncRNA. These were treated with either 1 µM Doxorubicin (DOXO) or vehicle control DMSO for 24 hours. (A) RT-qPCR results for *NOP14* normalized to *Cyclophilin A* and control GapmeR + DMSO control. Error bars represent SD (n≥3). * p<0.05; ** p<0.01; *** p<0.001 compared to control GapmeR, unpaired two-sided t-test. (B) Western blot results for PARP, NOP14, and p53. GAPDH was used as a loading control. (C) Quantification of the NOP14 protein expression normalized to GAPDH and control GapmeR + DMSO control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control GapmeR, unpaired two-sided t-test. RT-qPCR results for (D) *NOP14-AS1* and (E) *CDKN1A* normalized to *Cyclophilin A* and control GapmeR + DMSO control. Error bars represent SD (n≥3; except for *CDKN1A*, for which n≥2). * p<0.05; ** p<0.01; *** p<0.001 compared to control GapmeR, unpaired two-sided t-test.

Data also used for a manuscript under review for publication: Goyal et al.

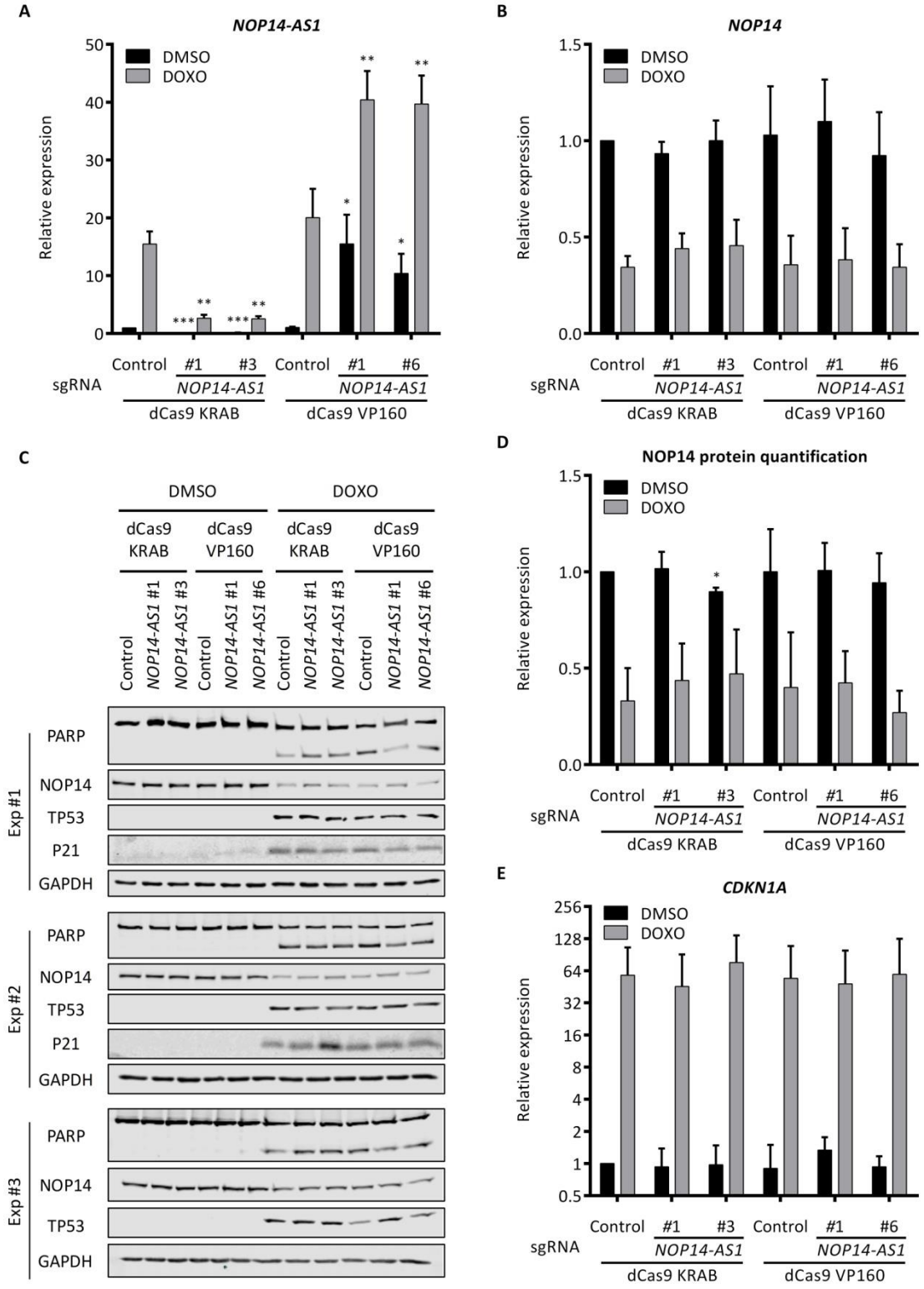


Figure 29: *NOP14-AS1* transcription modulation in NCI-H460 cells has no impact on *NOP14* expression

NCI-H460 cells were transduced with lentidCas9-KRAB-PURO iv sgRNA or lentidCas9-VP160-PURO iv sgRNA, respectively, containing either a control sgRNA or one of the two indicated sgRNAs targeting the *NOP14-AS1* promoter. These were then treated with either 1 µM Doxorubicin (DOXO) or vehicle control DMSO for 24 hours. (A) RT-qPCR results for *NOP14* normalized to *Cyclophilin A* and control sgRNA + DMSO controls. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test. (B) Western blot results for PARP, NOP14, and p53 and p21. GAPDH was used as a loading control. (C) Quantification of the NOP14

protein expression normalized to GAPDH and control sgRNA + DMSO control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test.
Data also used for a manuscript under review for publication: Goyal et al.

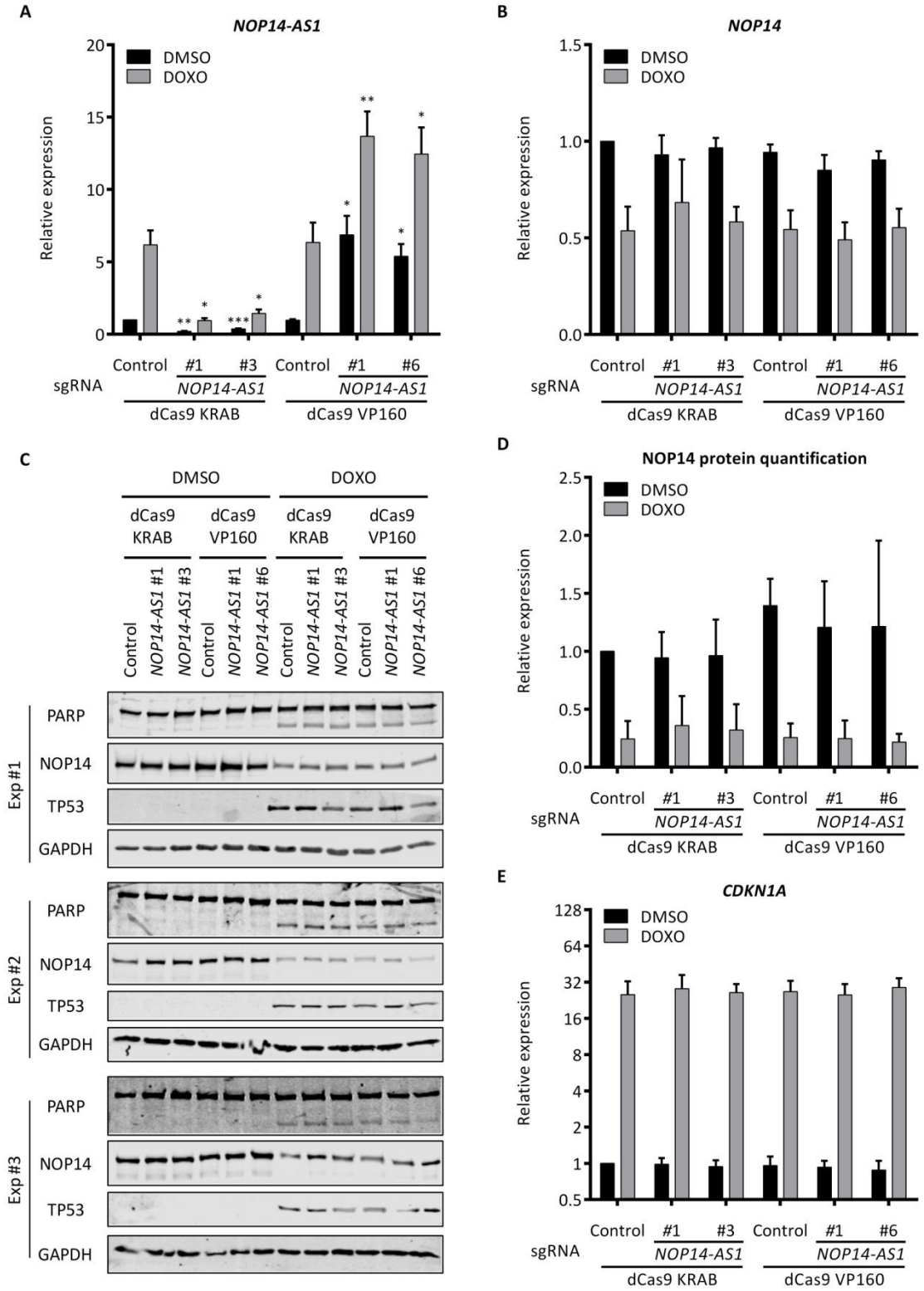


Figure 30: *NOP14-AS1* transcription modulation in A549 cells has no impact on *NOP14* expression
A549 cells were transduced with lentidCas9-KRAB-PURO iv sgRNA or lentidCas9-VP160-PURO iv sgRNA, respectively, containing either a control sgRNA or one of the two indicated sgRNAs targeting the *NOP14-AS1* promoter. These were

then treated with either 1 µM Doxorubicin (DOXO) or vehicle control DMSO for 24 hours. (A) RT-qPCR results for *NOP14* normalized to *Cyclophilin A* and control sgRNA + DMSO controls. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test. (B) Western blot results for PARP, NOP14, and p53 and p21. GAPDH was used as a loading control. (C) Quantification of the NOP14 protein expression normalized to GAPDH and control sgRNA + DMSO control. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to control sgRNA, unpaired two-sided t-test.
Data also used for a manuscript under review for publication: Goyal et al.

4. Discussion

4.1. CRISPRi for lncRNA knockdown: Lessons from the *NOP14-AS1* locus

Besides achieving a successful knockdown, *NOP14-AS1* targeting using dCas9/dCas9-KRAB (**Figure 7**) revealed two things about the CRISPRi system: (1) dCas9-KRAB was more efficient as compared to dCas9 and (2) dCas9/dCas9-KRAB mediated knockdown of a lncRNA can affect transcription of a neighboring gene. It was not surprising that dCas9-KRAB was a more potent inhibitor of *NOP14-AS1*, as this phenomenon was already observed by others while using this system for knockdown of protein-coding genes[96]. dCas9 is suggested to inhibit transcription initiation or elongation by sterically hindering the RNA polymerase[95]. Fusing an additional transcriptional repressor domain of KRAB to dCas9 was expected to make it more effective at transcriptional inhibition. Nevertheless, the simultaneous knockdown of *MFSD10*, while targeting *NOP14-AS1* was unexpected. Two possible explanations could be invoked: either this was an artifact of CRISPRi or a consequence of an *in cis*-regulatory relationship between *NOP14-AS1* and *MFSD10*. An ASO-mediated knockdown of *NOP14-AS1*, did not affect the *MFSD10* expression ruling out any transcript-dependent *in cis* effect of *NOP14-AS1* on *MFSD10* expression. However, it was still possible that the reduced *NOP14-AS1* transcription could lead to reduced *MFSD10* expression.

To further demonstrate that this phenomenon was indeed an artifact of the technique and not a locus-specific effect, this system was tested for five additional loci: *TP53*/*WRAP53*, *HOXD1*/*HOXD-AS1*, *RB1*/*LINC00441*, *MNX1*/*MNX1-AS1* and *HOXC11*/*HOTAIR* (**Figure 8, Figure 9, Figure 10, Figure 11, Figure 12, Figure 13**). In all these cases, a dCas9-KRAB mediated knockdown of the intended target resulted in simultaneous knockdown of the neighboring gene, while RNAi- or ASO-based approaches did not show such effects. Furthermore, the results from *HOTAIR* knockdown suggested that dCas9-KRAB could only target expression from the promoters which are in proximity to that of the intended target (**Figure 13)**. A recent study on the genome-scale repression of protein-coding genes using CRISPRi concluded that dCas9-KRAB repressed promoters when recruited within a span of up to 2 kb. These results are in agreement with the observations from *HOXC11*/*HOTAIR* locus.

Given that lncRNA loci are often found overlapping or in close proximity to other genes, it is possible that the use of CRISPRi to specifically knockdown a lncRNA is not possible for a vast proportion of lncRNAs. Such lncRNAs could be called “non-CRISPRable”. For example, all lncRNAs arising from bidirectional promoters are non-CRISPRable (**Figure 31A**). Similarly, all lncRNAs whose promoters lie in a close proximity to the promoters of other genes are also non-CRISPRable (**Figure 31B**). Since dCas9/dCas9-KRAB can also affect transcription elongation by sterically blocking the movement of RNA Polymerase, internal lncRNAs are also non-CRISPRable (**Figure 31C**). Only those lncRNAs, which do not fall in any of the above-mentioned categories, could be targeted by CRISPR-based approaches without directly targeting other genes in the same locus (**Figure 31D**).

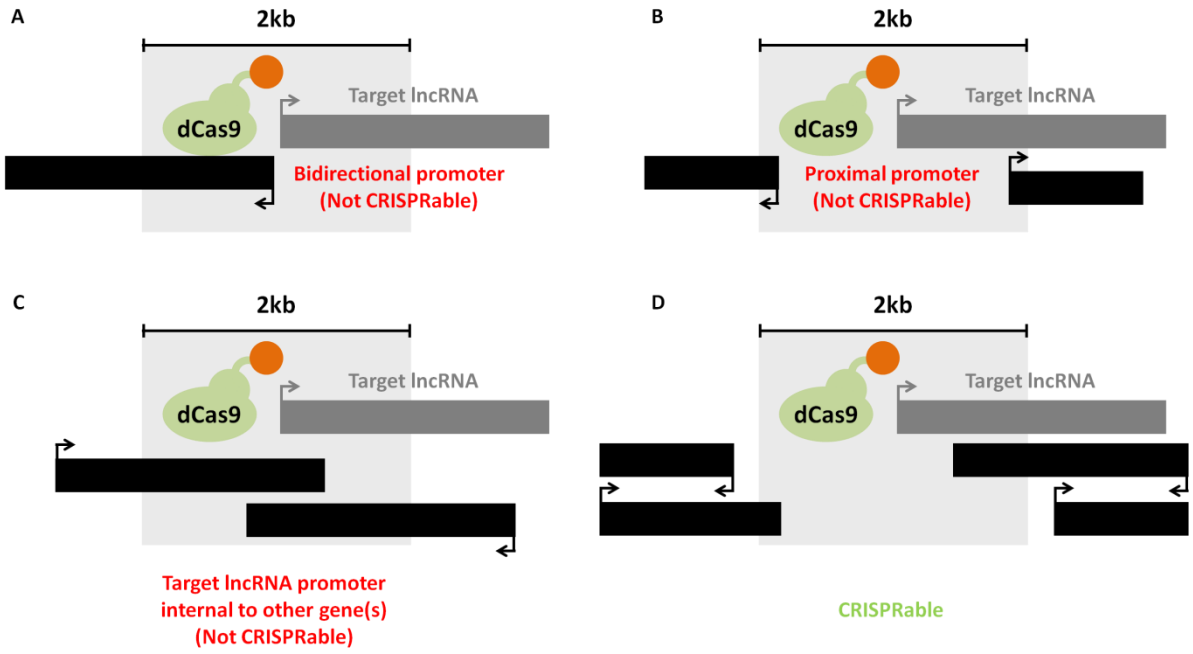


Figure 31: lncRNA CRISPRability based on its genomic context
Schematic representation of dCas9-KRAB-mediated knockdown of a lncRNA arising from (A) a bi-directional promoter, (B) a promoter proximal to the promoter of another gene, (C) an internal promoter or (D) an independent promoter with no proximal promoter.

Similarly, a Cas9-mediated excision of full-length lncRNA loci, though straightforward for intergenic lncRNAs, is impossible for intragenic lncRNAs overlapping with another gene as this will also lead to deletion of the sequences of those genes. Also, deletion of lncRNA promoters can only be performed for those lncRNAs which arise from external and independent promoters. For lncRNAs arising from internal/bidirectional promoters, such a deletion could also affect their overlapping/neighboring genes. In some cases, it might be possible to generate a functional knockout by deleting a part of the lncRNA which is distant from its promoter and not overlapping with any other gene[145]. However, for many lncRNAs, their functional domain lies in the promoter proximal or the part overlapping with other gene[34, 43]. In such cases, deletion of the promoter distal elements would have no effect on the lncRNA function. Also, deleting a part of the lncRNA without removing its transcriptional start site (TSS) could potentially lead to the generation of a new gene body with potentially new functions.

A genome-wide bioinformatic analysis to determine the fraction of CRISPRable lncRNAs was performed in the lab by Ksenia Myacheva. She found that only 38% of the total annotated lncRNAs were CRISPRable whereas 62% lncRNAs were not safely amenable to CRISPR/Cas9-mediated approaches (20% due to the bidirectional promoter, 7% due to proximal promoters, and 35% due to the internal promoter)[116]. A recent study utilizing dCas9-KRAB to identify lncRNA genes essential for cell proliferation found a total of 439 hits. However, out of these 439 lncRNAs, 169 were located within a 1 kb distance from the promoter a protein-coding which was also essential for cell proliferation in a different CRISPRi screen for essential protein-coding genes[104, 106]. This further validates the argument that CRISPRi should be used carefully for lncRNAs. In summary, the CRISPRi system provides an important and essential tool that also comparatively easily allows genome-wide screens.

However, the limitations of this system need to be taken into account in order to obtain relevant and biologically significant results for the genes targeted. The complexity of the targeted gene locus needs to be evaluated for sgRNA design and data analysis. Also, to avoid false positives, validation using orthogonal techniques like RNAi-based methods using siRNAs, siPOOLs or shRNAs or RNase H-mediated RNA degradation using ASOs or GapmeRs should complement the CRISPR-based experiments.

4.2. *NOP14-AS1*: A DNA damage induced lncRNA

LncRNAs play an important role in diverse cellular processes and the DDR is one of them. With an aim to identify novel lncRNAs involved in DDR, the Diederichs lab performed a microarray expression analysis of lncRNAs in cells treated with DNA damaging agents. This analysis identified lncRNA *NOP14-AS1* to be induced upon DNA damage. One of the aims of this Ph.D. project was to further characterize the regulation of *NOP14-AS1* in DNA damage.

In order to characterize the *NOP14-AS1* regulation in DNA damage, it was important to characterize its full-length sequence first. RACE analysis performed on cDNA from HepG2 cells identified that this gene gave rise to several splice isoforms arising from a total of three distinct TSSs (**Figure 15**). Multiple splice isoforms and a degenerate 5' end are a common feature of lncRNAs arising from such complex loci[146, 147]. Analysis of the publically available RNA-seq data for several cell lines corroborated the observations from RACE analysis (**Figure 15**). Analysis of the Encode CAGE data revealed that the first TSS accounted for most of the expression from the *NOP14-AS1* locus even though the number of clones obtained from 5'RACE was equally distributed for all the three TSSs (**Figure 15**). One possible explanation for this observation could be that since this analysis was performed on cDNA obtained from HepG2 cells treated with Etoposide, it could be that the second and third TSSs are transcriptionally more active upon DNA damage induction while the first TSS is constitutively active. Another possibility is that RACE analysis is not quantitative and is often biased towards detection of shorter isoforms. Nevertheless, this data proved to be of immense use in designing not only the RT-qPCR amplicons used to detect *NOP14-AS1* expression but also the LNA antisense GapmeRs as well as the sgRNAs used to knockdown/overexpress *NOP14-AS1* later during the course of this dissertation project. This analysis also advocates for exploitation of publically available RNA-seq and CAGE data for lncRNA research in combination with RACE and RT-qPCR experiments.

Treatments in four cell lines using a panel of five DNA damaging drugs established that the *NOP14-AS1* induction was not a cell type- or drug-specific effect (**Figure 16**). NCI-H460 cells showed the highest induction of *NOP14-AS1* and were therefore used as the preferred cell line model for the rest of the study. DNA damage-dependent *NOP14-AS1* induction was a transcriptional response (**Figure 17**) as was the case for most of the DNA damage induced lncRNAs. *NOP14-AS1* was induced in A549, NCI-H460 as well as HepG2 cells upon treatment with Nutlin-3 (**Figure 18A**). Also, Etoposide and Doxorubicin treatment in wt HCT-116 cells resulted in *NOP14-AS1* induction, while *TP53*-deficient HCT-116 cells showed no such induction (**Figure 18C**). Moreover, *TP53* knockdown using dCas9-KRAB or an siPOOL in NCI-H460 and A549 cells also resulted in a loss of *NOP14-*

AS1 induction upon Doxorubicin treatment (**Figure 22**). These results indicated that p53 was necessary and sufficient for *NOP14-AS1* upon DNA damage and add *NOP14-AS1* to the growing list of lncRNAs that are p53 regulated, such as *lincRNA-p21*, *PANDAR*, *DINO* etc.[51]. However, it is still not determined whether *NOP14-AS1* is a direct target of p53 as no canonical p53-binding sites could be detected in the *NOP14-AS1* promoter using the p53scan tool[148]. A p53 chip-seq could be performed to determine whether p53 binds to the *NOP14-AS1* promoter. The p53 response element thus determined could then be mutated (using Cas9 for example) and this could also serve as a loss-of-function model for *NOP14-AS1* in DNA damage. Most of the lncRNAs induced by p53 were also shown to play a role in modulating the p53 pathway[51]. This could argue for a role of *NOP14-AS1* in the p53 pathway.

4.3. *NOP14-AS1* and *NOP14* inverse co-regulation

Time course expression analysis in NCI-H460, A549, and HepG2 cells treated with one of the three DNA damaging drugs (Doxorubicin, Etoposide, and Cisplatin) revealed a striking inverse co-regulation of *NOP14* (mRNA as well as protein) and *NOP14-AS1* (**Figure 19, Figure 20, Figure 21**). Moreover, the observed *NOP14-AS1* and *NOP14* regulation were p53-dependent (**Figure 22**). Such DNA damage induced inverse co-regulation of a sense-antisense pair was previously observed for *INK4b* and the lncRNA *ANRIL*[42]. While *INK4b* expression was reduced, *ANRIL* expression was elevated upon DNA damage (caused by Etoposide, Bleomycin, and Neocarzinostatin). The same study concluded that *ANRIL* was induced via the ATM-E2F pathway and was responsible for *INK4b* repression observed upon DNA damage, establishing that a stimulus-induced inverse co-regulation of a sense-antisense pair of mRNA and lncRNA could indicate a functional interplay between them. Another example of a sense-antisense pair which shows a stimulus-dependent co-regulation is that of the *IL-1*α mRNA and its antisense lncRNA *AS-IL1*α. *AS-IL1*α is induced in bone marrow-derived macrophages upon Lipopolysaccharide (LPS) stimulation. Concurrent to the *AS-IL1*α induction, *IL-1*α transcription is also enhanced in an *AS-IL1*α-dependent manner[149].

Antisense lncRNAs often regulate the expression of their overlapping sense protein-coding genes through diverse mechanisms. At the level of transcription, they can induce promoter DNA methylation[22, 24, 141], recruit histone modifying enzymes[21, 44, 142, 144, 150-152], directly interfere and block the transcriptional machinery via transcriptional interference[28, 140, 153, 154], or regulate sense mRNA splicing[143, 155, 156]. Post-transcriptionally, they can bind to their sense mRNA and increase its stability by masking miRNA-binding sites[32] or enhance its translation by recruiting additional factors[34]. Although they can also act *in trans* to regulate the expression of other genes, antisense lncRNAs are suggested to act more frequently *in cis* due to their proximity to their overlapping sense genes[19, 157]. The inverse co-regulation of the sense-antisense pair of *NOP14* and *NOP14-AS1* and their co-dependence on p53 was a strong indicator of a possible functional relationship between them and therefore demanded further interrogation.

4.4. *NOP14* and its role in *NOP14-AS1* regulation

NOP14 knockdown in NCI-H460 as well as A549 cells using siRNAs, had no effect on *NOP14-AS1* expression (**Figure 23, Figure 24**). One possible explanation could be that either NOP14 does not affect *NOP14-AS1* expression or *NOP14* mRNA regulates the *NOP14-AS1* expression *in cis* similar to several other mRNAs[158], and such a regulatory relationship could not be uncovered by the RNAi-mediated knockdown. In line with this argument, a dCas9-KRAB-mediated repression of *NOP14* transcription resulted in a moderate *NOP14-AS1* induction (**Figure 25**). This induction (4-fold) was however not comparable to the much stronger induction (22-fold) observed upon Doxorubicin treatment, despite similar expression levels of *NOP14* (**Figure 25**). Moreover, there was no further increase in *NOP14-AS1* induction in the *NOP14* knockdown cells upon Doxorubicin treatment, despite the increased *NOP14* repression. These data indicate that the *NOP14-AS1* induction upon Doxorubicin treatment was due to an independent mechanism. *NOP14-AS1* induction observed upon *NOP14* knockdown using dCas9-KRAB was likely due to reduced transcriptional interference resulting from a reduction in *NOP14* transcription. In line with this argument, a plasmid-based rescue of *NOP14* expression in the *NOP14* knockdown cells did not result in a loss of *NOP14-AS1* induction (**Figure 27**). These results are of particular significance, as they demonstrate that CRISPRi can be used to study transcriptional interference between overlapping genes.

Interestingly, *NOP14* knockdown using siRNAs, as well as dCas9-KRAB, resulted in the induction of *CDKN1A* (**Figure 23, Figure 24, Figure 25**). This could be explained by the p53 stabilization upon *NOP14* knockdown, which could also be reversed upon ectopic rescue of *NOP14* expression (**Figure 25**, **Figure 27**). *NOP14* is a nucleolar protein where it is involved in the biogenesis of the small ribosomal subunit[159, 160]. It is a part of complex which is required for nucleolar localization of the EMG1, an RNA methyltransferase that is required for 18S rRNA processing[160-162]. The detected downregulation of *NOP14* could correspond to the nucleolar stress-mediated p53 stabilization[163] or DNA damage-induced inhibition of rRNA synthesis[164, 165].

4.5. *NOP14-AS1* and its role in *NOP14* regulation

An LNA antisense GapmeR-based knockdown of *NOP14-AS1* did not produce any significant impact on *NOP14* expression in the presence or absence of Doxorubicin (**Figure 28**) indicating that *NOP14* repression upon DNA damage was not *NOP14-AS1* lncRNA transcript driven. In the previous studies on *ANRIL* and *AS-IL1α*, an RNAi-based knockdown was sufficient to reverse the effect of these lncRNAs on their respective overlapping protein-coding genes[42, 149]. An RNAi-based knockdown of *NOP14-AS1* was also attempted using several siRNAs and also an siPOOL, but these reagents did not give a satisfactory knockdown. However, the ASO-based knockdown resulted in decreased *NOP14-AS1* levels and this should ideally be sufficient to at least partially reverse the repression of *NOP14* if mediated by the *NOP14-AS1* transcript. Again, two possible explanations for this observation could be invoked: either *NOP14* expression was independent of *NOP14-AS1* or the DNA damage-induced *NOP14-AS1* transcription and not the transcript itself was responsible for *NOP14* repression. To dissect these possibilities, a CRISPRi-mediated transcription repression and a CRISPRa-mediated

transcriptional activation of *NOP14-AS1* also did not have any impact on *NOP14* expression (**Figure 29, Figure 30**) making it clear that *NOP14-AS1* had no role in the regulation of *NOP14* in DNA damage. Moreover, unlike *NOP14* knockdown, *NOP14-AS1* knockdown cells did not show any obvious growth defects. Also, no effects on p53, PARP or *CDKN1A* expression were observed.

The CRISPRi-based loss-of-function model should be able to uncover any possible regulatory relationship between *NOP14-AS1* and *NOP14*. However, to rule out that any low-level expression from the *NOP14-AS1* could still impact *NOP14* expression, Berta Duran Arqué generated a *NOP14-AS1* deletion clone in Hek293 cells using Cas9. Hek293 cells were co-transfected with Cas9 and two sgRNAs spanning the non-overlapping part of the *NOP14-AS1* gene (**Figure 32A**). The clonal selection followed by genotyping identified one homozygous (clone #51) and four heterozygous deletion clones (**Figure 32A**).

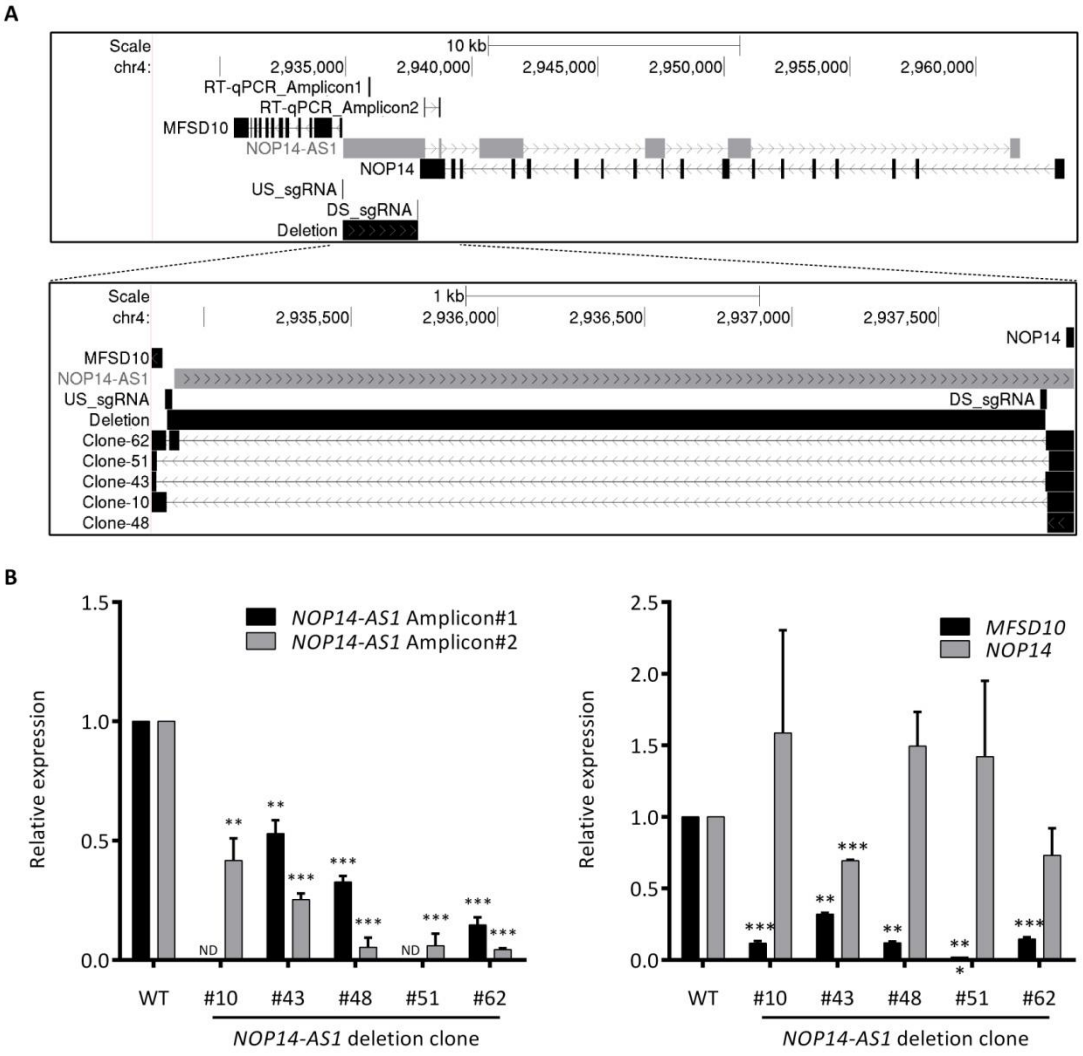


Figure 32: *NOP14-AS1* deletion in Hek293 using CRISPR/Cas9

A: Upper panel: Schematic representation of the *NOP14-AS1* genomic locus depicting the sgRNAs used to delete *NOP14-AS1* as well as the RT-qPCR amplicons used to detect *NOP14-AS1* expression. Lower panel: Hek293 cells were transfected with PX458-2X sgRNA plasmid, which expressed Cas9 as well as the two indicated sgRNAs. 48 hours post-transfection, cells were single cell sorted and allowed to grow for 2 weeks. The clones thus obtained were screened for NOP14-AS1 deletion. Sanger sequencing results of the five homozygous deletion clones obtained.

B,C: RT-qPCR results for (B) *NOP14-AS1* as measured using two different amplicons and (C) *MFSD10* and *NOP14* normalized to *Cyclophilin* A and Hek293 wild type (wt) cells. Error bars represent SD (n=3). * p<0.05; ** p<0.01; *** p<0.001 compared to wt cells, unpaired two-sided t-test.
Data used here has been generated by Berta Duran Arque under the supervision of Ashish Goyal and published as a part of the following manuscript: Goyal et al., Nucleic Acids Res, 2016[116].

Expression analysis revealed that *NOP14-AS1* expression was significantly reduced in all deletion clones. Clone #51 had no detectable *NOP14-AS1* expression as measured using RT-qPCR Amplicon#1. However, there was some residual *NOP14-AS1* expression as measured using Amplicon#2 (**Figure 32B**). This could be because Amplicon#2 lies outside the deleted region. Despite significant knockdown of *NOP14-AS1* in all five clones, *NOP14* expression was not significantly changed (**Figure 32B,C**). *MFSD10* expression was affected in these clones possibly due to deletion of promoter elements (**Figure 32C**) again advocating for caution while using Cas9 to knockout lncRNA genes.

Taken together, these experiments did not detect any role of *NOP14-AS1* in the regulation of *NOP14*. It is quite possible that *NOP14-AS1* plays a role in DNA damage response further downstream which is yet to be discovered. Given that its regulation upon DNA damage is controlled by p53, it could play a role in the p53 pathway as *lincRNA-p21, PANDAR*, *DINO*, *lincRNA-ROR* or *PINT*. However, no changes in p53, as well as CDKN1A expression, were observed upon *NOP14-AS1* knockdown or overexpression. This might indicate a further downstream role of this lncRNA similar to *lincRNA-p21* or *PANDAR*, which do not regulate the p53 levels but instead regulate the expression of selected p53 target genes. Since the loss-of-function and gain-of-function models for *NOP14-AS1* have been established, these could be used to perform gene expression analysis using microarray or RNA-seq to identify genes that are affected by *NOP14-AS1* regulation. The gene sets identified could be scanned for any overlaps with p53 target genes. There is also a possibility to find protein interaction partners of *NOP14-AS1* using techniques like RAP-MS (RNA antisense purification- mass spectrometry)[166] or ChiRP-MS (comprehensive identification of RNA-binding proteins by mass spectrometry)[167, 168].

4.6. Conclusions

Previously, *NOP14-AS1* was identified as a DNA damage-regulated lncRNA and that it may be inversely co-regulated with its neighboring protein-coding gene *NOP14*. In this study, I analyze the regulatory mechanisms of this induction and whether this sense-antisense pair of *NOP14* and *NOP14-AS1* shares a *cis*-regulatory relationship. I demonstrate that *NOP14-AS1* is a p53-inducible lncRNA which undergoes transcriptional induction upon DNA damage. I confirm the significant inverse co-regulation of *NOP14-AS1* and *NOP14* in multiple cell lines using multiple DNA-damaging drugs and demonstrate that both co-dependent on p53 for regulation upon DNA damage. Using the state-of-the-art CRISPR/Cas9 system, I was able to demonstrate that neither *NOP14* nor *NOP14-AS1* regulate the expression of each other. In the process of establishing this system, I experimentally demonstrate a major limitation of the CRISPRi system for lncRNA knockdown: CRISPRi targeting of a lncRNA arising from a complex locus can unintentionally affect the expression of its neighboring

genes. These results demonstrate that not all lncRNA-mRNA sense-antisense pairs necessarily regulate each other even if their expression is correlated and affected by the same pathways, and call for caution when functional interpretations are drawn based on the correlation between sense-antisense pairs in genome-wide expression studies.

Bibliography

1. Crick, F.H., *On protein synthesis.* Symp Soc Exp Biol, 1958. **12**: p. 138-63.
2. RajBhandary, U.L. and C. Kohrer, *Early days of tRNA research: discovery, function, purification and sequence analysis.* J Biosci, 2006. **31**(4): p. 439-51.
3. Palade, G.E., *A small particulate component of the cytoplasm.* J Biophys Biochem Cytol, 1955. **1**(1): p. 59-68.
4. Weinberg, R.A. and S. Penman, *Small molecular weight monodisperse nuclear RNA.* J Mol Biol, 1968. **38**(3): p. 289-304.
5. Dreyfuss, G., L. Philipson, and I.W. Mattaj, *Ribonucleoprotein particles in cellular processes.* J Cell Biol, 1988. **106**(5): p. 1419-25.
6. Maxwell, E.S. and M.J. Fournier, *The small nucleolar RNAs.* Annu Rev Biochem, 1995. **64**: p. 897-934.
7. Kruger, K., P.J. Grabowski, A.J. Zaug, J. Sands, D.E. Gottschling, and T.R. Cech, *Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena.* Cell, 1982. **31**(1): p. 147-57.
8. Guerrier-Takada, C., K. Gardiner, T. Marsh, N. Pace, and S. Altman, *The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme.* Cell, 1983. **35**(3 Pt 2): p. 849-57.
9. Brockdorff, N., A. Ashworth, G.F. Kay, V.M. McCabe, D.P. Norris, P.J. Cooper, S. Swift, and S. Rastan, *The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus.* Cell, 1992. **71**(3): p. 515-26.
10. Sarah Djebali, C.A.D., Angelika Merkel, Alex Dobin, Timo Lassmann, Ali Mortazavi, Andrea Tanzer, Julien Lagarde, Wei Lin, Felix Schlesinger, Chenghai Xue, Georgi K. Marinov, Jainab Khatun, Brian A. Williams, Chris Zaleski, Joel Rozowsky, Maik Röder, Felix Kokocinski, Rehab F. Abdelhamid, Tyler Alioto, Igor Antoshechkin, Michael T. Baer, Nadav S. Bar, Philippe Batut, Kimberly Bell, Ian Bell, Sudipto Chakrabortty, Xian Chen, Jacqueline Chrast, Joao Curado, Thomas Derrien, Jorg Drenkow, Erica Dumais, Jacqueline Dumais, Radha Duttagupta, Emilie Falconnet, Meagan Fastuca, Kata Fejes-Toth, Pedro Ferreira, Sylvain Foissac, Melissa J. Fullwood, Hui Gao, David Gonzalez, Assaf Gordon, Harsha Gunawardena, Cedric Howald, Sonali Jha, Rory Johnson, Philipp Kapranov, Brandon King, Colin Kingswood, Oscar J. Luo, Eddie Park, Kimberly Persaud, Jonathan B. Preall, Paolo Ribeca, Brian Risk, Daniel Robyr, Michael Sammeth, Lorian Schaffer, Lei-Hoon See, Atif Shahab, Jorgen Skancke, Ana Maria Suzuki, Hazuki Takahashi, Hagen Tilgner, Diane Trout, Nathalie Walters, Huaien Wang, John Wrobel, Yanbao Yu, Xiaoan Ruan, Yoshihide Hayashizaki, Jennifer Harrow, Mark Gerstein, Tim Hubbard, Alexandre Reymond, Stylianos E. Antonarakis, Gregory Hannon, Morgan C. Giddings, Yijun Ruan, Barbara Wold, Piero Carninci, Roderic Guigó & Thomas R. Gingeras, *Landscape of transcription in human cells.* Nature, 2012. **489**(7414): p. 101-8.
11. Ma, L., V.B. Bajic, and Z. Zhang, *On the classification of long non-coding RNAs.* RNA Biol, 2013. **10**(6): p. 925-33.

12. St Laurent, G., C. Wahlestedt, and P. Kapranov, *The Landscape of long noncoding RNA classification.* Trends Genet, 2015. **31**(5): p. 239-51.
13. Derrien, T., R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D.G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J.B. Brown, L. Lipovich, J.M. Gonzalez, M. Thomas, C.A. Davis, R. Shiekhattar, T.R. Gingeras, T.J. Hubbard, C. Notredame, J. Harrow, and R. Guigo, *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.
14. Fatica, A. and I. Bozzoni, *Long non-coding RNAs: new players in cell differentiation and development.* Nat Rev Genet, 2014. **15**(1): p. 7-21.
15. Huarte, M., *The emerging role of lncRNAs in cancer.* Nat Med, 2015. **21**(11): p. 1253-61.
16. Kitagawa, M., K. Kitagawa, Y. Kotake, H. Niida, and T. Ohhata, *Cell cycle regulation by long non-coding RNAs.* Cell Mol Life Sci, 2013. **70**(24): p. 4785-94.
17. Dhamija, S. and S. Diederichs, *From junk to master regulators of invasion: lncRNA functions in migration, EMT and metastasis.* Int J Cancer, 2016. **139**(2): p. 269-80.
18. Bonasio, R. and R. Shiekhattar, *Regulation of transcription by long noncoding RNAs.* Annu Rev Genet, 2014. **48**: p. 433-55.
19. Pelechano, V. and L.M. Steinmetz, *Gene regulation by antisense transcription.* Nat Rev Genet, 2013. **14**(12): p. 880-93.
20. Kornienko, A.E., P.M. Guenzl, D.P. Barlow, and F.M. Pauler, *Gene regulation by the act of long non-coding RNA transcription.* BMC Biol, 2013. **11**: p. 59.
21. Yap, K.L., S. Li, A.M. Munoz-Cabello, S. Raguz, L. Zeng, S. Mujtaba, J. Gil, M.J. Walsh, and M.M. Zhou, *Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a.* Mol Cell, 2010. **38**(5): p. 662-74.
22. Yu, W., D. Gius, P. Onyango, K. Muldoon-Jacobs, J. Karp, A.P. Feinberg, and H. Cui, *Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA.* Nature, 2008. **451**(7175): p. 202-6.
23. Pasmant, E., A. Sabbagh, M. Vidaud, and I. Bieche, *ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS.* FASEB J, 2011. **25**(2): p. 444-8.
24. Tufarelli, C., J.A. Stanley, D. Garrick, J.A. Sharpe, H. Ayyub, W.G. Wood, and D.R. Higgs, *Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease.* Nat Genet, 2003. **34**(2): p. 157-65.
25. Ji, P., S. Diederichs, W. Wang, S. Boing, R. Metzger, P.M. Schneider, N. Tidow, B. Brandt, H. Buerger, E. Bulk, M. Thomas, W.E. Berdel, H. Serve, and C. Muller-Tidow, *MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer.* Oncogene, 2003. **22**(39): p. 8031-41.
26. Gutschner, T., M. Hammerle, M. Eissmann, J. Hsu, Y. Kim, G. Hung, A. Revenko, G. Arun, M. Stentrup, M. Gross, M. Zornig, A.R. MacLeod, D.L. Spector, and S. Diederichs, *The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells.* Cancer Res, 2013. **73**(3): p. 1180-9.

27. Yang, L., C. Lin, W. Liu, J. Zhang, K.A. Ohgi, J.D. Grinstein, P.C. Dorrestein, and M.G. Rosenfeld, *ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs.* Cell, 2011. **147**(4): p. 773-88.

28. Hongay, C.F., P.L. Grisafi, T. Galitski, and G.R. Fink, *Antisense transcription controls cell fate in Saccharomyces cerevisiae.* Cell, 2006. **127**(4): p. 735-45.

29. Gelfand, B., J. Mead, A. Bruning, N. Apostolopoulos, V. Tadigotla, V. Nagaraj, A.M. Sengupta, and A.K. Vershon, *Regulated antisense transcription controls expression of cell-type-specific genes in yeast.* Mol Cell Biol, 2011. **31**(8): p. 1701-9.

30. Rinn, J.L., M. Kertesz, J.K. Wang, S.L. Squazzo, X. Xu, S.A. Brugmann, L.H. Goodnough, J.A. Helms, P.J. Farnham, E. Segal, and H.Y. Chang, *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.* Cell, 2007. **129**(7): p. 1311-23.

31. Geisler, S. and J. Coller, *RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts.* Nat Rev Mol Cell Biol, 2013. **14**(11): p. 699-712.

32. Faghihi, M.A., F. Modarresi, A.M. Khalil, D.E. Wood, B.G. Sahagan, T.E. Morgan, C.E. Finch, G. St Laurent, 3rd, P.J. Kenny, and C. Wahlestedt, *Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase.* Nat Med, 2008. **14**(7): p. 723-30.

33. Faghihi, M.A., M. Zhang, J. Huang, F. Modarresi, M.P. Van der Brug, M.A. Nalls, M.R. Cookson, G. St-Laurent, 3rd, and C. Wahlestedt, *Evidence for natural antisense transcript-mediated inhibition of microRNA function.* Genome Biol, 2010. **11**(5): p. R56.

34. Carrieri, C., L. Cimatti, M. Biagioli, A. Beugnet, S. Zucchelli, S. Fedele, E. Pesce, I. Ferrer, L. Collavin, C. Santoro, A.R. Forrest, P. Carninci, S. Biffo, E. Stupka, and S. Gustincich, *Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat.* Nature, 2012. **491**(7424): p. 454-7.

35. Thomson, D.W. and M.E. Dinger, *Endogenous microRNA sponges: evidence and controversy.* Nat Rev Genet, 2016. **17**(5): p. 272-83.

36. Cesana, M., D. Cacchiarelli, I. Legnini, T. Santini, O. Sthandier, M. Chinappi, A. Tramontano, and I. Bozzoni, *A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA.* Cell, 2011. **147**(2): p. 358-69.

37. Feng, J., W.D. Funk, S.S. Wang, S.L. Weinrich, A.A. Avilion, C.P. Chiu, R.R. Adams, E. Chang, R.C. Allsopp, J. Yu, and et al., *The RNA component of human telomerase.* Science, 1995. **269**(5228): p. 1236-41.

38. Sancar, A., L.A. Lindsey-Boltz, K. Unsal-Kacmaz, and S. Linn, *Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints.* Annu Rev Biochem, 2004. **73**: p. 39-85.

39. Liu, Y. and X. Lu, *Non-coding RNAs in DNA damage response.* Am J Cancer Res, 2012. **2**(6): p. 658-75.

40. Wang, X., S. Arai, X. Song, D. Reichart, K. Du, G. Pascual, P. Tempst, M.G. Rosenfeld, C.K. Glass, and R. Kurokawa, *Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription.* Nature, 2008. **454**(7200): p. 126-30.

41. Kurokawa, R., *Promoter-associated long noncoding RNAs repress transcription through a RNA binding protein TLS.* Adv Exp Med Biol, 2011. **722**: p. 196-208.

42. Wan, G., R. Mathur, X. Hu, Y. Liu, X. Zhang, G. Peng, and X. Lu, *Long non-coding RNA ANRIL (CDKN2B-AS) is induced by the ATM-E2F1 signaling pathway.* Cell Signal, 2013. **25**(5): p. 1086-95.

43. Mahmoudi, S., S. Henriksson, M. Corcoran, C. Mendez-Vidal, K.G. Wiman, and M. Farnebo, *Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage.* Mol Cell, 2009. **33**(4): p. 462-71.

44. O'Leary, V.B., S.V. Ovsepian, L.G. Carrascosa, F.A. Buske, V. Radulovic, M. Niyazi, S. Moertl, M. Trau, M.J. Atkinson, and N. Anastasov, *PARTICLE, a Triplex-Forming Long ncRNA, Regulates Locus-Specific Methylation in Response to Low-Dose Irradiation.* Cell Rep, 2015. **11**(3): p. 474-85.

45. Sharma, V., S. Khurana, N. Kubben, K. Abdelmohsen, P. Oberdoerffer, M. Gorospe, and T. Misteli, *A BRCA1-interacting lncRNA regulates homologous recombination.* EMBO Rep, 2015. **16**(11): p. 1520-34.

46. Zhang, Y., Q. He, Z. Hu, Y. Feng, L. Fan, Z. Tang, J. Yuan, W. Shan, C. Li, X. Hu, J.L. Tanyi, Y. Fan, Q. Huang, K. Montone, C.V. Dang, and L. Zhang, *Long noncoding RNA LINP1 regulates repair of DNA double-strand breaks in triple-negative breast cancer.* Nat Struct Mol Biol, 2016. **23**(6): p. 522-30.

47. Lee, S., F. Kopp, T.C. Chang, A. Sataluri, B. Chen, S. Sivakumar, H. Yu, Y. Xie, and J.T. Mendell, *Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins.* Cell, 2016. **164**(1-2): p. 69-80.

48. Feldstein, O., T. Nizri, T. Doniger, J. Jacob, G. Rechavi, and D. Ginsberg, *The long non-coding RNA ERIC is regulated by E2F and modulates the cellular response to DNA damage.* Mol Cancer, 2013. **12**(1): p. 131.

49. Beckerman, R. and C. Prives, *Transcriptional regulation by p53.* Cold Spring Harb Perspect Biol, 2010. **2**(8): p. a000935.

50. Riley, T., E. Sontag, P. Chen, and A. Levine, *Transcriptional control of human p53-regulated genes.* Nat Rev Mol Cell Biol, 2008. **9**(5): p. 402-12.

51. Chaudhary, R. and A. Lal, *Long noncoding RNAs in the p53 network.* Wiley Interdiscip Rev RNA, 2016.

52. Subramanian, M., M.F. Jones, and A. Lal, *Long Non-Coding RNAs Embedded in the Rb and p53 Pathways.* Cancers (Basel), 2013. **5**(4): p. 1655-75.

53. Grossi, E., Y. Sanchez, and M. Huarte, *Expanding the p53 regulatory network: LncRNAs take up the challenge.* Biochim Biophys Acta, 2016. **1859**(1): p. 200-8.

54. Huarte, M., *p53 partners with RNA in the DNA damage response.* Nat Genet, 2016. **48**(11): p. 1298-1299.

55. Huarte, M., M. Guttman, D. Feldser, M. Garber, M.J. Koziol, D. Kenzelmann-Broz, A.M. Khalil, O. Zuk, I. Amit, M. Rabani, L.D. Attardi, A. Regev, E.S. Lander, T. Jacks, and J.L. Rinn, *A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response*. Cell, 2010. **142**(3): p. 409-19.

56. Hung, T., Y. Wang, M.F. Lin, A.K. Koegel, Y. Kotake, G.D. Grant, H.M. Horlings, N. Shah, C. Umbricht, P. Wang, B. Kong, A. Langerod, A.L. Borresen-Dale, S.K. Kim, M. van de Vijver, S. Sukumar, M.L. Whitfield, M. Kellis, Y. Xiong, D.J. Wong, and H.Y. Chang, *Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters.* Nat Genet, 2011. **43**(7): p. 621-9.

57. Schmitt, A.M., J.T. Garcia, T. Hung, R.A. Flynn, Y. Shen, K. Qu, A.Y. Payumo, A. Peres-da-Silva, D.K. Broz, R. Baum, S. Guo, J.K. Chen, L.D. Attardi, and H.Y. Chang, *An inducible long noncoding RNA amplifies DNA damage signaling.* Nat Genet, 2016. **48**(11): p. 1370-1376.

58. Zhang, A., N. Zhou, J. Huang, Q. Liu, K. Fukuda, D. Ma, Z. Lu, C. Bai, K. Watabe, and Y.Y. Mo, *The human long non-coding RNA-RoR is a p53 repressor in response to DNA damage.* Cell Res, 2013. **23**(3): p. 340-50.

59. Marin-Bejar, O., F.P. Marchese, A. Athie, Y. Sanchez, J. Gonzalez, V. Segura, L. Huang, I. Moreno, A. Navarro, M. Monzo, J. Garcia-Foncillas, J.L. Rinn, S. Guo, and M. Huarte, *Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2.* Genome Biol, 2013. **14**(9): p. R104.

60. Liu, X., D. Li, W. Zhang, M. Guo, and Q. Zhan, *Long non-coding RNA gadd7 interacts with TDP-43 and regulates Cdk6 mRNA decay.* EMBO J, 2012. **31**(23): p. 4415-27.

61. Wan, G., X. Hu, Y. Liu, C. Han, A.K. Sood, G.A. Calin, X. Zhang, and X. Lu, *A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation.* EMBO J, 2013. **32**(21): p. 2833-47.

62. Negishi, M., S.P. Wongpalee, S. Sarkar, J. Park, K.Y. Lee, Y. Shibata, B.J. Reon, R. Abounader, Y. Suzuki, S. Sugano, and A. Dutta, *A new lncRNA, APTR, associates with and represses the CDKN1A/p21 promoter by recruiting polycomb proteins.* PLoS One, 2014. **9**(4): p. e95216.

63. Prensner, J.R., W. Chen, M.K. Iyer, Q. Cao, T. Ma, S. Han, A. Sahu, R. Malik, K. Wilder-Romans, N. Navone, C.J. Logothetis, J.C. Araujo, L.L. Pisters, A.K. Tewari, C.E. Canman, K.E. Knudsen, N. Kitabayashi, M.A. Rubin, F. Demichelis, T.S. Lawrence, A.M. Chinnaiyan, and F.Y. Feng, *PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer.* Cancer Res, 2014. **74**(6): p. 1651-60.

64. Sanchez, Y., V. Segura, O. Marin-Bejar, A. Athie, F.P. Marchese, J. Gonzalez, L. Bujanda, S. Guo, A. Matheu, and M. Huarte, *Genome-wide analysis of the human p53 transcriptional network unveils a lncRNA tumour suppressor signature.* Nat Commun, 2014. **5**: p. 5812.

65. Gazy, I., D.A. Zeevi, P. Renbaum, S. Zeligson, L. Eini, D. Bashari, Y. Smith, A. Lahad, M. Goldberg, D. Ginsberg, and E. Levy-Lahad, *TODRA, a lncRNA at the RAD51 Locus, Is Oppositely Regulated to RAD51, and Enhances RAD51-Dependent DSB (Double Strand Break) Repair.* PLoS One, 2015. **10**(7): p. e0134120.

66. Adriaens, C., L. Standaert, J. Barra, M. Latil, A. Verfaillie, P. Kalev, B. Boeckx, P.W. Wijnhoven, E. Radaelli, W. Vermi, E. Leucci, G. Lapouge, B. Beck, J. van den Oord, S. Nakagawa, T. Hirose, A.A. Sablina, D. Lambrechts, S. Aerts, C. Blanpain, and J.C. Marine, *p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity.* Nat Med, 2016. **22**(8): p. 861-8.

67. Diaz-Lagares, A., A.B. Crujeiras, P. Lopez-Serra, M. Soler, F. Setien, A. Goyal, J. Sandoval, Y. Hashimoto, A. Martinez-Cardus, A. Gomez, H. Heyn, C. Moutinho, J. Espada, A. Vidal, M. Paules, M. Galan, N. Sala, Y. Akiyama, M. Martinez-Iniesta, L. Farre, A. Villanueva, M. Gross, S. Diederichs, S. Guil, and M.

Esteller, *Epigenetic inactivation of the p53-induced long noncoding RNA TP53 target 1 in human cancer.* Proc Natl Acad Sci U S A, 2016. **113**(47): p. E7535-E7544.

68. Mojica, F.J., C. Ferrer, G. Juez, and F. Rodriguez-Valera, *Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning.* Mol Microbiol, 1995. **17**(1): p. 85-93.
69. Mojica, F.J., C. Diez-Villasenor, E. Soria, and G. Juez, *Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria.* Mol Microbiol, 2000. **36**(1): p. 244-6.
70. Jansen, R., J.D. Embden, W. Gaastra, and L.M. Schouls, *Identification of genes that are associated with DNA repeats in prokaryotes.* Mol Microbiol, 2002. **43**(6): p. 1565-75.
71. Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D.A. Romero, and P. Horvath, *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.
72. Lander, E.S., *The Heroes of CRISPR.* Cell, 2016. **164**(1-2): p. 18-28.
73. Bhaya, D., M. Davison, and R. Barrangou, *CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation.* Annu Rev Genet, 2011. **45**: p. 273-97.
74. Makarova, K.S., Y.I. Wolf, O.S. Alkhnbashi, F. Costa, S.A. Shah, S.J. Saunders, R. Barrangou, S.J. Brouns, E. Charpentier, D.H. Haft, P. Horvath, S. Moineau, F.J. Mojica, R.M. Terns, M.P. Terns, M.F. White, A.F. Yakunin, R.A. Garrett, J. van der Oost, R. Backofen, and E.V. Koonin, *An updated evolutionary classification of CRISPR-Cas systems.* Nat Rev Microbiol, 2015. **13**(11): p. 722-36.
75. Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier, *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.
76. Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, X. Wu, W. Jiang, L.A. Marraffini, and F. Zhang, *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.
77. Wang, H., M. La Russa, and L.S. Qi, *CRISPR/Cas9 in Genome Editing and Beyond.* Annu Rev Biochem, 2016.
78. Lieber, M.R., *The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway.* Annu Rev Biochem, 2010. **79**: p. 181-211.
79. Jasin, M. and R. Rothstein, *Repair of strand breaks by homologous recombination.* Cold Spring Harb Perspect Biol, 2013. **5**(11): p. a012740.
80. Xiao, A., Z. Wang, Y. Hu, Y. Wu, Z. Luo, Z. Yang, Y. Zu, W. Li, P. Huang, X. Tong, Z. Zhu, S. Lin, and B. Zhang, *Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish.* Nucleic Acids Res, 2013. **41**(14): p. e141.
81. Essletzbichler, P., T. Konopka, F. Santoro, D. Chen, B.V. Gapp, R. Kralovics, T.R. Brummelkamp, S.M. Nijman, and T. Burckstummer, *Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line.* Genome Res, 2014. **24**(12): p. 2059-65.
82. Han, J., J. Zhang, L. Chen, B. Shen, J. Zhou, B. Hu, Y. Du, P.H. Tate, X. Huang, and W. Zhang, *Efficient in vivo deletion of a large imprinted lncRNA by CRISPR/Cas9.* RNA Biol, 2014. **11**(7): p. 829-835.

83. Aparicio-Prat, E., C. Arnan, I. Sala, N. Bosch, R. Guigo, and R. Johnson, *DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs.* BMC Genomics, 2015. **16**: p. 846.

84. Yin, Y., P. Yan, J. Lu, G. Song, Y. Zhu, Z. Li, Y. Zhao, B. Shen, X. Huang, H. Zhu, S.H. Orkin, and X. Shen, *Opposing Roles for the lncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell Differentiation.* Cell Stem Cell, 2015. **16**(5): p. 504-16.

85. Deng, C., Y. Li, L. Zhou, J. Cho, B. Patel, N. Terada, J. Bungert, Y. Qiu, and S. Huang, *HoxBlinc RNA Recruits Set1/MLL Complexes to Activate Hox Gene Expression Patterns and Mesoderm Lineage Development.* Cell Rep, 2016. **14**(1): p. 103-14.

86. Welsh, I.C., H. Kwak, F.L. Chen, M. Werner, L.S. Shopland, C.G. Danko, J.T. Lis, M. Zhang, J.F. Martin, and N.A. Kurpios, *Chromatin Architecture of the Pitx2 Locus Requires CTCF- and Pitx2-Dependent Asymmetry that Mirrors Embryonic Gut Laterality.* Cell Rep, 2015. **13**(2): p. 337-49.

87. Zhu, S., W. Li, J. Liu, C.H. Chen, Q. Liao, P. Xu, H. Xu, T. Xiao, Z. Cao, J. Peng, P. Yuan, M. Brown, X.S. Liu, and W. Wei, *Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library.* Nat Biotechnol, 2016. **34**(12): p. 1279-1286.

88. Durruthy-Durruthy, J., V. Sebastiano, M. Wossidlo, D. Cepeda, J. Cui, E.J. Grow, J. Davila, M. Mall, W.H. Wong, J. Wysocka, K.F. Au, and R.A. Reijo Pera, *The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming.* Nat Genet, 2016. **48**(1): p. 44-52.

89. Groff, A.F., D.B. Sanchez-Gomez, M.M. Soruco, C. Gerhardinger, A.R. Barutcu, E. Li, L. Elcavage, O. Plana, L.V. Sanchez, J.C. Lee, M. Sauvageau, and J.L. Rinn, *In Vivo Characterization of Linc-p21 Reveals Functional cis-Regulatory DNA Elements.* Cell Rep, 2016. **16**(8): p. 2178-86.

90. Paralkar, V.R., C.C. Taborda, P. Huang, Y. Yao, A.V. Kossenkov, R. Prasad, J. Luan, J.O. Davies, J.R. Hughes, R.C. Hardison, G.A. Blobel, and M.J. Weiss, *Unlinking an lncRNA from Its Associated cis Element.* Mol Cell, 2016.

91. Bassett, A.R., A. Akhtar, D.P. Barlow, A.P. Bird, N. Brockdorff, D. Duboule, A. Ephrussi, A.C. Ferguson-Smith, T.R. Gingeras, W. Haerty, D.R. Higgs, E.A. Miska, and C.P. Ponting, *Considerations when investigating lncRNA function in vivo.* eLife, 2014. **3**: p. e03058.

92. Ran, F.A., P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott, and F. Zhang, *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-308.

93. Yang, H., H. Wang, C.S. Shivalila, A.W. Cheng, L. Shi, and R. Jaenisch, *One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering.* Cell, 2013. **154**(6): p. 1370-9.

94. Gutschner, T., M. Baas, and S. Diederichs, *Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases.* Genome Res, 2011. **21**(11): p. 1944-54.

95. Qi, L.S., M.H. Larson, L.A. Gilbert, J.A. Doudna, J.S. Weissman, A.P. Arkin, and W.A. Lim, *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**(5): p. 1173-83.

96. Gilbert, L.A., M.H. Larson, L. Morsut, Z. Liu, G.A. Brar, S.E. Torres, N. Stern-Ginossar, O. Brandman, E.H. Whitehead, J.A. Doudna, W.A. Lim, J.S. Weissman, and L.S. Qi, *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* Cell, 2013. **154**(2): p. 442-51.

97. Margolin, J.F., J.R. Friedman, W.K. Meyer, H. Vissing, H.J. Thiesen, and F.J. Rauscher, 3rd, *Kruppel-associated boxes are potent transcriptional repression domains.* Proc Natl Acad Sci U S A, 1994. **91**(10): p. 4509-13.

98. Ayer, D.E., C.D. Laherty, Q.A. Lawrence, A.P. Armstrong, and R.N. Eisenman, *Mad proteins contain a dominant transcription repression domain.* Mol Cell Biol, 1996. **16**(10): p. 5772-81.

99. Cong, L., R. Zhou, Y.C. Kuo, M. Cunniff, and F. Zhang, *Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains.* Nat Commun, 2012. **3**: p. 968.

100. Larson, M.H., L.A. Gilbert, X. Wang, W.A. Lim, J.S. Weissman, and L.S. Qi, *CRISPR interference (CRISPRi) for sequence-specific control of gene expression.* Nat Protoc, 2013. **8**(11): p. 2180-96.

101. Maeder, M.L., S.J. Linder, V.M. Cascio, Y. Fu, Q.H. Ho, and J.K. Joung, *CRISPR RNA-guided activation of endogenous human genes.* Nat Methods, 2013. **10**(10): p. 977-9.

102. Perez-Pinera, P., D.D. Kocak, C.M. Vockley, A.F. Adler, A.M. Kabadi, L.R. Polstein, P.I. Thakore, K.A. Glass, D.G. Ousterout, K.W. Leong, F. Guilak, G.E. Crawford, T.E. Reddy, and C.A. Gersbach, *RNA-guided gene activation by CRISPR-Cas9-based transcription factors.* Nat Methods, 2013. **10**(10): p. 973-6.

103. Chavez, A., J. Scheiman, S. Vora, B.W. Pruitt, M. Tuttle, E. P R Iyer, S. Lin, S. Kiani, C.D. Guzman, D.J. Wiegand, D. Ter-Ovanesyan, J.L. Braff, N. Davidsohn, B.E. Housden, N. Perrimon, R. Weiss, J. Aach, J.J. Collins, and G.M. Church, *Highly efficient Cas9-mediated transcriptional programming.* Nat Meth, 2015. **12**(4): p. 326-328.

104. Gilbert, L.A., M.A. Horlbeck, B. Adamson, J.E. Villalta, Y. Chen, E.H. Whitehead, C. Guimaraes, B. Panning, H.L. Ploegh, M.C. Bassik, L.S. Qi, M. Kampmann, and J.S. Weissman, *Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation.* Cell, 2014. **159**(3): p. 647-61.

105. Konermann, S., M.D. Brigham, A.E. Trevino, J. Joung, O.O. Abudayyeh, C. Barcena, P.D. Hsu, N. Habib, J.S. Gootenberg, H. Nishimasu, O. Nureki, and F. Zhang, *Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex.* Nature, 2015. **517**(7536): p. 583-8.

106. Liu, S.J., M.A. Horlbeck, S.W. Cho, H.S. Birk, M. Malatesta, D. He, F.J. Attenello, J.E. Villalta, M.Y. Cho, Y. Chen, M.A. Mandegar, M.P. Olvera, L.A. Gilbert, B.R. Conklin, H.Y. Chang, J.S. Weissman, and D.A. Lim, *CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells.* Science, 2017. **355**(6320).

107. Lenstra, T.L., A. Coulon, C.C. Chow, and D.R. Larson, *Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription.* Mol Cell, 2015. **60**(4): p. 597-610.

108. Ghosh, S., C. Tibbit, and J.L. Liu, *Effective knockdown of Drosophila long non-coding RNAs by CRISPR interference.* Nucleic Acids Res, 2016.

109. Beiter, T., E. Reich, R.W. Williams, and P. Simon, *Antisense transcription: a critical look in both directions.* Cell Mol Life Sci, 2009. **66**(1): p. 94-112.

110. Sanjana, N.E., O. Shalem, and F. Zhang, *Improved vectors and genome-wide libraries for CRISPR screening.* Nat Methods, 2014. **11**(8): p. 783-4.

111. Guilinger, J.P., D.B. Thompson, and D.R. Liu, *Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.* Nat Biotechnol, 2014. **32**(6): p. 577-82.

112. Chen, B., L.A. Gilbert, B.A. Cimini, J. Schnitzbauer, W. Zhang, G.W. Li, J. Park, E.H. Blackburn, J.S. Weissman, L.S. Qi, and B. Huang, *Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system.* Cell, 2013. **155**(7): p. 1479-91.

113. Cheng, A.W., H. Wang, H. Yang, L. Shi, Y. Katz, T.W. Theunissen, S. Rangarajan, C.S. Shivalila, D.B. Dadon, and R. Jaenisch, *Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system.* Cell Res, 2013. **23**(10): p. 1163-71.

114. Hsu, P.D., D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, V. Agarwala, Y. Li, E.J. Fine, X. Wu, O. Shalem, T.J. Cradick, L.A. Marraffini, G. Bao, and F. Zhang, *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat Biotechnol, 2013. **31**(9): p. 827-32.

115. Bunz, F., A. Dutriaux, C. Lengauer, T. Waldman, S. Zhou, J.P. Brown, J.M. Sedivy, K.W. Kinzler, and B. Vogelstein, *Requirement for p53 and p21 to sustain G2 arrest after DNA damage.* Science, 1998. **282**(5393): p. 1497-501.

116. Goyal, A., K. Myacheva, M. Gross, M. Klingenberg, B. Duran Arque, and S. Diederichs, *Challenges of CRISPR/Cas9 applications for long non-coding RNA genes.* Nucleic Acids Res, 2016.

117. Hammerle, M., T. Gutschner, H. Uckelmann, S. Ozgur, E. Fiskin, M. Gross, B. Skawran, R. Geffers, T. Longerich, K. Breuhahn, P. Schirmacher, G. Stoecklin, and S. Diederichs, *Posttranscriptional destabilization of the liver-specific long noncoding RNA HULC by the IGF2 mRNA-binding protein 1 (IGF2BP1).* Hepatology, 2013. **58**(5): p. 1703-12.

118. Gutschner, T., M. Hammerle, N. Pazaitis, N. Bley, E. Fiskin, H. Uckelmann, A. Heim, M. Grobeta, N. Hofmann, R. Geffers, B. Skawran, T. Longerich, K. Breuhahn, P. Schirmacher, B. Muhleck, S. Huttelmaier, and S. Diederichs, *Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1) is an important protumorigenic factor in hepatocellular carcinoma.* Hepatology, 2014. **59**(5): p. 1900-11.

119. Kent, W.J., *BLAT--the BLAST-like alignment tool.* Genome Res, 2002. **12**(4): p. 656-64.

120. Cabili, M.N., M.C. Dunagin, P.D. McClanahan, A. Biaesch, O. Padovan-Merhar, A. Regev, J.L. Rinn, and A. Raj, *Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution.* Genome Biol, 2015. **16**: p. 20.

121. 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.

122. Washietl, S., I.L. Hofacker, M. Lukasser, A. Huttenhofer, and P.F. Stadler, *Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome.* Nat Biotechnol, 2005. **23**(11): p. 1383-90.

123. Weinberg, M.S. and K.V. Morris, *Transcriptional gene silencing in humans.* Nucleic Acids Res, 2016. **44**(14): p. 6505-17.

124. Bennett, C.F. and E.E. Swayze, *RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform.* Annu Rev Pharmacol Toxicol, 2010. **50**: p. 259-93.

125. Vickers, T.A., S. Koo, C.F. Bennett, S.T. Crooke, N.M. Dean, and B.F. Baker, *Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis.* J Biol Chem, 2003. **278**(9): p. 7108-18.

126. Luo, S., J.Y. Lu, L. Liu, Y. Yin, C. Chen, X. Han, B. Wu, R. Xu, W. Liu, P. Yan, W. Shao, Z. Lu, H. Li, J. Na, F. Tang, J. Wang, Y.E. Zhang, and X. Shen, *Divergent lncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells.* Cell Stem Cell, 2016. **18**(5): p. 637-52.

127. Hannus, M., M. Beitzinger, J.C. Engelmann, M.T. Weickert, R. Spang, S. Hannus, and G. Meister, *siPools: highly complex but accurately defined siRNA pools eliminate off-target effects.* Nucleic Acids Res, 2014. **42**(12): p. 8049-61.

128. Musahl, A.S., X. Huang, S. Rusakiewicz, E. Ntini, A. Marsico, G. Kroemer, O. Kepp, and U.A. Orom, *A long non-coding RNA links calreticulin-mediated immunogenic cell removal to RB1 transcription.* Oncogene, 2015. **34**(39): p. 5046-54.

129. Tsai, M.C., O. Manor, Y. Wan, N. Mosammaparast, J.K. Wang, F. Lan, Y. Shi, E. Segal, and H.Y. Chang, *Long noncoding RNA as modular scaffold of histone modification complexes.* Science, 2010. **329**(5992): p. 689-93.

130. Li, L., B. Liu, O.L. Wapinski, M.C. Tsai, K. Qu, J. Zhang, J.C. Carlson, M. Lin, F. Fang, R.A. Gupta, J.A. Helms, and H.Y. Chang, *Targeted disruption of Hotair leads to homeotic transformation and gene derepression.* Cell Rep, 2013. **5**(1): p. 3-12.

131. el-Deiry, W.S., T. Tokino, V.E. Velculescu, D.B. Levy, R. Parsons, J.M. Trent, D. Lin, W.E. Mercer, K.W. Kinzler, and B. Vogelstein, *WAF1, a potential mediator of p53 tumor suppression.* Cell, 1993. **75**(4): p. 817-25.

132. Kent, W.J., C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler, *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

133. Zhu, Y.Y., E.M. Machleder, A. Chenchik, R. Li, and P.D. Siebert, *Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction.* Biotechniques, 2001. **30**(4): p. 892-7.

134. Dunham I, K.A., Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee BK, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shoresh N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Dunham I, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kellis M, Khatun J, Kheradpour P, Kundaje A, Lassmann T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SC, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Pazin MJ, Lowdon RF, Dillon LA, Adams LB, Kelly CJ, Zhang J, Wexler JR, Green ED, Good PJ, Feingold EA, Bernstein BE, Birney E, Crawford GE, Dekker J, Elnitski L, Farnham PJ, Gerstein M, Giddings MC, Gingeras TR, Green ED, Guigó R, Hardison RC, Hubbard TJ, Kellis M, Kent W, Lieb JD, Margulies EH, Myers RM, Snyder M, Stamatoyannopoulos JA, Tenenbaum

SA, Weng Z, White KP, Wold B, Khatun J, Yu Y, Wrobel J, Risk BA, Gunawardena HP, Kuiper HC, Maier CW, Xie L, Chen X, Giddings MC, Bernstein BE, Epstein CB, Shoresh N, Ernst J, Kheradpour P, Mikkelsen TS, Gillespie S, Goren A, Ram O, Zhang X, Wang L, Issner R, Coyne MJ, Durham T, Ku M, Truong T, Ward LD, Altshuler RC, Eaton ML, Kellis M, Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Batut P, Bell I, Bell K, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena HP, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Li G, Luo OJ, Park E, Preall JB, Presaud K, Ribeca P, Risk BA, Robyr D, Ruan X, Sammeth M, Sandhu KS, Schaeffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Hayashizaki Y, Harrow J, Gerstein M, Hubbard TJ, Reymond A, Antonarakis SE, Hannon GJ, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR, Rosenbloom KR, Sloan CA, Learned K, Malladi VS, Wong MC, Barber GP, Cline MS, Dreszer TR, Heitner SG, Karolchik D, Kent W, Kirkup VM, Meyer LR, Long JC, Maddren M, Raney BJ, Furey TS, Song L, Grasfeder LL, Giresi PG, Lee BK, Battenhouse A, Sheffield NC, Simon JM, Showers KA, Safi A, London D, Bhinge AA, Shestak C, Schaner MR, Kim SK, Zhang ZZ, Mieczkowski PA, Mieczkowska JO, Liu Z, McDaniell RM, Ni Y, Rashid NU, Kim MJ, Adar S, Zhang Z, Wang T, Winter D, Keefe D, Birney E, Iyer VR, Lieb JD, Crawford GE, Li G, Sandhu KS, Zheng M, Wang P, Luo OJ, Shahab A, Fullwood MJ, Ruan X, Ruan Y, Myers RM, Pauli F, Williams BA, Gertz J, Marinov GK, Reddy TE, Vielmetter J, Partridge E, Trout D, Varley KE, Gasper C, Bansal A, Pepke S, Jain P, Amrhein H, Bowling KM, Anaya M, Cross MK, King B, Muratet MA, Antoshechkin I, Newberry KM, McCue K, Nesmith AS, Fisher-Aylor KI, Pusey B, DeSalvo G, Parker SL, Balasubramanian S, Davis NS, Meadows SK, Eggleston T, Gunter C, Newberry J, Levy SE, Absher DM, Mortazavi A, Wong WH, Wold B, Blow MJ, Visel A, Pennachio LA, Elnitski L, Margulies EH, Parker SC, Petrykowska HM, Abyzov A, Aken B, Barrell D, Barson G, Berry A, Bignell A, Boychenko V, Bussotti G, Chrast J, Davidson C, Derrien T, Despacio-Reyes G, Diekhans M, Ezkurdia I, Frankish A, Gilbert J, Gonzalez JM, Griffiths E, Harte R, Hendrix DA, Howald C, Hunt T, Jungreis I, Kay M, Khurana E, Kokocinski F, Leng J, Lin MF, Loveland J, Lu Z, Manthravadi D, Mariotti M, Mudge J, Mukherjee G, Notredame C, Pei B, Rodriguez JM, Saunders G, Sboner A, Searle S, Sisu C, Snow C, Steward C, Tanzer A, Tapanari E, Tress ML, van Baren MJ, Walters N, Washietl S, Wilming L, Zadissa A, Zhang Z, Brent M, Haussler D, Kellis M, Valencia A, Gerstein M, Reymond A, Guigó R, Harrow J, Hubbard TJ, Landt SG, Frietze S, Abyzov A, Addleman N, Alexander RP, Auerbach RK, Balasubramanian S, Bettinger K, Bhardwaj N, Boyle AP, Cao AR, Cayting P, Charos A, Cheng Y, Cheng C, Eastman C, Euskirchen G, Fleming JD, Grubert F, Habegger L, Hariharan M, Harmanci A, Iyengar S, Jin VX, Karczewski KJ, Kasowski M, Lacroute P, Lam H, Lamarre-Vincent N, Leng J, Lian J, Lindahl-Allen M, Min R, Miotto B, Monahan H, Moqtaderi Z, Mu XJ, O'Geen H, Ouyang Z, Patacsil D, Pei B, Raha D, Ramirez L, Reed B, Rozowsky J, Sboner A, Shi M, Sisu C, Slifer T, Witt H, Wu L, Xu X, Yan KK, Yang X, Yip KY, Zhang Z, Struhl K, Weissman SM, Gerstein M, Farnham PJ, Snyder M, Tenenbaum SA, Penalva LO, Doyle F, Karmakar S, Landt SG, Bhanvadia RR, Choudhury A, Domanus M,

Ma L, Moran J, Patacsil D, Slifer T, Victorsen A, Yang X, Snyder M, Auer T, Centanin L, Eichenlaub M, Gruhl F, Heermann S, Hoeckendorf B, Inoue D, Kellner T, Kirchmaier S, Mueller C, Reinhardt R, Schertel L, Schneider S, Sinn R, Wittbrodt B, Wittbrodt J, Weng Z, Whitfield TW, Wang J, Collins PJ, Aldred SF, Trinklein ND, Partridge EC, Myers RM, Dekker J, Jain G, Lajoie BR, Sanyal A, Balasundaram G, Bates DL, Byron R, Canfield TK, Diegel MJ, Dunn D, Ebersol AK, Frum T, Garg K, Gist E, Hansen R, Boatman L, Haugen E, Humbert R, Jain G, Johnson AK, Johnson EM, Kutyavin TV, Lajoie BR, Lee K, Lotakis D, Maurano MT, Neph SJ, Neri FV, Nguyen ED, Qu H, Reynolds AP, Roach V, Rynes E, Sabo P, Sanchez ME, Sandstrom RS, Sanyal A, Shafer AO, Stergachis AB, Thomas S, Thurman RE, Vernot B, Vierstra J, Vong S, Wang H, Weaver MA, Yan Y, Zhang M, Akey JM, Bender M, Dorschner MO, Groudine M, MacCoss MJ, Navas P, Stamatoyannopoulos G, Kaul R, Dekker J, Stamatoyannopoulos JA, Dunham I, Beal K, Brazma A, Flicek P, Herrero J, Johnson N, Keefe D, Lukk M, Luscombe NM, Sobral D, Vaquerizas JM, Wilder SP, Batzoglou S, Sidow A, Hussami N, Kyriazopoulou-Panagiotopoulou S, Libbrecht MW, Schaub MA, Kundaje A, Hardison RC, Miller W, Giardine B, Harris RS, Wu W, Bickel PJ, Banfai B, Boley NP, Brown JB, Huang H, Li Q, Li JJ, Noble WS, Bilmes JA, Buske OJ, Hoffman MM, Sahu AD, Kharchenko PV, Park PJ, Baker D, Taylor J, Weng Z, Iyer S, Dong X, Greven M, Lin X, Wang J, Xi HS, Zhuang J, Gerstein M, Alexander RP, Balasubramanian S, Cheng C, Harmanci A, Lochovsky L, Min R, Mu XJ, Rozowsky J, Yan KK, Yip KY, Birney E., The ENCODE Project Consortium,, *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

135. Rosenbloom, K.R., C.A. Sloan, V.S. Malladi, T.R. Dreszer, K. Learned, V.M. Kirkup, M.C. Wong, M. Maddren, R. Fang, S.G. Heitner, B.T. Lee, G.P. Barber, R.A. Harte, M. Diekhans, J.C. Long, S.P. Wilder, A.S. Zweig, D. Karolchik, R.M. Kuhn, D. Haussler, and W.J. Kent, *ENCODE data in the UCSC Genome Browser: year 5 update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D56-63.

136. Vassilev, L.T., B.T. Vu, B. Graves, D. Carvajal, F. Podlaski, Z. Filipovic, N. Kong, U. Kammlott, C. Lukacs, C. Klein, N. Fotouhi, and E.A. Liu, *In vivo activation of the p53 pathway by small-molecule antagonists of MDM2.* Science, 2004. **303**(5659): p. 844-8.

137. Villegas, V.E. and P.G. Zaphiropoulos, *Neighboring gene regulation by antisense long non-coding RNAs.* Int J Mol Sci, 2015. **16**(2): p. 3251-66.

138. Johnsson, P., A. Ackley, L. Vidarsdottir, W.O. Lui, M. Corcoran, D. Grander, and K.V. Morris, *A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells.* Nat Struct Mol Biol, 2013. **20**(4): p. 440-6.

139. Villamizar, O., C.B. Chambers, J.M. Riberdy, D.A. Persons, and A. Wilber, *Long noncoding RNA Saf and splicing factor 45 increase soluble Fas and resistance to apoptosis.* Oncotarget, 2016. **7**(12): p. 13810-26.

140. Xue, Z., Q. Ye, S.R. Anson, J. Yang, G. Xiao, D. Kowbel, N.L. Glass, S.K. Crosthwaite, and Y. Liu, *Transcriptional interference by antisense RNA is required for circadian clock function.* Nature, 2014. **514**(7524): p. 650-3.

141. Latos, P.A., F.M. Pauler, M.V. Koerner, H.B. Senergin, Q.J. Hudson, R.R. Stocsits, W. Allhoff, S.H. Stricker, R.M. Klement, K.E. Warczok, K. Aumayr, P. Pasierbek, and D.P. Barlow, *Airn transcriptional*

overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science, 2012. **338**(6113): p. 1469-72.

142. Modarresi, F., M.A. Faghihi, M.A. Lopez-Toledano, R.P. Fatemi, M. Magistri, S.P. Brothers, M.P. van der Brug, and C. Wahlestedt, *Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation.* Nat Biotechnol, 2012. **30**(5): p. 453-9.

143. Beltran, M., I. Puig, C. Pena, J.M. Garcia, A.B. Alvarez, R. Pena, F. Bonilla, and A.G. de Herreros, *A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition.* Genes Dev, 2008. **22**(6): p. 756-69.

144. Pandey, R.R., T. Mondal, F. Mohammad, S. Enroth, L. Redrup, J. Komorowski, T. Nagano, D. Mancini-Dinardo, and C. Kanduri, *Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation.* Mol Cell, 2008. **32**(2): p. 232-46.

145. Sauvageau, M., L.A. Goff, S. Lodato, B. Bonev, A.F. Groff, C. Gerhardinger, D.B. Sanchez-Gomez, E. Hacisuleyman, E. Li, M. Spence, S.C. Liapis, W. Mallard, M. Morse, M.R. Swerdel, M.F. D'Ecclessis, J.C. Moore, V. Lai, G. Gong, G.D. Yancopoulos, D. Frendewey, M. Kellis, R.P. Hart, D.M. Valenzuela, P. Arlotta, and J.L. Rinn, *Multiple knockout mouse models reveal lincRNAs are required for life and brain development.* eLife, 2013. **2**: p. e01749.

146. Kapranov, P., J. Drenkow, J. Cheng, J. Long, G. Helt, S. Dike, and T.R. Gingeras, *Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays.* Genome Res, 2005. **15**(7): p. 987-97.

147. Kapranov, P., A.T. Willingham, and T.R. Gingeras, *Genome-wide transcription and the implications for genomic organization.* Nat Rev Genet, 2007. **8**(6): p. 413-23.

148. Smeenk, L., S.J. van Heeringen, M. Koeppel, M.A. van Driel, S.J. Bartels, R.C. Akkers, S. Denissov, H.G. Stunnenberg, and M. Lohrum, *Characterization of genome-wide p53-binding sites upon stress response.* Nucleic Acids Res, 2008. **36**(11): p. 3639-54.

149. Chan, J., M. Atianand, Z. Jiang, S. Carpenter, D. Aiello, R. Elling, K.A. Fitzgerald, and D.R. Caffrey, *Cutting Edge: A Natural Antisense Transcript, AS-IL1alpha, Controls Inducible Transcription of the Proinflammatory Cytokine IL-1alpha.* J Immunol, 2015. **195**(4): p. 1359-63.

150. Matsui, M., Y. Chu, H. Zhang, K.T. Gagnon, S. Shaikh, S. Kuchimanchi, M. Manoharan, D.R. Corey, and B.A. Janowski, *Promoter RNA links transcriptional regulation of inflammatory pathway genes.* Nucleic Acids Res, 2013. **41**(22): p. 10086-109.

151. Postepska-Igielska, A., A. Giwojna, L. Gasri-Plotnitsky, N. Schmitt, A. Dold, D. Ginsberg, and I. Grummt, *LncRNA Khps1 Regulates Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure.* Mol Cell, 2015. **60**(4): p. 626-36.

152. Zhang, X., H. Li, J.C. Burnett, and J.J. Rossi, *The role of antisense long noncoding RNA in small RNA-triggered gene activation.* RNA, 2014. **20**(12): p. 1916-28.

153. Martianov, I., A. Ramadass, A. Serra Barros, N. Chow, and A. Akoulitchev, *Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript.* Nature, 2007. **445**(7128): p. 666-70.

154. Stojic, L., M. Niemczyk, A. Orjalo, Y. Ito, A.E. Ruijter, S. Uribe-Lewis, N. Joseph, S. Weston, S. Menon, D.T. Odom, J. Rinn, F. Gergely, and A. Murrell, *Transcriptional silencing of long noncoding RNA GNG12-AS1 uncouples its transcriptional and product-related functions.* Nat Commun, 2016. **7**: p. 10406.

155. Munroe, S.H. and M.A. Lazar, *Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA.* J Biol Chem, 1991. **266**(33): p. 22083-6.

156. Hastings, M.L., H.A. Ingle, M.A. Lazar, and S.H. Munroe, *Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring antisense RNA.* J Biol Chem, 2000. **275**(15): p. 11507-13.

157. Guttman, M. and J.L. Rinn, *Modular regulatory principles of large non-coding RNAs.* Nature, 2012. **482**(7385): p. 339-46.

158. Engreitz, J.M., J.E. Haines, E.M. Perez, G. Munson, J. Chen, M. Kane, P.E. McDonel, M. Guttman, and E.S. Lander, *Local regulation of gene expression by lncRNA promoters, transcription and splicing.* Nature, 2016. **539**(7629): p. 452-455.

159. Liu, P.C. and D.J. Thiele, *Novel stress-responsive genes EMG1 and NOP14 encode conserved, interacting proteins required for 40S ribosome biogenesis.* Mol Biol Cell, 2001. **12**(11): p. 3644-57.

160. Warda, A.S., B. Freytag, S. Haag, K.E. Sloan, D. Gorlich, and M.T. Bohnsack, *Effects of the Bowen-Conradi syndrome mutation in EMG1 on its nuclear import, stability and nucleolar recruitment.* Hum Mol Genet, 2016.

161. Meyer, B., J.P. Wurm, P. Kotter, M.S. Leisegang, V. Schilling, M. Buchhaupt, M. Held, U. Bahr, M. Karas, A. Heckel, M.T. Bohnsack, J. Wohnert, and K.D. Entian, *The Bowen-Conradi syndrome protein Nep1 (Emg1) has a dual role in eukaryotic ribosome biogenesis, as an essential assembly factor and in the methylation of Psi1191 in yeast 18S rRNA.* Nucleic Acids Res, 2011. **39**(4): p. 1526-37.

162. Eschrich, D., M. Buchhaupt, P. Kotter, and K.D. Entian, *Nep1p (Emg1p), a novel protein conserved in eukaryotes and archaea, is involved in ribosome biogenesis.* Curr Genet, 2002. **40**(5): p. 326-38.

163. James, A., Y. Wang, H. Raje, R. Rosby, and P. DiMario, *Nucleolar stress with and without p53.* Nucleus, 2014. **5**(5): p. 402-26.

164. Kruhlak, M., E.E. Crouch, M. Orlov, C. Montano, S.A. Gorski, A. Nussenzweig, T. Misteli, R.D. Phair, and R. Casellas, *The ATM repair pathway inhibits RNA polymerase I transcription in response to chromosome breaks.* Nature, 2007. **447**(7145): p. 730-4.

165. Calkins, A.S., J.D. Iglehart, and J.B. Lazaro, *DNA damage-induced inhibition of rRNA synthesis by DNA-PK and PARP-1.* Nucleic Acids Res, 2013. **41**(15): p. 7378-86.

166. McHugh, C.A., C.K. Chen, A. Chow, C.F. Surka, C. Tran, P. McDonel, A. Pandya-Jones, M. Blanco, C. Burghard, A. Moradian, M.J. Sweredoski, A.A. Shishkin, J. Su, E.S. Lander, S. Hess, K. Plath, and M. Guttman, *The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3.* Nature, 2015. **521**(7551): p. 232-6.

167. Chu, C., Q.C. Zhang, S.T. da Rocha, R.A. Flynn, M. Bharadwaj, J.M. Calabrese, T. Magnuson, E. Heard, and H.Y. Chang, *Systematic discovery of Xist RNA binding proteins.* Cell, 2015. **161**(2): p. 404-16.

168. Roth, A. and S. Diederichs, *Molecular biology: Rap and chirp about X inactivation.* Nature, 2015. **521**(7551): p. 170-1.

Appendix

4.7. Supplementary Figures

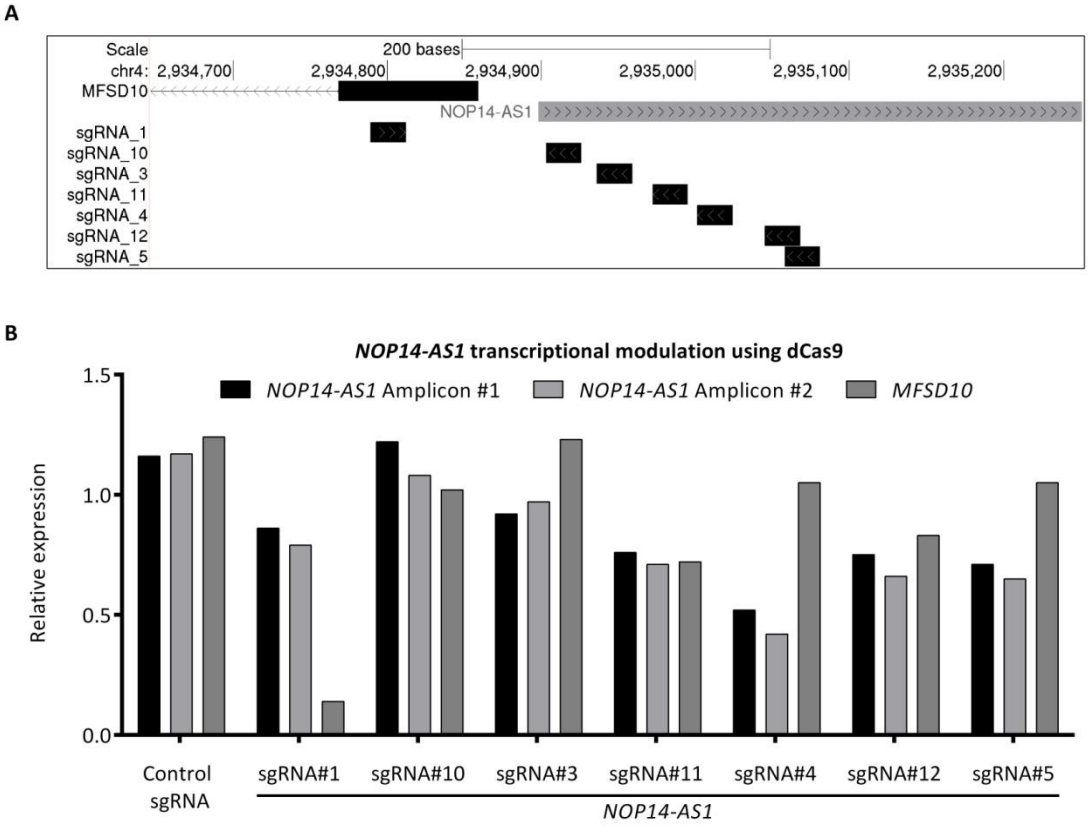


Figure 33: *NOP14-AS1* transcriptional modulation using dCas9

A: Schematic representation of the *MFSD10* / *NOP14-AS1* genomic locus depicting the sgRNAs used to target *NOP14-AS1*. **B:** A549 cells expressing dCas9 were transduced with either Control sgRNA or one of the seven indicated sgRNAs targeting *NOP14-AS1* promoter. RT-qPCR results for *NOP14-AS1* normalized to *Cyclophilin A* and control sgRNA. Data obtained from a single experiment. Not followed up upon due to the superior knockdown results with KRAB.

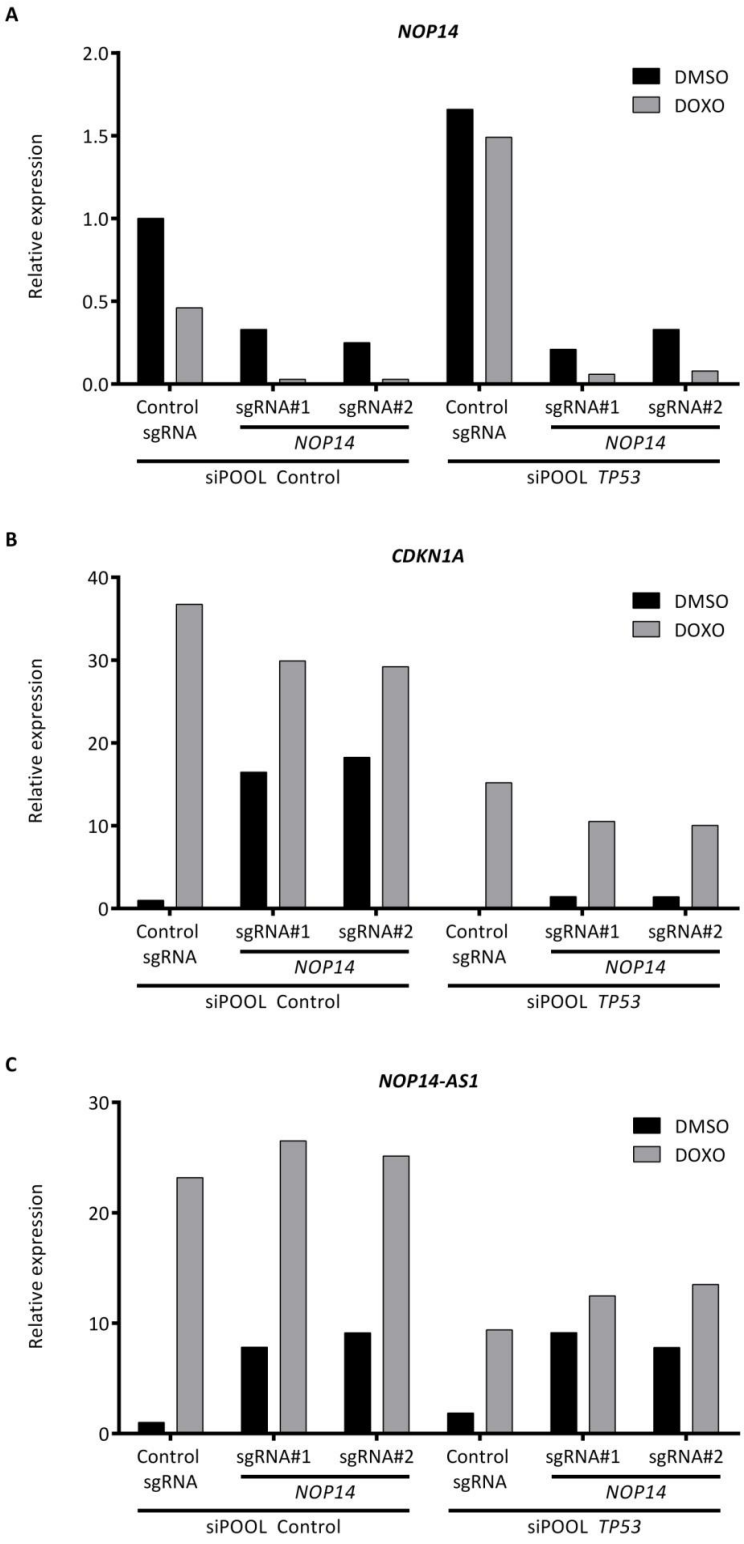


Figure 34: *CDKN1A* induction upon *NOP14* knockdown is p53-dependent

NCI-H460 cells were transduced with lentidCas9-KRAB-PURO iv sgRNA containing either a control sgRNA or one of the two independent sgRNAs targeting *NOP14*. These were transfected with either a control siPOOL or an siPOOL targeting *TP53* following which these treated with either 1 µM Doxorubicin (DOXO) or vehicle control DMSO for 24 hours. RT-qPCR results for (A) *NOP14*, (B) *CDKN1A* and (C) *NOP14-AS1* normalized to *Cyclophilin A* and control sgRNA + siPOOL control + DMSO control. Data obtained from a single experiment.

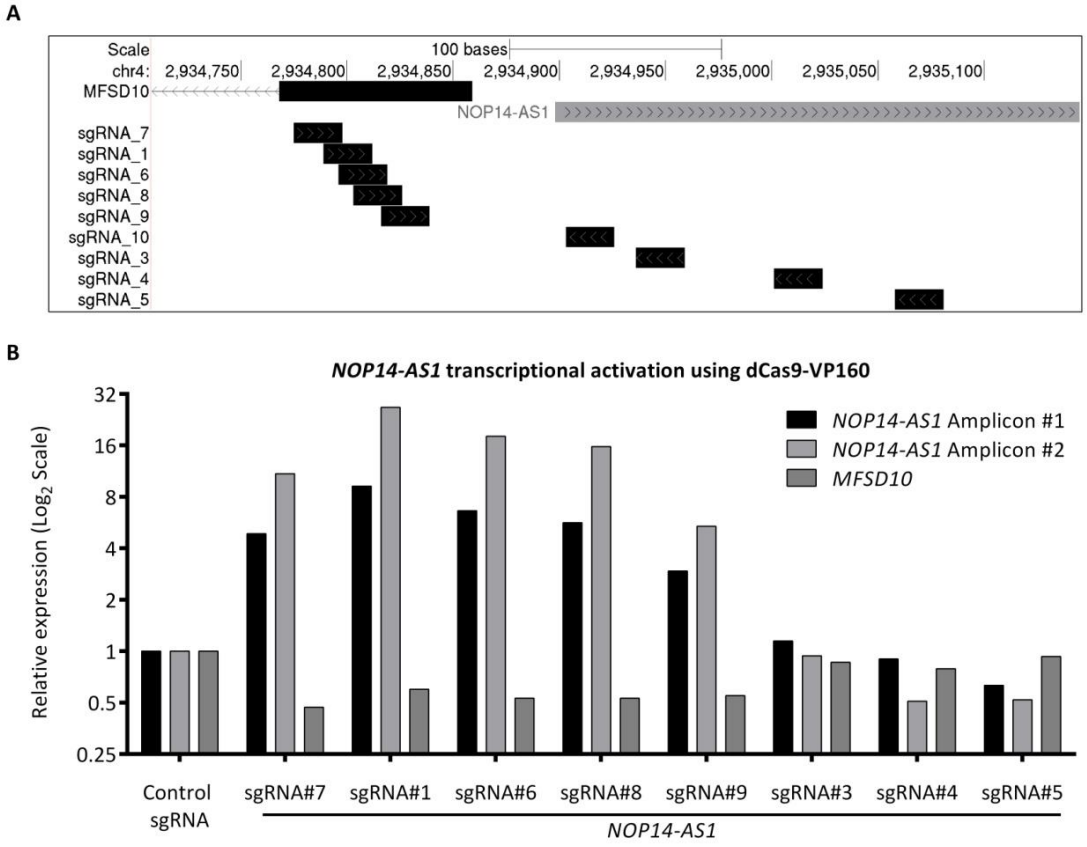


Figure 35: *NOP14-AS1* transcriptional activation using dCas9-VP160

A: Schematic representation of the *MFSD10* / *NOP14-AS1* genomic locus depicting the sgRNAs used to target *NOP14-AS1*.

B: A549 cells expressing dCas9-VP160 were transduced with either Control sgRNA or one of the eight indicated sgRNAs targeting *NOP14-AS1* promoter. RT-qPCR results for *NOP14-AS1* normalized to *Cyclophilin A* and control sgRNA. Data obtained from a single experiment. Not followed up upon as the aim was to identify two most efficient sgRNAs.

4.8. Supplementary sequences

Table 9: Common sequence features of the plasmids used in this study

Name	Sequence
dCas9	ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCG ACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAA GAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACC GCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGA TGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAA GCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACC ATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGG CCCTGGCCCACATGATCAAGTTCCGGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAG CGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATC AACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAA ATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCT GGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAG GACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTC TGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCAC CAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTG AAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACG GCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCT GGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAG CGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGC GGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCG CATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGC GAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCA TCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCT GTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAG CCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTC CGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGAC AAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGT TTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGAT GAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATC CGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACT TCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGG CCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATC CTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCG TGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAA GCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACC CAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGG AACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCCATCGTGCCTCAGAGCTTTCTGAAGGA CGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCC TCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCC AGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTT CATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGG ATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCA AGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCA CGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAA AGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGG AAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGAT TACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATC GTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCG TGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGA TAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTG GCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAG AGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGC CAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTG GAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGC

Table 9: Common sequence features of the plasmids used in this study

Name	Sequence
	CCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGA TAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATC AGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACA AGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCT GGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAA GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGT CTCAGCTGGGAGGCGAC
XTEN-Linker	AGCGGCAGCGAGACACCCGGCACAAGCGAGTCTGCCACCCCCGAGTCT
KRAB	AGAACACTGGTGACCTTCAAGGATGTATTTGTGGACTTCACCAGGGAGGAGTGGAAGCTGCTGGACA CTGCTCAGCAGATCGTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCTGGTTTCCTTGGGTTA TCAGCTTACTAAGCCAGATGTGATCCTCCGGTTGGAGAAGGGAGAAGAGCCCTGGCTGGTG
VP160	GACGCGCTGGACGATTTCGATCTCGACATGCTGGGTTCTGATGCCCTCGATGACTTTGACCTGGATA TGTTGGGAAGCGACGCATTGGATGACTTTGATCTGGACATGCTCGGCTCCGATGCTCTGGACGATTT CGATCTCGATATGTTAGGGTCAGACGCACTGGATGATTTCGACCTTGATATGTTGGGAAGCGATGCC CTTGATGATTTCGACCTGGACATGCTCGGCAGCGACGCCCTGGACGATTTCGATCTGGACATGCTGG GGTCCGATGCCTTGGATGATTTTGACTTGGATATGCTGGGGAGTGATGCCCTGGACGACTTTGACCT GGACATGCTGGGCTCCGATGCGCTCGATGACTTCGATTTGGATATGTTGGGATCC
P2A	GCAACAAACTTCTCTCTGCTGAAACAAGCCGGAGATGTCGAAGAGAATCCTGGACCG
PURO^{R}	ACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCGCGACGACGTCCCCAGGGCCGTACGCACCCTCG CCGCCGCGTTCGCCGACTACCCCGCCACGCGCCACACCGTCGATCCGGACCGCCACATCGAGCGGGT CACCGAGCTGCAAGAACTCTTCCTCACGCGCGTCGGGCTCGACATCGGCAAGGTGTGGGTCGCGGAC GACGGCGCCGCGGTGGCGGTCTGGACCACGCCGGAGAGCGTCGAAGCGGGGGCGGTGTTCGCCGAGA TCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCCTCCT GGCGCCGCACCGGCCCAAGGAGCCCGCGTGGTTCCTGGCCACCGTCGGAGTCTCGCCCGACCACCAG GGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGGTGCCCG CCTTCCTGGAGACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTCACCGTCACCGC CGACGTCGAGGTGCCCGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTGA
BLAST^{R}	ATGGCCAAGCCTTTGTCTCAAGAAGAATCCACCCTCATTGAAAGAGCAACGGCTACAATCAACAGCA TCCCCATCTCTGAAGACTACAGCGTCGCCAGCGCAGCTCTCTCTAGCGACGGCCGCATCTTCACTGG TGTCAATGTATATCATTTTACTGGGGGACCTTGTGCAGAACTCGTGGTGCTGGGCACTGCTGCTGCT GCGGCAGCTGGCAACCTGACTTGTATCGTCGCGATCGGAAATGAGAACAGGGGCATCTTGAGCCCCT GCGGACGGTGCCGACAGGTGCTTCTCGATCTGCATCCTGGGATCAAAGCCATAGTGAAGGACAGTGA TGGACAGCCGACGGCAGTTGGGATTCGTGAATTGCTGCCCTCTGGTTATGTGTGGGAGGGCTAA

Table 10: Sanger sequencing results of RACE clones

Name	Sequence
5' RACE Clone# 1	GGGATATGGCTCCCTGGCCAGCTTGTGAAGTTCAAGGGGCTCGTTACTTCTTACAGCTCACAGTCCC TCCCGGGTCCAGCACACCTGAGGGCTGACCTCCCTCTTGATTAGTGCAGGGGAAAAGCACTTTAAAA AGCTGTGTCATGCGTAACTTGCACATCTTCTGATTCTGACCTGAAACTGCATCCCAGGTCGGCACAG ACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCCA
5' RACE Clone# 2	ACAGCTCACAGTCCCTCCCGGGTCCAGCACACCTGAGGGCTGACCTCTCTCTTGATTAGTGCAGGGG AAAAGCACTTTAAAAAGCTGTGTCATGCGTAACTTGCACATCTTCTGATTCTGACCTGAAACTGCAT CCCAGGTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAA CCCCAGCACCCA
5' RACE Clone# 3	AGTGGCGGGGTTCTCGCCGGTCCTGCGGGTCGCTGCTCGGCTGTGGGTTAGCAGCACCCTGAGCGGA GCGGAAGAGCTGCCCGCCGCCGCGTAACCCCGAGGCCGAGACTCCTGGCCTGGCCTCCTTGGGCCCA CGACGCCTGCCCCGCGGAGCGGACCCATGGGCCTGAGGTCGGCACAGACGCAACCCAGTCTCCTCCC ACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCCA
5' RACE Clone# 4	AGTGGCATTCGGGATATGGCTCCCTGGCCAGCTTGTGAAGTTCAAGGGGCTCGTTACTTCTTACAGC TCACAGTCCCTCCCGGGTCCAGCACACCTGAGGGCTGACCTCCCTCTTGATTAGTGCAGGGGAAAAG CACTTTAAAAAGCTGTGTCATGCGTAACTTGCACATCTTCTGATTCTGACCTGAAACTGCATCCCAG GTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCA GCACCCA

Table 10: Sanger sequencing results of RACE clones

Name	Sequence
5' RACE Clone# 5	GGGATATGGCTCCCTGGCCAGCTTGTGAAGTTCAAGGGGCTCGTTACTTCTTACAGCTCACAGTCCC TCCCGGGTCCAGCACACCTGAGGGCTGACCTCCCTCTTGATTAGTGCAGGGGAAAAACACTTTAAAA AGCTGTGTCATGCGTAACTTGCACATCTTCTGATTCTGACCTGAAACTGCATCCCAGGTCGGCACAG ACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCC
5' RACE Clone# 6	AGTGGCATTCGGGATATGGCTCCCTGGCCAGCTTGTGAAGTTCAAGGGGCTCGTTACTTCTTACAGC TCACAGTCCCTCCCGGGTCCAGCACACCTGAGGGCTGACCTCCCTCCTGATTAGTGCAGGGGAAAAG CACTTTAAAAAGCTGTGTCATGCGTAACTTGCACATCTTCTGATTCTGACCTGAAACTGCATCCCAG GTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCA GCACCCA
5' RACE Clone# 7	AGTGGCATTCGGGATATGGCTCCCTGGCCAGCTTGTGAAGTTCAAGGGGCTCGTTACTTCTTACAGC TCACAGTCCCTCCCGGGTCCAGCACACCTGAGGGCTGACCTCCCTCTTGATTAGTGCAGGGGAAAAG CACTTTAAAAAGCTGTGTCATGCGTAACTTGCACATCTTCTGATTCTGACCTGAAACTGCATCCCAG GTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCA GCACCCA
5' RACE Clone# 8	CCCAACAAGCCTTGCACCACGAGGAAGTGGCAGTGGCATTCGGGATATGGCTCCCTGGCCAGCTTGT GAAGTTCAAGGGGCTCGTTACTTCTTACAGCTCACAGTCCCTCCCGGGTCCAGCACACCTGAGGGCT GACCTCCCTCTTGATTAGTGCAGGGGAAAAGCACTTTAAAAAGCTGTGTCATGCGTAACTTGCACAT CTTCTGATTCTGACCTGAAACTGCATCCCAGGTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCC GAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCCA
5' RACE Clone# 9	GGGAGACCGAGTAATAATAAAACCCCAGTCTCCCACACAGCTAGCTCTGCGTGAATTACCTTTTCTC TGTTGCAAATGCCCTTTCTCGATAAATTGGCTGTCTAGGTAACAGGGAAGGTGAACCCACTGGGCAG TTACACTTTTATCCACCCTTCCAGCTGGATCACGCGCCCTCTTCAATGACATGACACGCAGCTAGAG TGTATGAGCTTGCTGCCGGCACCTTCCGTCCTGGATCGCATAGCAAGGCCTCAATGGACAGCCAGGT CGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGC ACCCA
5' RACE Clone# 10	AGTGGCATTCGGGATATGGCTCCCTGGCCAGCTTGTGAAGTTCAAGGGGCTCGTTACTTCTTACAGC TCACAGTCCCTCCCGGGTCCAGCACACCTGAGGGCTGACCTCCCTCTTGATTAGTGCAGGGGAAAAG CACTTTAAAAAGCTGTGTCATGCGTAACTTGCACATCTTCTGATTCTGACCTGAAACTGCATCCCAG GTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCA GCACCCA
5' RACE Clone# 11	AGTGTTTGCCCTGGCTTGTAGCCCAGCGGGGGCTCCGTCATCCCCTGGGGAATGCCTCGGCCTGCAA GACCGCATACCGCATTGGAACAGGGAAACCACCTACTTCAGCACCTCCCTCAGCAAGGTGGCAGGTC CCAACAAGCCTTGCACCACGAGGAAGTGGCAGTGGCATTCGGGATATGGCTCCCTGGCCAGCTTGTG AAGTTCAAGGGGCTCGTTACTTCTTACAGCTCACAGTCCCTCCCGGGTCCAGCACACCTGAGGGCTG ACCTCCCTCTTGATTAGTGCAGGGGAAAAGCACCTTAAAAAGCTGTGTCATGCGTAACTTGCACATC TTCTGATTCTGACCTGAAACTGCATCCCAGGTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCTG AGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCCA
5' RACE Clone# 12	ATTGGAACAGGGAAACCACCTACTTCAGCACCTCCCTCAGCAAGGTGGCAGGTCCCAACAAGCCTTG CACCACGAGGAAGTGGCAGTGGCATTCGGGATATGGCTCCCTGGCCAGCTTGTGAAGTTCAAGGGGC TCGTTACTTCTTACAGCTCACAGTCCCTCCCGGGTCCAGCACACCTGAGGGCTGACCTCCCTCTTGA TTAGTGCAGGGGAAAAGCACTTTAAAAAGCTGTGTCATGCGTAACTTGCACATCTTCTGATTCTGAC CTGAAACTGCATCCCAGGTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTC CTGAATGGAGAACCCCAGCACCCA
5' RACE Clone# 13	CTCGCCGGTCCTGCGGGTCGCTGCTCGGCTGTGGGTTAGCAGCACCCTGAGCGGAGCGGAAGAGCTG CCCGCCGCCGCGTAACCCCGAGGCCGAGACTCCTGGCCTGGCCTCCTTGGGCCCACGACGCCTGCCC CGCGGAGCGGACCCATGGGCCTGAGGTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGT GCAGTGTCCTGAATGGAGAACCCCAGCACCCA
5' RACE Clone# 14	GACCGAGTAATAATAAAACTCCAGTCTCCCACACAGCTAGCTCTGCGTGAATTACCTTTTCTCTGTT GCAAATGCCCTTTCTTGATAAATTGGCTGTCTAGGTAACAGGGAAGGTGAACCCACTGGGCAGTTAC ACTTTTATCCACCCTTCCAGCTGGATCACGCGCCCTCTTCAATGACATGACACGCAGCTAGAGTGTA TGAGCTTGCTGCCGGCACCTTCCGTCCTGGATCGCATAGCAAGGCCTCAATGGACAGCCAGGTGCCT GGTCCGGTCCCACTTTTTTTTCTGTCCTGTTCTCGTTTCCTCTTTGCTTGGCCAAAGAGGGTTGGGC CAGGGTCCTGAGTGCAAGGGCTTGGCTTGGTGAATTAAAGTGGCAGTAGCCGTGTGGCAGAGAAAAT TTGTAGATGGGTGAAATCCTGGGCCTACACAGTGACAAAGTGTATATAGACAGCCAGCCTAGTAGGT GGATGTGACAACCCCAGACTCTTCTAGATTTGTGGATCCGTGCACCTGCCAGACCTGTCTCTGAGCC ATCTGGGGAGGCTGGGTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCT GAATGGAGAACCCCAGCACCCA
5' RACE	GGGAGACCGAGTAATAATAAAACTCCAGTCTCCCACACAGCTAGCTCTGCGTGAATTACCCTTTCTC

Table 10: Sanger sequencing results of RACE clones

Name	Sequence
Clone# 15	TGTTGCAAATGCCCTTTCTTGATAAATTGGCTGTCTAGGTAACAGGGAAGGTGGACCCACTGGGCAG TTACACTTTTATCCACCCTTCCAGCTGGATCACGCGCCCTCTTCAATGACATGACACGCAGCTAGAG TGTATGAGCTTGCTGCCGGCACCTTCCGTCCTGGATCGCATAGCAAGGCCTCAATGGACAGCCAGGT GCCTGGTCCGGTCCCACTTTTTTTTCTGTCCTGTTCTCGTTTCCTCTTTGCTTGGCCAAAGAGGGTT GGGCCAGGGTCCTGAGTGCAAGGGCTTGGCTTGGTGAATTAAAGTGGCAGTAGCCGTGTGGCAGAGA AAATTTGTAGATGGGTGAAATCCTGGGCCTACACAGTGACAAAGTGTATATAGACAGCCAGCCTAGT TGGTGGATGTGACAACCCCAGACTCTTCTAGATTTGTGGATCCGTGCACCTGCCAGACCTGTCTCTG AGCCATCTGGGGAGGCTGGGTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTG TCCTGAATGGAGAACCCCAGCACCCA
5' RACE Clone# 16	GCGGGTCGCTGCTCGGCTGTGGGTTAGCAGCACCCTGAGCGGAGCGGAAGAGCTGCCCGCCGCCGCG TAACCCCGAGGCCGAGACTCCTGGCCTGGCCTCCTTGGGCCCACGACGCCTGCCCCGCGGAGCGGAC CCATGGGCCTGAGGTTTGTGTGCGGGGCCTCAGGCGAGCGCTGGCGCAGCCTGTCCCCCATCGTCTG CGCCCTGCGGGCCTGGCCCTCCCCGTGGAAGGACCGCGGGCTTCGGGGCCCCACGCTGGGGGGCTGA CTCCTGCTACACACACAGCGGTGGCCGCGGGCCTGGGGCGGCAGCACAGCCATTCTGCTCGGGCTGC GTGAGCTGACCCCGCTCTCTGTGCACCTACTTTTGCTGCGTTTGGTGCTGAGCCTGAAGTCCCAGGA CAGGCGGCTGAGGGGGAAAGATGGTGTGAATTGGGTGGGCAGAGACAACTGAAAGCCACAGGACCCG GTTGGCACCTTTTTGGCCGCCTGCGCCTTTGCCTGCAGCCCCGCTCCCTGGGTCACCCGAAGGAGGA GCTGGCGCTCTTTCCGGGCGCAGAAGCACCCGACCCACGCTGGAGGAAACGGTCCCTCTGGGCAGGA GAAGCTTTGGGCCTGGCTGGAGCACACTCACGGAACCACGCCAACGCCTGGCCTGTAGCCAGCTCCC ACCTCCCCTCCCACCTTCTAAAATGCTTGCGGATCTCACGTGGCCGACCCTAACCAGATACACATGG GGAAAGGAATCAGGTGGCACACATGCCGCCTTTACTCCATAGAGAACTTCCAATTGTGGGTCGGCAC AGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCCA
5' RACE Clone# 17	GTGGGTTAGCAGCACCCTGAGCGGAGCGGAAGAGCTGCCCGCCGCCGCGTAACCCCGAGGCCGAGAC TCCTGGCCTGGCCTCCTTGGGCCCACGACGCCTGCCCCGCGGAGCGGACCCATGGGCCTGAGGTTTG TGTGCGGGGCCTCAGGCGAGCGCTGGCGCAGCCTGTCCCCCATCGTCTGCGCCCTGCGGGCCTGGCC CTCCCCGTGGAAGGACCGCGGGCTTCGGGGCCCCACGCTGGGGGGCTGACTCCTGCCACACACACAG CGGTGGCCGCGGGCCTGGGGCGGCAGCACAGCCATTCTGCTCGGGCTGCGTGAGCTGACCCCGCTCT CTGTGCACCTACTTTTGCTGCGTTTGGTGCTGAGCCTGAAGTCCCAGGACAGGCGGCTGAGGGGGAA AGATGGTGTGAATTGGGTGGGCAGAGACAACTGAAAGCCACAGGACCCGGTTGGCACCTTTTTGGCC GCCTGCGCCTTTGCCTGCAGCCCCGCTCCCTGGGTCACCCGAAGGAGGAGCTGGCGCTCTTTCCGGG CGCAGAAGCACCCGACCCACGCTGGAGGAAACGGTCCCTCTGGGCAGGAGAAGCTTTGGGCCTGGCT GGAGCACACTCACGGAACCACGCCAACGCCTGGCCTGTAGCCAGCTCCCACCTCCCCTCCCACCTTC TAAAATGCTTGCGGATCTCACGTGGCCGACCCTAACCAGATACACATGGGGAAAGGAATCAGGTGGC ACACATACCGCCTTTACTCCATAGAGAACTTCCAATCGTGGGTCGGCACAGACGCAACCCAGTCTCC TCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCCA
5' RACE Clone# 18	CCTGCGGGTCGCTGCTCGGCTGTGGGTTGGCAGCACCCTGAGCGGAGCGGAAGAGCTGCCCGCCGCC GCGTAACCCCGAGGCCGAGACTCCTGGCCTGGCCTCCTTGGGCCCACGACGCCTGCCCCGCGGAGCG GACCCATGGGCCTGAGGTTTGTGTGCGGGGCCTCAGGCGAGCGCTGGCGCAGCCTGTCCCCCATCGT CTGCGCCCTGCGGGCCTGGCCCTCCCCGTGGAAGGACCGCGGGCTTCGGGGCCCCACGCTGGGGGGC TGACTCCTGTCACACACACAGCGGTGGCCGCGGGCCTGGGGCGGTAGCACAGCCATTCTGCTCGGGC TGCGTGAGCTGACCCCGCTCTCTGTGCACCTACTTTTGCTGCGTTTGGTGCTGAGCCTGAAGTCCCA GGACAGGCGGCTGAGGGGGAAAGATGGTGTGAATTGGGTGGGCAGAGACAACTGAAAGCCACAGGAC CCGGTTGGCACCTTTTTGGCCGCCTGCGCCTTTGCCTGCAGCCCCGCTCCCTGGGTCACCCGAAGGA GGAGCCGGCGCTCTTTCCGGGCGCAGAAGCACCCGACCCACGCTGGAGGAAACGGTCCCTCTGGGCA GGAGAAGCTTTGGGCCTGGCTGGAGCACACTCACGGAACCACGCCAACGCCTGGCCTGTAGCCAGCT CCCACCTCCCCTCCCACCTTCTAAAATGCTTGCGGATCTCACGTGGCCGACCCTAACCAGATACACA TGGGGAAAGGAATCAGGTGGCACACATACCGCCTTTACTCCATAGAGAACTTCCAATTGTGGGTCGG CACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACC CA
5' RACE Clone# 19	CTCTTCCGGCCGGAGTCAGTGGCGGGGTTCTCGCCGGTCCTGCGGGTCGCTGCTTGGCTGTGGGTTA GCAGCACCCTGAGCGGAGCGGAAGAGCTGCCCGCCGCCGCGTAACCCCGAGGCCGAGACTCCTGGCC TGGCCTCCTTGGGCCCACGACGCCTGCCCCGCGGAGCGGACCCATGGGCCTGAGGTTTGTGTGCGGG GCCTCAGGCGAGCGCTGGCGCAGCCTGTCCCCCATCGTCTGCGCCCTGCGGGCCTGGCCCTCCCCGT GGAAGGACCGCGGGCTTCGGGGCCCCACGCTGGGGGGCTGACTCCTGCCACACACACAGCGGTGGCC GCGGGCCTGGGGCGGCAGCACAGCCATTCTGCTCGGGCTGCGTGAGCTGACCCCGCTCTCTGTGCAC CTACTTTTGCTGCGTTTGGTGCTGAGCCTGAAGTCCCAGGACAGGCGGCTGAGGGGGAAAGATGGTG TGAATTGGGTGGGCAGAGACAACTGAAAGCCACAGGACCCGGTTGGCACCTTTTTGGCCGCCTGCGC CTTTGCCTGCAGCCCCGCTCCCTGGGTCACCCGAAGGAGGAGCTGGCGCTCTTTCCGGGCGCAGAAG

Table 10: Sanger sequencing results of RACE clones

Name	Sequence
	CACCCGACCCACGCTGGAGGAAACGGTCCCTCTGGGCAGGAGAAGCTTTGGGCCTGGCTGGAGCACA CTCACGGAACCACGCCAACGCCTGGCCTGTAGCCAGCTCCCACCTCCCCTCCCACCTTCTAAAATGC TTGCGGATCTCACGTGGCCGACCCTAACCAGATACACATGGGGAAAGGAATCAGGTGGCACACATAC CGCCTTTACTCCATAGAGAACTTCCAATTGTGGGTAAGAACAAAGCCCCACCTCTTCCTAATAGATG CAGTTTTCCACCTGCAGCCAAAAAGACAACCTGTTATTAGCTTTTTTCCCAGAGGACAGGCATCTCT ATGAAACTGCCATTGCAAAATTTGCCGACAGTGAGATAACCAGCTCCATTTTGCTTCTAACCTCCAT GCCCTCCTTGTTTACTGCTGCACGCAGGTTGAACTGTGGGAGGAACTTAGTTTAAAACAAAGATGAT AACAGCCCTTTCCCAAAGGAAACCTTTCCTGGGGATTAGCCTGCTTTTGTGGGACCAACAGATTAGC CACAAGATTAGAAACTACGGTTGAAGAGTCACGCAGCGGGAGGCTACAGGATTCTGACCCTCCCCAA ACTGCTCCTGGCGATAACATCACCATCGTAAAGCCTAAGATCAGCACTTGGTGGGTCACCTGGCACC ACCCAGGCTCATCTGATGTTGTGGCACCCAGGAACTGACTCAGTGCGAAAACAGCTTCGACTCCCTG TGATTTCATCTCTGACCTGACCAATGGGTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGA GTGCAGTGTCCTGAATGGAGAACCCCAGCACCCA
5' RACE Clone# 20	CCTGCGGGTCGCTGCTTGGCTGTGGGTTAGCAGCACCCTGAGCGGAGCGGAAGAGCTGCCCGCCGCC GCGTAACCCCGAGGCCGAGACTCCTGGCCTGGCCTCCTTGGGCCCACGACGCCTGCCCCGCGGAGCG GACCCATGGGCCTGAGGTTTGTGTGCGGGGCCTCAGGCGAGCGCTGGCGCAGCCTGTCCCCCATCGT CTGCGCCCTGCGGGCCTGGCCCTCCCCGTGGAAGGACCGCGGGCTTCGGGGCCCCACGCTGGGGGGC TGACTCCTGCCACACACACAGCGGTGGCCGCGGGCCTGGGGCGGCAGCACAGCCATTCTGCTCGGGC TGCGTGAGCTGACCCCGCTCTCTGTGCACCTACTTTTGCTGCGTTTGGTGCTGAGCCTGAAGTCCCA GGACAGGCGGCTGAGGGGGAAAGATGGTGTGAATTGGGTGGGCAGAGACAACTGAAAGCCACAGGAC CCGGTTGGCACCTTTTTGGCCGCCTGCGCCTTTGCCTGCAGCCCCGCTCCCTGGGTCACCCGAAGGA GGAGCTGGCGCTCTTTCCGGGCGCAGAAGCACCCGACCCACGCTGGAGGAAACGGTCCCTCTGGGCA GGAGAAGCTTTGGGCCTGGCTGGAGCACACTCACGGAACCACGCCAACGCCTGGCCTGTAGCCAGCT CCCACCTCCCCTCCCACCTTCTAAAATGCTTGCGGATCTCACGTGGCCGACCCTAACCAGATACACA TGGGGAAAGGAATCAGGTGGCACACATACCGCCTTTACTCCATAGAGAACTTCCAATTGTGGGTAAG AACAAAGCCCCACCTCTTCCTAATAGATGCAGTTTTCCACCTGCAGCCAAAAAGACAACCTGTTATT AGCTTTTTTCCCAGAGGACAGGCATCTCTATGAAACTGCCATTGCAAAATTTGCCGACAGTGAGATA ACCAGCTCCATTTTGCTTCTAACCTCCATGCCCTCCTTGTTTACTGCTGCACGCAGGTTGAACTGTG GGAGGAACTTAGTTTAAAACAAAGATGATAACAGCCCTTTCCCAAAGGAAACCTTTCCTGGGGATTA GCCTGCTTTTGTGGGACCAACAGATTAGCCACAAGATTAGAAACTACGGTTGAAGAGTCACGCAGCG GGAGGCTACAGGATTCTGACCCTCCCCAAACTGCTCCTGGCGATAACATCACCATCGTAAAGCCTAA GATCAGCACTTGGTGGGTCACCTGGCACCACCCAGGCTCATCTGATGTTGTGGCACCCAGGAACTGA CTCAGTGCGAAAACAGCTTCGACTCCCTGTGATTTCATCTCTGACCTGACCAATGGTCGGCACAGAC GCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCCA
5' RACE Clone# 21	GAGACCGAGTAATAATAAAACTACAGTCTCCCACACAGCTAGCTCTGCGTGAATTACCTTTTCTCTG TTGCAAATGCNCTTTCTTGATAAATNNGCTGTCTAGGCAACAGGGAAGGTGAACCCACTGGGCAGTT ACACTTTTATCCACCCTTCCAGCTGGATCACGCGCCCTCTTCAATGACATGACACGCAGCTAGAGTG TATGAGCTTGCTGCCGGCACCTTCCGTCCTGGATCGCATAGCAAGGCCTCAATGGACAGCCAGGTGC CTGGTCCGGTCCCACTNNTTTTTCTGTCCTGTTCTCGTTTCCTCTTTGCTTGGCCAAAGAGGNTTGG GCCAGGGTCCTGAGTGCAAGGGCTTGGCTTAGTGAATTAAAGTGGCAGTAGCCGTGTGGCAGAGAAA ATNNNTAGATGGGTGAAATCCTGGGCCTNCACAGTGACAAAGTGTATATAGACAGCCAGCCTAGTTG GTGGATGTGACAACCCCAGACTCTTCTNGATTTGTGGATCCGTGCACCTGCCAGACCTGTCTCTGNG CCATCTGGGGAGGNNGGGTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGNGAGTGCAGTGTC CTGAATGGAGAACCCCAGCACCC
5' RACE Clone# 22	ACAGTGAGATAACCAGCTCCATTTTGCTTCTAACCTCCATGCCCTCCTTGTTTACTGCTGCACGCAG GTTGAACTGTGGGAGGAACTTAGTTTAAAACAAAGATGATAACAGCCCTTTCCCAAAGGAAACCTTT CCTGGGGATTAGCCTGCTTTTGTGGGACCAACAGATTAGCCACAAGATTAGAAACTACGGTTGAAGA GTCACGCAGCGGGAGGCTACAGGATTCTGACCCTCCCCAAACTGCTCCTGGCGATAACATCACCATC GTAAAGCCTAAGATCAGCACTTGGTGGGTCACCTGGCACCACCCAGGCTCATCTGATGTTGTGGCAC CCAGGAACTGACTCAGTGCGAAAACAGCTTCGACTCCCTGTGATTTCATCTCTGACCTGACCAATGG GTACTTCTGACTCACTGGCTGCCCCCCACCCACCAGCTTATCTTTAAAAACTCCAAACCCCAGATGC TCAGGGAGACCGAGTAATAATAAAACTCCAGTCTCCCACACAGCTAGCTCTGCGTGAATTACCTTTT CTCTGTTGCAAATGCCCTTTCTTGATAAATTGGCTGTCTAGGTAACAGGGAAGGTGAACCCACTGGG CAGTTACACTTTTATCCACCCTTCCAGCTGGATCACGCGCCCTCTTCAATGACATGACACGCAGCTA GAGTGTATGAGCTTGCTGCCGGCACCTTCCGTCCTGGATCGCATAGCAAGGCCTCAATGGACAGCCA GGTGCCTGGTCCGGTCCCACTTTTTTTTCTGTCCTGTTCTCGTTTCCTCTTTGCTTGGCCAAAGAGG GTTGGGCCAGGGTCCTGAGTGCAAAGGCTTGGCTTGGTGAATTAAAGTGGCAGTAGCCGTGTGGCAG AGAAAATTTGTAGATGGGTGAAATCCTGGGCCTACACAGTGACAAAGTGTATATAGACAGCCAGCCT

Table 10: Sanger sequencing results of RACE clones

Name	Sequence
	AGTTGGTGGATGTGACAACCCCAGACTCTTCTAGATTTGTGGATCCGTGCACCTGCCAGACCTGTCT CTGAGCCATCTGGGGAGGCTGGGTAAGATTTCCCCCGTCAGGGCCTGAGCCTCATCTCCAAGCTGCT GGCTTCAGCCCAGCTGCCACCAGCGCTGTCTCTGTCTTGCGCTTTCCAGAGAGAGACTGACTCATTT TGAGCTCCAGGCCTACCAGGCAGCCCAACCGCGGCAGTAACTGCCTCTCACACTGAGGATGGCTCAT GCTCTGGTCCAGCCATGGCTTGAACACCTCCCAACAAGAACAAAGCCAGACTCATTCTCCCTAGCAT CTTCTTCCTTGCTCAAAGATCCCACACCCCGTTAGAAGCTGGAACTGGGTCCTTGTCACCTTCCTCA CCTCCCCAGTCGTAGTCTGGTTAATCCTTACATCTCACCACGGTCACCGTGGCTAGATGTGACCTCC GCGATGCTTCAGGCACGTACCAGGGCAGCCTGGGCAGGGCTGGTCACTGTGATCTCATGAGGTGTGA ACTTGACCCCGGCCTTCTTTCCTGGGATGCCGGCTAAGGACACAAGTTCAGTGTTTGCCCTGGCTTG TAGCCCAGCGGGGGCTCCGTCATCCCCTGGGGAATGCCTCGGCCTGCAAGACCGCATACCGCATTGG AACAGGGAAACCACCTACTTCAGCACCTCCCTCAGCAAGGTGGCAGGTCCCAACAAGCCTTGCACCA CGAGGAAGTGGCAGTGGCATTCGGGATATGGCTCCCTGGCCAGCTTGTGAAGTTCAAGGGGCTCGTT ACTTCTTACAGCTCACAGTCCCTCCCGGGTCCAGCACACCTGAGGGCTGACCTCCCTCTTGATTAGT GCAGGGGAAAAGCACTTTAAAAAGCTGTGTCATGCGTAACTTGCACATCTTCTGATTCTGACCTGAA ACTGCATCCCAGGTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAA TGGAGAACCCCAGCACCCA
5' RACE Clone# 23	AGGGAGACCGAGTAATAATAAAACTCCAGTCTCCCACACAGCTAGCTCTGCGTGAATTACCTTTTCT CTGTTGCAAATGCCCTTTCTTGATAAATTGGCTGTCTAGGTAACAGGGAAGGTGAACCCACTGGGCA GTTACACTTTTATCCACCCTTCCAGCTGGATCACGCGCCCTCTTCAATGACATGACACGCAGCTAGA GTGTATGAGCTTGCTGCCGGCACCTTCCGTCCTGGATCGCATAGCAAGGCCTCAATGGACAGCCAGG TGCCTGGTCCGGTCCCACTTTTTTTTCTGTCCTGTTCTCGTTTCCTCTTTGCTTGGCCAAAGAGGGT TGGGCCAGGGTCCTGAGTGCAAGGGCTTGGCTTGGTGAATTAAAGTGGCAGTAGCCGTGTGGCAGAG AAAATTTGTAGATGGGTGAAATCCTGGGCCTACACAGTGACAAAGTGTATATAGACAGCCAGCCTAG TTGGTGGATGTGACAACCCCAGACTCTTCTAGATTTGTGGATCCGTGCACCTGCCAGACCTGTCTCT GAGCCATCTGGGGAGGCTGGGTCGGCACAGACGCAACCCAGTCTCCTCCCACCTCCGAGAGTGCAGT GTCCTGAATGGAGAACCCCAGCACCCA
3' RACE Clone# 1	GTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCCAGAAGCATCCACGG AAGGACTAAATTCTACACATGAAGGCAAACGTCCGTGGTTGACAGAAATTACACTGGCCAGAATCCC CAGTCCCCATGAGGCTTGTCCAGACGCAGTGAACCAGTCGCAGCTGATAACACACAGACCATTCCCG ATCCCAGAGGTGCATTTCAGGATTCATTCTATTTCATCAGAGACGGGGTCTTCCCGTGTTGCCCAGG CTGGTCTCGAACTCCTGGGCTGAAGCAATCTTCCTGCCTTGGCCTCCCAGAGGGTTGGAATTGCAGA TGTGAGGTGTGTCCTGGAACCCTGGCTGCACCCACAGATACAGCCACTCCTGGGCCCATGTGGATGT AACCTTCGAGTGCCGGAATGCCTCGTCTCCATGGAGACCAAACTCAAGGGCCTGACTGGCCCAGTCT ACCGAGACCATGAGACCTGCTGTGCCCCTCGGCCAGGCCAACCCAAATTCTTTTAAATAAAAGGCAC AAGCCACAGCCA
3' RACE Clone# 2	GTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCCAGAAGCATCCACTG AAGGACTAAATTCTACACATGAAGGCAAACGTCCGTGGTTGACAGAAATTACACTGGCCAGAATCCC CAGTCCCCATGAGGCTTGTCCAGACGCAGTGAACCAGTCGCAGCTGATAACACACAGACCATTCCCG ATCCCAGAGGTGCATTTCAGGATTCATTCTATTTCATCAGAGACGGGGTCTTCCTGTGTTGCCCAGG CTGGTCTCGAACTCCTGGGCTGAAGCAATCTTCCTGCCTTGGCCTCCCAGAGGGTTGGAATTGCAGA TGTGAGGTGTGTCCTGGAACCCTGGCTGCACCCACAGATACAGCCACTCCTGGGCCCATGTGGATGT AACCTTCGAGTGCCGGAATGCCTCGTCTCCATGGAGACCAAACTCAAGGGCCTGACTGGCCCAGTCT ACCGAGACCATGAGACCTGCTGTGCCCCTCGGCCAGGCCAACCCAAATTCTTTTAAATAAAAGGCAC AAGCCACAGCCACACCAGC
3' RACE Clone# 3	GTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCCAGAAGCATCCACTG AAGGACTAAATTCTACACATGAAGGCAAACGTCCGTGGTTGACAGAAATTACACTGGCCAGAATCCC CAGTCCCCACGAGGCTTGTCCAGACGCAGTGAACCAGTCGCAGCTGATAACACACAGACCATTCCCG ATCCCAGAGGTGCATTTCAGGATTCATTCTATTTCATCAGAGACGGGGTCTTCCTGTGTTGCCCAGG CTGGTCTCGAACTCCTGGGCTGAAGCAATCTTCCTGCCTTGGCCTCCCAGAGGGTTGGAATTGCAGA TGTGAGGTGTGTCCTGGAACCCTGGCTGCACCCACAGATACAGCCACTCCTGGGCCCATGTGGATGT AACCTTCGAGTGCCGGAATGCCTCGTCTCCATGGAGACCAAACTCAAGGGCCTGACTGGCCCAGACT ACCGAGACCATGAGACCTGCTGTGCCCCTCGGCCAGGCCAACCCAAATTCTTTTAAATAAAAGGCAC AAGCCACAGC
3' RACE Clone# 4	GTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCCAGAAGCATCCACTG AAGGACTAAATTCTACACATGAAGGCAAACGTCCGTGGTTGACAGAAATTACACTGGCCAGAATCCC CAGTCCCCATGAGGCTTGTCCAGACGCAGTGAACCAGTCGCAGCTGATAACACACAGACCATTCCCG ATCCCAGAGGTGCATTTCAGGATTCATTCTATTTCATCAGAGACGGGGTCTTCCTGTGTTGCCCAGG CTGGTCTCGAACTCCTGGGCTGAAGCAATCTTCCTGCCTTGGCCTCCCAGAGGGTTGGAATTGCAGA

Table 10: Sanger sequencing results of RACE clones

Name	Sequence
	TGTGAGGTGTGTCCTGGAACCCTGGCTGCACCCACAGATACAGCCACTCCTGGGCCCATGTGGATGT
	AACCTTCGAGTGCCGGAATGCCTCGTCTCCATGGAGACCAAACTCAAGGGCCTGACTGGCCCAGTCT
	ACCGAGACCATGAGACCTGCTGTGCCCCTCGGCCAGGCCAACCCAAATTCTTTTAAATAAAAGGCAC
	AAGCCACAGCCACACCAGCAGGAAGTGGATACAGGATAAGGGGCAGGGCTCACATCTCATCCTCATC
	AGCCTTTGACCAGGCGACGGAGTCTGTGATAGCTAGCTTATCTGCTCTATGCTCCAGGGACCTGGCC
	GGAACGAGCCGAGTTGCTGTCCTCATGGAACTCTCATCCTAGTGGGGCAGAGACAAGGACGGTGTAA
	CGTGCCCGGGGGTCAGCACCACAGACAAAACGCAGGGAGCACGGCACAGTGTGGGCCAGGGAAGGCC
	TGGCTTCCCAAGGGTGTGGGCGCGAGCAGCATGGCTGTGGGGAGTTCTAGGTGCAGAGAACTACTCC
	AGAGACGAAGCCAGGCCTGCGAGGAGGCCGGAGCGCATGAGGGGAAGGCACGAGGCAAGGCTGCCAG
	GGCTGTGCTGGTGGTGCGGGACGGTCCTGTGGGGGTAGCTGCGGAGATGGCACTGATGCCTTCGGCT
	CCAGCAGCATCCTGGGCTGGTGCGGGCCAATAAGGGGCAGGGTTGGGGCAGGATGAAAGGCACAAGC
	CACAGCCATGTCAAGCAGGAAGTGGATGCAGGGTACAGCCTGTCAGGGTCAGGGTGAAAGGCAGGGA
	GGCGGGTAGGCCATAGGTGAATTTGGAGCAGCCTGTAGAGGACAGGATGTGCCATCAAAGACTTGTT
	CTTGGGTCCAGCTGAGACTGTCCCTTGGGGCTTCCTGTCACCTTGTGCAGGATAGTGAGACCATCCT
	TGGAAAACCCATCGGGACTGCAGAGCGCGGCTCAGAGCTGACTGTCCAGGCCTGGCAGGGGCTGACC
	CCTGGGCGCTGGTGCACCCTCCGTGCGGCCTCTTGGAGGAAGCACACAAGGCGACAGTTCCCAGCTG
	CCGCCTGAGATGCAGCATCTGCTTTACTCCAGCCCAGCACTGCTCACCGTCACTGCCCACTCTTATG
	ATTTTACTTCCCTTCATATTTGTGATTTTGCAAAGATGGTGTTGACTTACAGCTGCAGCATCAAATG
	ACTGACTGCCACCAGCTTGGCCCCAGCTCAGGGACATGTCTGAGCCCAGGCAGAGCACCCTAGTGGC
	CGGGAAGCCTCACCTGGAGCTCCTGCGGGTGGCTGCAGTCCGCCAGGTGATCCGTGAGGAGGGCTTG
	GAGAGGCCCCATGATGGCGTGGAAGGATGGCAGGGACCCGTACATGAGCACGCAGCGCTTCAGCAGG
	GCCAGGCCCACAGCCAGGCAGGACAGTCTGTGAGGGCAGGAGGCAAGAGGAGGTCCAACTTGTTCTG
	TTTGTCACTGACAAAACCAATAACGTGGCAAAAATGAACTTCCCCACTTCCACTAGGCTGAAAACTC
	CAGTGTGGCCACCTTGGCCGGCCAGGGTTGGGCCCATGCAACCACGGGCAAGGCAGGCTCAGGGTCA
	GGCCTAACGCAGGAGGGGAGGCTGTGGGAACCGCAGCGCCAGCTGGGATGGGGCGGCTCCATTTTTG
	GTTCCAAATTATGATTT
	GTCTCCTCCCACCTCCGAGAGTGCAGTGTCCTGAATGGAGAACCCCAGCACCCAGAAGCATCCACTG
	AAGGACTAAATTCTACACATGAAGGCAAACGTCCGTGGTTGACAGAAATTACACTGGCCAGAATCCC
	CAGTCCCCATGAGGCTTGTCCAGACGCAGTGAACCAGTCGCAGCTGATAACACACAGACCATTCCCG
	ATCCCAGAGGTGCATTTCAGGATTCATTCTATTTCATCAGAGACGGGGTCTTCCTGTGTTGCCCAGG
	CTGGTCTCGAACTCCTGGGCTGAAGCAATCTTCCTGCCTTGGCCTCCCAGAGGGTTGGAATTGCAGA
	TGTGAGGTGTGTCCTGGAACCCTGGCTGCACCCACAGATACAGCCACTCCTGGGCCCATGTGGATGT
	AACCTTCGAGTGCCGGAATGCCTCGTCTCCATGGAGACCAAACTCAAGGGCCTGACTGGCCCAGTCT
	ACCGAGACCATGAGACCTGCTGTGCCCCTCGGCCAGGCCAACCCAAATTCTTTTAAATAAAAGGCAC
	AAGCCACAGCCACACCAGCAGGAAGTGGATACAGGATAAGGGGCAGGGCTCACATCTCATCCTCATC
	AGCCTTTGACCAGGCGACGGAGTCTGTGATAGCTAGCTTATCTGCTCTATGCTCCAGGGACCTGGCC
	GGAACGAGCCGAGTTGCTGTCCTCATGGAACTCTCATCCTAGTGGGGCAGAGACAAGGACGGTGTAA
	CGTGCCCGGGGGTCAGCACCACAGACAAAACGCAGGGAGCACGGCACAGTGTGGGCCAGGGAAGGCC
	TGGCTTCCCAAGGGTGTGGGCGCGAGCAGCATGGCTGTGGGGAGTTCTAGGTGCAGAGAACTACTCC
	AGAGACGAAGCCAGGCCTGCGAGGAGGCCGGAGCGCATGAGGGGAAGGCACGAGGCAAGGCTGCCAG
3' RACE Clone# 5	GGCTGTGCTGGTGGTGCGGGACGGTCCTGTGGGGGTAGCTGCGGAGATGGCACTGATGCCTTCGGCT
	CCAGCAGCATCCTGGGCTGGTGCGGGCCAATAAGGGGCAGGGTTGGGGCAGGATGAAAGGCACAAGC
	CACAGCCATGTCAAGCAGGAAGTGGATGCAGGGTACAGCCTGTCAGGGTCAGGGTGAAAGGCAGGGA
	GGCGGGTAGGCCATAGGTGAATTTGGAGCAGCCTGTAGAGGACAGGATGTGCCATCAAAGACTTGTT
	CTTGGGTCCAGCTGAGACTGTCCCTTGGGGCTTCCTGTCACCTTGTGCAGGATAGTGAGACCATCCT
	TGGAAAACCCATCGGGACTGCAGAGCGCGGCTCAGAGCTGACTGTCCAGGCCTGGCAGGGGCTGACC
	CCTGGGCGCTGGTGCACCCTCCGTGCGGCCTCTTGGAGGAAGCACACAAGGCGACAGTTCCCAGCTG
	CCGCCTGAGATGCAGCATCTGCTTTACTCCAGCCCAGCACTGCTCACCGTCACTGCCCACTCTTATG
	ATTTTACTTCCCTTCATATTTGTGATTTTGCAAAGATGGTGTTGACTTACAGCTGCAGCATCAAATG
	ACTGACTGCCACCAGCTTGGCCCCAGCTCAGGGACATGTCTGAGCCCAGGCAGAGCACCCTAGTGGC
	CGGGAAGCCTCACCTGGAGCTCCTGCGGGTGGCTGCAGTCCGCCAGGTGATCCGTGAGGAGGGCTTG
	GAGAGGCCCCATGATGGCGTGGAAGGATGGCAGGGACCCGTACATGAGCACGCAGCGCTTCAGCAGG
	GCCAGGCCCACAGCCAGGCAGGACAGTCTGTGAGGGCAGGAGGCAAGAGGAGGTCCAACTTGTTCTG
	TTTGTCACTGACAAAACCAATAACGTGGCAAAAATGAACTTCCCCACTTCCACTAGGCTGAAAACTC
	CAGTGTGGCCACCTTGGCCGGCCAGGGTTGGGCCCATGCAACCACGGGCAAGGCAGGCTCAGGGTCA
	GGCCTAACGCAGGAGGGGAGGCTGTGGGAACCGCAGCGCCAGCTGGGATGGGGCGGCTCCATTTTTG
	GTTCCAAATTATGATTTGTGAAATCAAGCAAATGTATTTTGAAAATGAAAAGAAAAAGGGACAGCAA

Acknowledgements

I would like to thank Prof. Dr. Sven Diederichs for providing me the opportunity to work on this project and his constant supervision. I am very grateful to him for providing funding and the encouragement to participate in relevant scientific meetings. I particularly appreciate his swiftness in arranging an access to S2 laboratory so that I could perform experiments involving lentiviruses.

Next, I would like to thank Prof. Dr. Christine Clayton and PD Dr. Thomas Hofmann, for their support as members of my thesis advisory committee. Their suggestions were very helpful in designing the experiments during the course of this study. I would also like to thank Dr. Julien Béthune for kindly agreeing to be a member of my thesis examination commission.

I would also like to thank Dr. Kai Breuhan and Prof. Dr. Peter Schirmacher for providing me with the antibodies and laboratory facilities. I would like to thank all the members of Dr. Stephanie Rössler's and Dr. Kai Breuhan's lab for being so cooperative during my stay with them.

I would like to thank all the members of Diederichs's lab, for their inputs on this project. Besides the scientific discussions, I am really grateful to them for introducing me to their cultures, and for sharing awesome food and drinks from their countries.

A big thank you to Matthias Groß for his constant help and technical support during the entire course of this study. I will always remember the partying and the awewome "canoue" jokes shared with you and Marcel.

I would also like to thank Marcel for the exciting collaboration on several projects and his valuable inputs on this project. I really enjoyed our discussions on scientific as well as political matters.

Special thanks to Steffen Link and Linda Nötzold for the time spent with them outside the lab. I would always cherish the memories of your visit to my home in India.

I would like to thank my parents, without whose support it would have been impossible to even imagine being where I am today.

Finally, I would like to thank my lovely wife – Manu. You supported me through the highs and the lows of this project and were a constant source of encouragement.

Statement

I hereby declare that the submitted dissertation is my own work. I only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given. I did not yet present this thesis or parts thereof to a university as part of an examination or degree. I confirm that the declarations made above are correct. I am aware of the importance of a sworn affidavit and the criminal prosecution in case of a false or incomplete affidavit. I affirm that the above is the absolute truth to the best of my knowledge and that I have not concealed anything.

Bei der eingereichten Dissertation handelt es sich um meine eigenständig erbrachte Leistung. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Ich habe wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt. Die Richtigkeit dieser Erklärung bestätige ich. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

_____ _____

Place and date Ashish Goyal