Universidad de Navarra

Facultad de Ciencias

Somatic Copy-Number Alterations across Human Cancers from a LncRNA Perspective

Alejandro Athie Cuervo

Universidad de Navarra

Facultad de Ciencias

Somatic Copy-Number Alterations across Human Cancers from a LncRNA Perspective

Memoria presentada por D./Dª Alejandro Athie Cuervo para aspirar al grado de Doctor por la Universidad de Navarra

El presente trabajo ha sido realizado bajo mi dirección en el Departamento de Terapia Génica y Regulación de la Expresión Génica y autorizo su presentación ante el Tribunal que lo ha de juzgar.

Pamplona, 20 de junio de 2017.

Dra. Maite Huarte Martínez

FIGURE INDEX

FIGURES INTRODUCTION

FIGURES RESULTS

FIGURES DISCUSSION

TABLE INDEX

LIST OF ABBREVIATIONS AND ACRONYMS

Abbreviation	Definition
7-AAD	7-aminoactinomycin D
aCGH	array comparative genomic hybridization
ADC	adenocarcinoma
ASO	antisense oligonucleotide
AS-UCHL1	ubiquitin carboxy-terminal hydrolase L1
BAFFR	B cell-activating factor receptor
BCA	bicinchoninic acid
BFB	break-fusion-bridge
BMI1	B lymphoma Mo-MLV insertion region 1 homolog
BP	biological processes
BRCA2	breast cancer type 2 susceptibility protein
BSA	Bovine serum albumine
CAGE	cap analysis of gene expression
Cas9	CRISPR associated protein 9
CBP	creb binding protein
CCLE	cance cell line encyclopedia
CCND1	cyclin-d1
CDKN1A	cyclin-dependent kinase inhibitor 1a
CGH	comparative genome hybridization
ChIP-seq	chromatin immunoprecipitation sequencing
ChIRP-MS	Chromatin Isolation by RNA Purification
CIN	chromosomal instability
CNA	copy number alteration
CPAT	coding-potential assessment tooL
CPC	coding potential calculator
CRISPR	clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
CV	crystal violet
CYLD	cylindromatosis
DAVID	database for annotation, visualization and integrated discovery
DC	dendritic cells
DD	death domain
DHFR	dihydroxyfolate reductase
DINO	damage induced noncoding
DLX2	distal-less homeobox 2
DMIN	double minute chromosome
DR	death receptor
DRCR	double rolling circle replication
DUB	deubiquitinase
ECL	enhanced chemiluminescence
EGFR	epidermal growth factor receptor

EMT	epithelial–mesenchymal transition
ENCODE	Encyclopedia of DNA elements
eRNA	enhancer RNA
FACS	fluorescence activated cell sorting
FAL1	focally amplified lncRNA on chromosome 1
FANTOM	functional annotation of the mammalian genome project
FDA	food and drug administration
Firre	functional intergenic repeating RNA element
FISH	fluorescence in situ hybridization
FLIP	flice-like inhibitory protein
Fostes	fork stalling and template switching
GAS5	growth arrest specific 5
GEO	gene expression omnibus
GISTIC	genome identification of significant target in cancer
GO	gene ontology
GR	glucocorticoid receptor
Her-2	human epidermal growth factor receptor-2
hnRNPK	Heterogeneous nuclear ribonucleoproein K
hnRNPU	Heterogeneous nuclear ribonucleoproein U
HNSC	head and neck squamous carcinoma
HOMD	homozygous deletion
HOTAIR	HOX transcript antisense RNA
HR	homologous recombination
HSF1	heat shock factor 1 protein
HSRs	homogeneous staining regions
HULC	highly upregulated in liver cancer
ICGC	international cancer genome consortium
IKB	inhibitor of kappa B
IL-1β	interleukin 1 beta
LBR	lamin-B receptor
LCC	larg-cell carcinoma
lncRNA	long noncoding RNA
lncRNA-LET	lncRNA low expression in tumors
LOH	loss of heterozygosity
LPS	lipopolysaccharide
LSD1	Lysine-Specific Demethylase 1
LUAD	lung adenocarcinoma
LUSC	lung squamous cell carcinoma
MALAT1	metastasis associated lung adenocarcinoma transcript 1
MeCP2	methyl-CpG-binding protein 2
MEKK3	mitogen-activated protein kinase kinase kinase 3
MF	molecular function
MIN	microsatellite instability
MMP	metalloproteinase

MMR	mismatch repair
mRNA	messenger RNA
MTS	3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium
MTX	methotrexate
NBs	nuclear bodies
NBAT-1	neuroblastoma-associated transcript 1
ncRNA-a	ncRNA-activating
NEAT1	nuclear paraspeckle assembly transcript 1
NELF	negative elongation factor
NEX	nuclear export sequence
NF90	nuclear factor 90
NFAT	nuclear factor of activated T-cells
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	next generation sequencing
NIK	NF-kappa-B-inducing kinase
NKILA	NF-KappaB interacting lncRNA
NLS	nuclear localization sequence
nt	nucleotide
NSCLC	non-small cell lung cancer
ORF	open reading frame
PANDAR	promoter of CDKN1A antisense DNA damage activated RNA
PCA3	prostate cancer antigen 3
PCAT-1	prostate cancer associated transcript 1
PCG	protein coding gene
PCGEM1	prostate-specific transcript 1
PDX	patient derived xenograft
piRNA	piwi-interacting RNAs
pol	polymerase
PRC2	Polycomb repressive complex
PSA	prostate specific antigen
PTMs	posstranslational modifications
PVT1	plasmacytoma variant translocation 1
qRT-PCR	quantiative real time polymerase chain reaction
RAP-MS	RNA Antisense Purification with Mass Spectrometry
RHD	Rel homology domain
Ribo-seq	ribosome profiling
RIP	RNA immunoprecipitation
RIP1	receptor-interacting protein 1
RMST	rhabdomyosarcoma 2 associated transcript
RNA-FISH	RNA fluorescente in situ hybridization
RNAi	RNA interference
rRNA	ribosomal RNA
SAMMSON	survival associated mitochondrial melamoma-specific oncogenic ncRNA

SART3	squamous cell carcinoma antigen recognized by T-cells 3
SCC	squamous cell carcinoma
SCF	skp, cullin, c-box containing complex
SCLC	small cell lung cancer
SCNA	somatic copy number alteration
sgRNA	single guide RNA
SHARP	SMRT/HDAC1-associated repressor protein
shRNA	short hairpin RNA
snoRNA	small nucleolar RNA
SNP	single nucleotide polymorphisms
snRNA	small nuclear RNA
SPHK1	sphingosine kinase 1
spliRNA	splicing sites RNA
SR	serine/arginine-rich
STAU1	staufen double-stranded RNA binding protein 1
TACE	tumor necrosis factor alpha converting enzyme
TAD	transactivation domain
TAK1	transforming growth factor beta-activated kinase 1
TCGA	the cancer genome atlas
TF	transcription factor
TINCR	terminal differentiation-induced ncRNA
TIR	transcription initiation regions
tiRNA	transcription initiation RNA
TLS	translocated in liposarcoma
TNFR1/2	tumor necrosis factor receptor 1/2
TNFα	tumor necrosis Afactor alpha
TRAF2	TNF receptor-associated factor 2
tRNA	transfer RNA
UTR	untranslated region
WGS	whole genome sequencing
XCI	X chromosome inactivation
Xic	X-inactivation centre
XIST	X inactive-specific transcript
YY1	yin yang 1

ABSTRACT

The genome of a tumor cell presents thousands of genomic alterations including base-substitutions and somatic copy number alterations (SCNAs). SCNAs comprise amplifications and deletions of big chromosomal regions usually containing hundreds of genes. Some of these regions harbor well-studied cancer drivers; however many others do not contain a known driver. The analysis of SCNA focusing on the non-coding genome helped us pinpoint a list of copy number altered long noncoding RNAs (lncRNAs). In order to validate our findings we experimentally characterized functionally and mechanistically a lncRNA amplified in lung cancer which we named LUAD-amp-1. LUAD-amp-1 acts as an oncogenic lncRNA, and its expression is induced by the transcription factor NF-κB upon TNFα treatment. Moreover, LUAD-amp-1 is implicated in the inhibition of a set of NF-κB regulated genes including TNFα itself. LUAD-amp-1 molecular mechanism relies on its association with SART3, altering its localization and modulating the nuclear translocation of its associated protein USP4.

1. INTRODUCTION

1.1 The RNA World

Back in 1986 Walter Gilbert [3] summarized the findings supporting the hypothesis of an 'RNA world' where RNA was the precursor molecule of all life on Earth. It was recognized that the first self-replicating system consisted of two components: RNA, responsible for storing the information, and proteins, which carry out the enzymatic activities. The discovery of RNAs which catalyze biochemical reactions such as phosphodiester bond cleavage in *E. Coli* [4] and self-splicing reaction in *Tetrahymena* [5] demonstrated that RNA could also carry out enzymatic reactions, making proteins dispensable in the primitive cells.

Even though there are many gaps remaining in the 'RNA world' hypothesis a fact that cannot be questioned is the diverse functionality of RNA for storing information as genes and catalyzing reactions as enzymes. These completely different functions demonstrate the potential that RNA has and the important roles that it plays in modern-day cells.

1.1.1 The Diverse Roles of RNA

Back in 1958 Francis Crick stated the central dogma of molecular biology describing the unidirectional flow of genetic information starting from DNA, followed by RNA, and finishing with proteins [6]. The intermediate step in this process is mediated by RNA, which acts as a messenger (mRNA) to carry the information from the nucleus to the cytoplasm. Crick suggested that once in the cytoplasm the RNA localized into 'microsomal particles', a nucleoprotein complex where 'the RNA forms the template and the protein supports and protects the RNA [6].' Now we know the microsomal particles he was referring to correspond to the ribosomes [7]; formed by proteins and mostly ribosomal RNAs (rRNAs) which catalyze the formation of new peptide bonds during protein synthesis.

Most importantly, the RNA also acts as a template for protein synthesis. However an open question remained on how the information encoded in the RNA template is translated to proteins. An additional hypothesis termed 'the adaptor hypothesis' proposed the existence of an adaptor molecule which carries the aminoacid and also recognizes the RNA template [6]. Around the same time in 1958, Zamecnik and Hoagland [8] discovered

this adaptor molecule termed transfer RNAs (tRNAs). So far three classes of RNAs were described: rRNAs, mRNAs and tRNAs, each one playing a specific role in the central dogma of molecular biology. The variety in functions demonstrated that RNA is more than just an intermediate molecule. In addition, many more types of RNAs remained to be discovered.

In the following years a new class of RNAs was discovered by biochemical fractionation experiments. This class of RNAs known as small nuclear RNAs (snRNAs) does not associate with the ribosome [9], and instead it is a component of the RNA splicing machinery. A similar class of small nucleolar RNAs (snoRNA) shares some characteristics with snRNAs such as its size and nuclear localization; however they are involved in posttranscriptional modifications of other RNAs. So far all the roles of the discovered RNA molecules were centered on protein synthesis.

1.1.2 RNAs as Regulatory Molecules

The idea that RNA could also act as a regulator came long time ago (1961), when Jacob and Monod speculated that the regulator/repressor of the lac operon could be a polyribonucleotide molecule [10]; further experiments showed that this was not the case, instead the repressor turned out to be a protein [11]. It was not until the discovery of microRNAs that a mechanism of RNA acting as a regulator was described. Lin-4 [12], and let-7 [13] are the founding members of this class of RNAs discovered in *C. elegans*. MicroRNAs exert their function by base pairing with the 3' untranslated region (UTR) of target RNAs to inhibit their translation and promote their degradation. One microRNA can regulate multiple targets. On the other hand, one RNA molecule can be regulated by several microRNAs. This interplay/network between RNA molecules adds up an extra level of gene regulation.

Few years later the RNA interference (RNAi) pathway was described and with it the potential of using RNA to silence gene expression was widely adopted. Since then a huge variety of small RNAs have been discovered starting with piwi-interacting RNAs (piRNAs) [14], involved in transposon silencing, and more recently small RNAs associated with transcription initiation (tiRNAs) [15] and splicing sites (spliRNAs) [16].

At this point it seemed that the regulatory RNA field was merely confined to RNAs of small size. It was not until the discovery of the X inactive-specific transcript (XIST), an RNA transcript with a critical function in X-chromosome inactivation, that long RNAs started to be explored [17, 18]. XIST can be easily confused with an mRNA since it is spliced and polyadenylated, however it lacks an open reading frame (ORF), and remains untranslated. These features defined a new class of RNAs known as long noncoding RNAs (lncRNAs). Examples of other lncRNAs (Rox [19], Tsix [20], Air [21]) started emerging in the coming years.

1.2 Discovery of LncRNAs

The advent of new sequencing technologies combined with the human genome project enabled the discovery and characterization of novel genes at the DNA and RNA levels.
Known protein coding gene exons account for less than 3% of the human genome [22]. The resulting 97% consists of intergenic and intronic sequences. Intergenic sequences were referred to as 'junk DNA' [23] since they largely consist of transposons and repetitive sequences that do not code for proteins and remain transcriptionally silent [24]. Later work demonstrated that intergenic regions were not gene deserts; in contrast, they are actively transcribed [25].

The first evidences of active transcription in intergenic DNA were obtained using microarrays with tiling probes [26] that mapped along entire chromosomes [27]. This approach combined with computational analysis helped demarcate actively transcribed chromatin domains [28].

Additional features obtained with chromatin immunoprecipitation coupled to high throughput deep sequencing (ChIP-seq) helped discover more transcriptional units [29], [30]. ChIP-seq enabled the mapping of chromatin marks along the entire genome. By analyzing these marks, clear patterns of transcriptional units were revealed: histone 3 lysine 4 trimethylation (H3K4me3) at gene promoters, and H3K36me3 in the gene bodies. Many of these transcriptional units were localized between protein genes; because of this they were named long intergenic (or intervening) noncoding RNA (lincRNAs).

Projects like FANTOM [25] and ENCODE [31] using additional technologies (cDNA cloning followed by Sanger sequencing, cap analysis of gene expression (CAGE), and RNA-seq) found out that more than 80% of the genome is actively transcribed [22]. So far 58,648 lncRNA loci have been annotated [32]. The standing question is whether all these transcripts have a biological function. A recent publication that integrates several features of lncRNAs such as conservation in transcription initiation regions (TIRs) and exons, overlap with trait-associated SNPs, and expression profiles, estimates that nearly 20,000 lncRNAs are potentially functional [33].

1.2.1 Classification and Characteristics of LncRNAs

LncRNAs have been arbitrarily defined as transcripts larger than 200 nucleotides in size, separating them from the small noncoding RNAs. Most of lncRNAs share similar features to mRNAs like RNA polymerase (pol) II dependent transcription, alternative splicing, 5' capping and 3' polyadenylation [34]. However, the major difference defining lncRNAs is their lack of protein coding potential.

Although lncRNAs conform a very heterogeneous class of transcripts, overall lncRNAs are shorter than protein coding genes, contain fewer exons, lack sequence conservation, have tissue-specific expression patterns, are lowly expressed [35] and less efficiently spliced than protein-coding genes [36].

According to its genomic location, lncRNAs can be classified into five groups: ***intergenic***, located between two protein-coding genes, which are the most abundant class of lncRNAs; ***intronic***, located inside the introns of protein coding-genes; ***divergent***, transcribed in the opposite direction of a protein-coding transcript without overlapping; ***antisense***, overlapping protein-coding genes but transcribed from the opposite strand; ***enhancer*** lncRNAs (eRNAs) are transcribed from enhancer regions and are usually monoexonic [37] (Figure 1) .

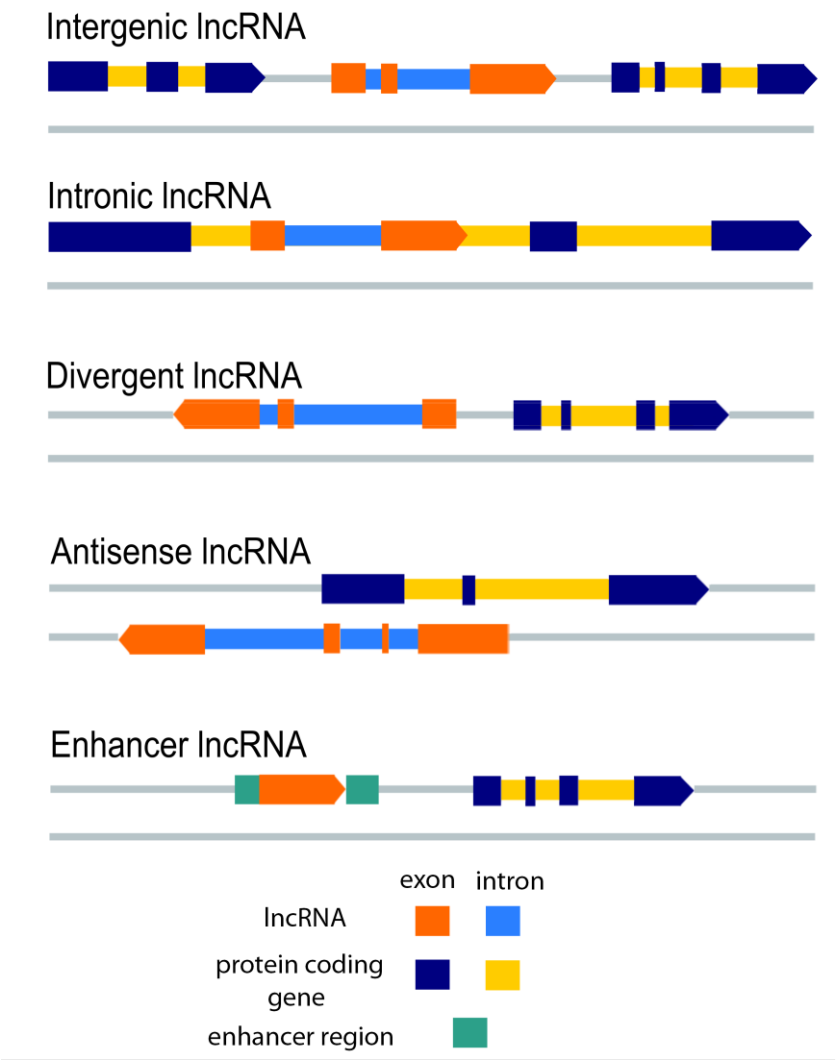

Figure 1. Classification of lncRNAs based on their genomic context. LncRNAs can be classified dependent on their genomic context. Intergenic lncRNAs are located between two protein-coding genes (PCG); intronic lncRNAs are encoded inside the intron of a PCG; divergent lncRNAs are transcribed in the opposite direction from a PCG without overlapping with it; antisense lncRNAs are transcribed from the opposite strand of a PCG, and can overlap it. Enhancer lncRNAs are transcribed from an enhancer region.

1.2.2 Distinguishing between coding and noncoding RNAs

A key aspect to define lncRNAs is their lack of protein coding potential. However, several criteria should be considered to properly separate coding from non-coding transcripts.

ORF length. ORFs can occur in any RNA sequence by chance; this is why the ORF length has become a fundamental criteria to distinguish lncRNAs from mRNAs [38]. The arbitrary threshold of > 300 nt (correspoding to 100 codons) has been used in annotation pipelines [39] based on the observation that > 95% of the known proteins are bigger than 100 aminoacids in length. A caveat for only using ORF length as a classification criteria is that > 300 nt ORFs can occur by chance in randomly generated RNA sequences bigger than 1000 nt. For example well characterized lncRNAs (*H19, Xist, KcnqOT1*) contain ORFs > 300 nt. Furthermore, under these criteria small proteins (< 100) can be misclassified as non-coding.

Nucleotide composition and substitution patterns. A particular feature of the ORFs encoding for proteins is their nucleotide composition defined by the non-random codon usage [40]. Several tools (CPC [41], CPAT [42]) exploit this criteria to distinguish between coding and noncoding transcripts. Furthermore, a nucleotide substitution pattern is also observed in protein coding sequences taking into consideration that these sequences are under selective pressure [43]. Multiple sequence alignments help identify a higher substitution frequency on the third position of codons [44].

Protein domains and similarity. Analysis of the possible products of transcripts in all three frames coupled to protein domain database searches [45] can help identify protein-coding transcripts. In addition, tools such as HMMER [46] enable the comparison of sequences to protein databases.

The combination of these criteria generates computational filters to predict the coding potential of a transcript. Complementary to the computational approach, experimental techniques such as ribosome profiling (Ribo-Seq) help map the RNAs associated with translating ribosomes. Surprisingly a substantial portion of the mouse lncRNAs efficiently associates with ribosomes [47], suggesting they could be translated into proteins.

Even though lncRNAs do associate with ribosomes they have a different ribosome occupancy profile compared to that of protein-coding transcripts. Upon encountering a stop codon, ribosomes are released from the translated transcript [48]; in Ribo-Seq experiments this process is reflected as a sharp decrease of signal after the stop codon. In contrast to protein-coding transcripts where translation termination can be identified, lncRNAs and other noncoding regions (5' UTR and 3' UTR) lack this signature [49]. So far integration analysis of mass spectrometry (MS) and RNA-seq data [50] have not identified peptides supporting lncRNAs translation. However, this could be due to the lack of sensibility of current MS technologies.

1.2.3 Evolution of lncRNAs

Evolutionary analyses could give some clues on the lncRNA functionality; these analyses have been hampered by the fact that lncRNA sequences evolved very rapidly compared to those of protein-coding genes [51] with > 70% of lncRNAs having no sequence similar orthologs in species that diverge > 50 million years ago [52]. Nevertheless lncRNA exons display weak but detectable signals of positive selection [53]. In addition, conservation analyses demonstrate that throughout evolution lncRNAs retain patches of highly conserved sequences [54], pointing out that these could be the functional elements of the transcript.

Even though there is a lack of sequence conservation some lncRNA homologs have been identified because they show syntenic conservation [52], [55]. Other lncRNAs like XIST are poorly conserved in most of its sequence but maintain the same exon-intron structure across different species.

An idea for explaining the low sequence conservation of lncRNAs is that their function might rely in the secondary structure level instead of the primary sequence. Recent publications using chemical probing followed by sequencing [56], demonstrate that indeed there is evolutionary conservation at the structure level [57] and that structural motifs are required for lncRNAs function [58].

1.2.4 Molecular Mechanisms of LncRNAs

LncRNAs mechanisms of action have been classified into *cis* acting or *trans* acting. The effects of *cis* acting lncRNAs are restricted to the chromosome from which they are transcribed, while *trans* acting lncRNAs function either by affecting genes on other chromosomes or even by travelling to the cytoplasm where they exert its function. Because of this *trans* acting lncRNAs are further classified into nuclear or cytoplasmic

Countless molecular mechanisms of lncRNAs have been described so far. Some of these mechanisms include: scaffolding protein complexes, recruiting proteins to specific regions in the genome, promoting loop formation between genomic regions, acting as decoys of miRNAs or proteins, stabilizing mRNAs, regulating proteins activity, among many others. Even though these mechanisms are not fully understood, what it is clear is that lncRNAs exert all these functions by interacting with DNA, other RNAs and proteins.

1.2.4.1 Epigenetic regulators

Epigenetic regulation mediated by lncRNAs is based on the capacity of lncRNAs to bind and act as scaffolds for chromatin modifying complexes and modulate their activity to reprogram the chromatin state. The best-studied lncRNA involved in epigenetic regulation is XIST. XIST plays an essential role in the mechanism of X chromosome inactivation (XCI), which results in dosage compensation of gene expression between female (XX) and male (XY).

The inactivation process is initiated by random expression of XIST from the X-inactivation centre (*Xic*) of one of the two female X chromosomes [59]. XIST spreads in *cis,* exclusively over the entire X-chromosome (Xi) from where it was transcribed, in a two-step mechanism first coating gene rich regions and then covering gene-poor domains [60]. It has been shown that during the spreading mechanism there is a proportional localization of polycomb repressive complex 2 (PRC2) and H3K27me3 to the sites where XIST bind, suggesting that PRC2 co-migrates with XIST to maintain the inactive state of Xi [61]. In addition to this working model, it has been demonstrated that XIST accumulates in a transcriptionally silent nuclear compartment, excluded of RNA pol II and enriched with heterochromatin marks [62] (Figure 2a).

The recent development of protocols to isolate RNA-protein complexes (ChIRP-MS [63], RAP-MS) [64]) enabled the discovery of novel XIST bound proteins, for example: SMRT/HDAC1-associated repressor protein (SHARP/SPEN), hnRNP-K, hnRNP-U, and lamin-B receptor (LBR) among others. Loss of function experiments demonstrated that SHARP binding to the A-repeat of XIST is required for Xi silencing.

Similar to XIST, other lncRNAs have been implicated in epigenetic regulation by similar mechanism of scaffolding and guiding chromatin modifying complexes [65] in *cis*. However, other lncRNAs use these mechanisms to regulate gene expression in *trans*, for example the lncRNA HOX transcript antisense RNA (HOTAIR). HOTAIR is transcribed from the HOXC locus and represses the expression of the HOXD cluster localized 40 kb apart [28], as well as genes located in other chromosomes [66]. It has been proposed that HOTAIR directly binds and guides PRC2 to target genes, and additionally acts as a scaffold between the PRC2 and Lysine-Specific Demethylase 1 (LSD1) complex. Although the significance of these findings is currently under active debate [67], HOTAIR mechanism of action paved the way for the discovery of many other *cis* and *trans* acting lncRNAs, which bind to chromatin modifying complexes. Some of the chromatin modifying complexes and the lncRNAs that associate with them are summarized in the following Table 1:

Chromatin Modifier	Function	LncRNAs	Reference
DNMT1	Methyltransferase, that binds and recognizes hemi-methylated DNA, inducing methylation and transcriptional silencing.	Dali, Dum, ecCEBP, DACOR1, LincRNA-p21	[68], [69], [70], [71], [72]
DNMT3b	Maintenance and *de novo* DNA methylation, which results in silencing.	Promoter-associated RNA (pRNA)	[73]
G9a	Histone methyltransferase catalyzes the methylation of H3K9, a repressive histone mark.	Kcnq1ot1, Air	[74], [75]
LSD1/CoREST	Catalyzes the demethylation of the H3K4me, acting as a corepressor of gene expression.	HOTAIR, HOXA-AS2	[76], [77]
Mediator	Multiprotein complex that acts as a transcriptional coactivator by looping chromatin, to bring distant regions of a chromosome into physical proximity.	ncRNA-activating (ncRNA-a)	[78]
MLL/WDR5	Histone methyltransferase complex, which methylates lysine 4 of histone 3, a mark of active transcription	HOTTIP, Fendrr	[79], [80], [81]
NoRC	Recruiter of enzymes for the formation of heterochromatin and silencing of rRNA genes.	NoRC-associated RNA	[82], [83]
PRC1	Proteins BMI1, RING1A/B. Recognizes methylated H3K27 and $H2A^{ub}$	XIST, MALAT1, ANRIL	[84], [85], [86]

PRC2	Histone methyltransferase complex formed by four core proteins: SUZ12, EED, RBBP4 and EZH2. The catalytic subunit EZH2 is responsible for trimethylating H3K27me3 a mark of silent chromatin.	XIST, HOTAIR, COLDAIR, MEG3, Kcnq1ot1, linc-Pint, ANRIL, lncRNA-ES1/2, MALAT1 many others…	[87], [28], [88], [89], [74], [90], [91], [92], [93]
SWI/SNF	Chromatin remodeling complex with ATPase activity. It slides nucleosomes, allowing access to the transcriptional machinery to activate gene expression.	SChLAP1, UCA1, NEAT1, Evf2, HIF1A-AS1, lncTCF7, COX2	[94], [95], [96], [97], [98], [99], [100]

Table 1 Chromatin modifiers and lncRNAs

1.2.4.2 Transcriptional regulators

In addition to chromatin modifying complexes, lncRNAs can also bind to transcription factors (TF). Even though transcription factors, by themselves, are able to recognize and bind DNA sequences, lncRNAs can modulate this association.

LncRNAs can act as a decoy and inhibit the DNA binding of their associated TF. For example, the lncRNA growth arrest-specific 5 (GAS5) [101] associates with the DNA-binding domain of the glucocorticoid receptor (GR), competing with the binding of GR to the DNA. Similar mechanisms have been described for other lncRNAs. *PANDAR* associates to nuclear transcription factor Y (NF-YA) [102]; *Jpx* binds to CTCF removing it from *Xist* promoter region [103]; and Lethe associates to nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) and inhibits its binding to the NF-κB response elements, resulting in a lack of expression of target genes (*IL6, IL8, NFKBIA*) [104] (Figure 2b).

Moreover, lncRNAs can also act as direct cofactors of TF, for example: *DINO* binds to p53 stabilizing it, what results in the enhancement of DNA damage signaling [105]. Another example is the lncRNA *HSR1* (heat shock RNA 1) acts as a coactivator of the heat shock factor 1 (HSF1) protein [106]. The lncRNA *EVF2* either associates to Distal-Less Homeobox

2 (DLX2) or to the methylCpG binding-protein 2 (MeCP2) resulting in transcriptional activation or repression of specific DNA regions, respectively [107], [108]. The lncRNA (rhabdomyosarcoma 2-associated transcript) *RMST* is required for the binding of SOX2 to the promoter of genes involved in neurogenesis. [109].

LncRNAs can also regulate TF activity indirectly. For example, a set of DNA damage induced lncRNAs, which are transcribed from the cyclin D1 gene promoter (*CCND1*), associate to the RNA-binding protein, translocated in liposarcoma (TLS). This association promotes the binding of TLS to the CREB-binding protein (CBP)/p300, what results in inhibition of the acetyltransferase activity of p300 followed by repression of the *CCND1* gene [110]. LncRNAs can also regulate indirectly the function of TF by altering its cellular localization. An example is the lncRNA *NRON* which binds the TF nuclear factor of activated T cells (NFAT), keeping it in the cytoplasm [111].

1.2.4.3 Nuclear architects

Besides the genomic regulation mediated by DNA sequences and epigenetic marks it has been recognized that the architectural organization of the cell nucleus also plays an important role. The nucleus is organized into domains that spatially arrange genomic sequences and concentrate protein complexes. Furthermore, these domains are associated with functional roles (DNA replication and repair, gene transcription, RNA processing and mRNA transport).

Recent studies have shown that several lncRNAs exploit their scaffolding and guiding mechanism to assemble nuclear domains. One example is the functional intergenic repeating RNA element (*Firre*) lncRNA [112] that forms a nuclear compartment involving its own locus in the X chromosome and other loci in chromosome 2, 9, 15, and 17. Interestingly these trans-genomic loci include many genes related to energy metabolism, suggesting that the formation of this nuclear compartment is not a random process. The establishment of these inter-chromosomal interactions depends on the binding of hnRNPU with the repetitive sequences present in *Firre* (Figure 2c). More striking is the case of large scale remodeling and repositioning of the Xi chromosome mediated by *XIST*. Upon XCI, Xi adopts a chromosomal architecture characterized by two macro-domains [113] and long range looping interactions. Deletion of *XIST* results in a conformational

change of Xi similar to the one of the active X (Xa), suggesting the important role of *XIST* not only in the establishment of Xi silencing but also on the maintenance of Xi's 3-D structure [114]. Furthermore, it has been show that XIST interaction with LBR mediates the re-localization of Xi to the nuclear lamina, a region where gene expression is silenced [115].

LncRNAs are also involved in the formation and function of nuclear bodies (NB). NBs are dynamic domains enriched for particular components (enzymes, substrates) and associated to specific loci. Well-characterized NBs include paraspeckles, nuclear speckles and nucleolus among others. Paraspeckles are NBs involved in the accumulation of adenosine-to-inosine (A-to-I) edited mRNAs. The precise mechanism of paraspeckle formation remains unclear; however the lncRNA (nuclear paraspeckle assembly transcript 1) NEAT1 is essential for their organization and maintenance [116]. Paraspeckle formation is dependent on: 1) NEAT1's act of transcription, and 2) NEAT's association with NONO, SFPQ, and PSPC1 [117]. One of the most abundant lncRNAs, the metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) is also associated with NBs, specifically to nuclear speckles. It has been proposed that MALAT1 recruits pre-mRNA splicing factors to the nuclear speckles, in particular serine-arginine rich (SR) proteins this way it regulates the gene expression of the associated transcripts [118]. In contrast to its neighbor gene NEAT1, MALAT1 is not essential for nuclear speckles formation. Genome wide mapping of the binding sites of NEAT1 and MALAT1 showed and association with actively transcribed loci, suggesting they could be acting as scaffolds to form domains enriched with active genes [119].

Nuclear architecture is also composed by chromatin loops that bring together enhancers with promoters resulting in transcriptional gene activation. Early studies reported the transcription at enhancers of a class of lncRNAs termed eRNAs [120]. Some eRNAs are involved in the looping formation mechanism, they accomplish this task by binding to different protein partners. Interestingly, a set of eRNAs mediates the recruitment of cohesins to enhancers facilitating the formation of stable enhancer-promoter loops [121]. For example the lncRNA CCAT1-L is transcribed from an enhancer region close to the *MYC* locus, and interacts with CCCTC-binding factor (CTCF); these interactions promote the

chromosomal looping between the enhancer and the *MYC* promoter resulting in upregulation of *MYC* transcription [122]. A similar mechanism has been reported for another class of lncRNA named ncRNA-a. In contrast to eRNA, ncRNAs-a are not transcribed from enhancer regions; however they have an enhancer-like function of activating other genes [123]. ncRNAs-a interact with the coactivator complex Mediator what results in the formation of DNA loops between the ncRNA-a loci and its targets [78].

In other cases it has been shown that eRNAs regulate the enhancer function independent of chromosome looping, implicating eRNAs in the regulation of chromatin accessibility of target promoters and the subsequent RNA pol II binding [124], [125]. Moreover, eRNAs can also regulate the process of RNA pol II pause release by acting as decoys and capturing the (negative elongation factor) NELF complex, which represses transcriptional elongation [126]. Another recently proposed mechanism termed 'transcription factor trapping' suggest that the transcript *per se* is involved in stabilizing the binding of TFs (YY1) to enhancer regions ([127]). Together these studies demonstrate that eRNAs act as regulators of gene activation at different stages.

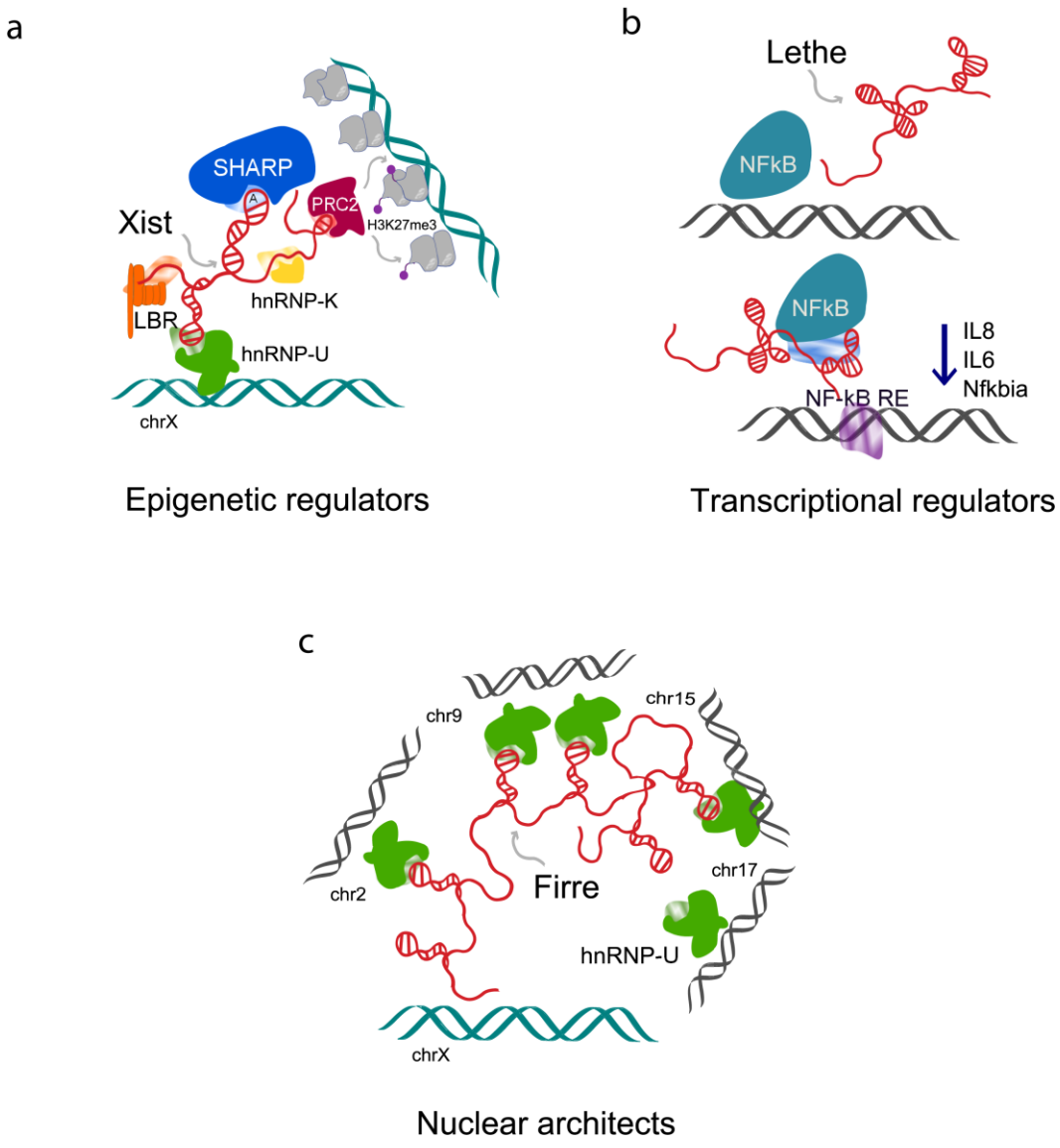

Figure 2. Proposed molecular mechanism for some lncRNAs acting as: epigenetic regulators, transcription factor regulators and nuclear architects. (a) Molecular mechanism proposed for chromosome X-inactivation mediated by the lncRNA *Xist*. *Xist* associates to the chromatin by binding to heterogeneous nuclear ribonucleoprotein U (hnRNPU) and promotes the recruitment of the SMRT/HDAC1-associated repressor protein (SHARP) through its A-repeat region (A). *Xist* also interacts with PRC2, complex in charge of depositing the epigenetic repressive mark H3K27me3. For the recruitment of PRC2, the heterogeneous nuclear ribonucleoprotein K (hnRNPK) is required. The nuclear lamin B receptor (LBR) also binds *Xist* and could be required for structural remodeling to start the transcriptional silencing. (b) The lncRNA *Lethe* is transcriptionally regulated by NF-κB; moreover Lethe exerts its function by directly binding to the transcription factor NF-κB and blocking its association with the NF-κB response elements (RE), inhibiting the induction of inflammatory genes (*IL8, IL6, Nfkbia*). (c) The lncRNA *Firre* is anchored to chromatin by interacting with hnRNPU. It is transcribed from the X chromosome and simultaneously interacts with chromosome 2, 9, 15 and 17. It has been proposed to act as a scaffold that organizes chromosomal architecture.

All the mechanisms previously described explain the role of lncRNAs in regulation of gene expression at the transcriptional level. However lncRNAs are also involved in posttranscriptional regulation processes such as mRNA stability, translational control and protein modifications.

1.2.4.4 mRNA stabilizers

By exploiting the capacity to recognize specific sequences lncRNAs can associate to mRNAs. This association can have a positive or negative impact on mRNA stability. For example, the half-STAU1-binding site RNAs (1/2-sbsRNAs) act by base pairing to the 3' UTR of target mRNAs via an Alu element, what results in the formation of a dsRNA regions. The dsRNA then recruits the protein staufen1 (STAU1), which promotes mRNA decay [128].

In contrast to the decay mechanism, STAU1 can also stabilize mRNAs by interacting with the lncRNA *TINCR* [129]. *TINCR* binds to a group of mRNAs involved in differentiation via a 25-nucleotide motif called the *TINCR* box. The exact mechanism on how mRNA stabilization, mediated by STAU1-*TINCR*, is accomplished requires more investigation. However these two examples (1/2-sbsRNAs and *TINCR*) demonstrate the potential lncRNAs have as mRNA stability regulators.

LncRNAs can also regulate mRNA stability indirectly, by acting as decoys of miRNAs or proteins, preventing the binding to their targets. Examples of this mechanism include the lncRNA *linc-MD1*, *linc-ROR* and *NORAD*. (Figure 3a)

1.2.4.5 Translation regulators

Translational regulation is an important step in gene expression where lncRNAs are also involved. LncRNAs can act either as promoters or inhibitors of translation.

The antisense transcript (*AS-Uchl1*) enhances the translation of the ubiquitin carboxyl-terminal esterase L1 (Uchl1) mRNA, resulting in an increase of UCHL1 protein levels [130]. The proposed mechanism to explain this enhancement involves the binding of *AS-Uchl1* with an overlapping region in the 5' of Uchl1 and a sequence repeat (SINEB2) present in *AS-Uchl1*, what results in the targeting of Uchl1 mRNA to active polysomes. In contrast, lncRNAs can also downregulate the levels of a protein by inhibiting its translation; such is

the case of lincRNA-p21, which directly binds *CTNNB1* and *JUNB* mRNAs suppressing their translation [131] (Figure 3b).

1.2.4.6 Regulation of protein posttranslational modifications

An extra level of regulation that can also be modulated by cytoplasmic lncRNAs involves protein posttranslational modifications (PTMs). One example is lnc-DC (dendritic cells) that promotes STAT3 signaling by directly interacting with STAT3 to prevent its dephosphorylation. Phosphorylated STAT3 is then translocated to the nucleus to promote the transcriptional program of DC differentiation. Another case is the NF-κB Interacting LncRNA (*NKILA*) that negatively regulates the NF-κB pathway by interfering with IκB phosphorylation [132] (Figure 3c).

With all the myriad of mechanisms described above it is clear that lncRNAs are important players in many biological processes ranging from cell proliferation, differentiation and development. Moreover, dysregulation of lncRNAs expression has been associated to diverse human diseases including cancer.

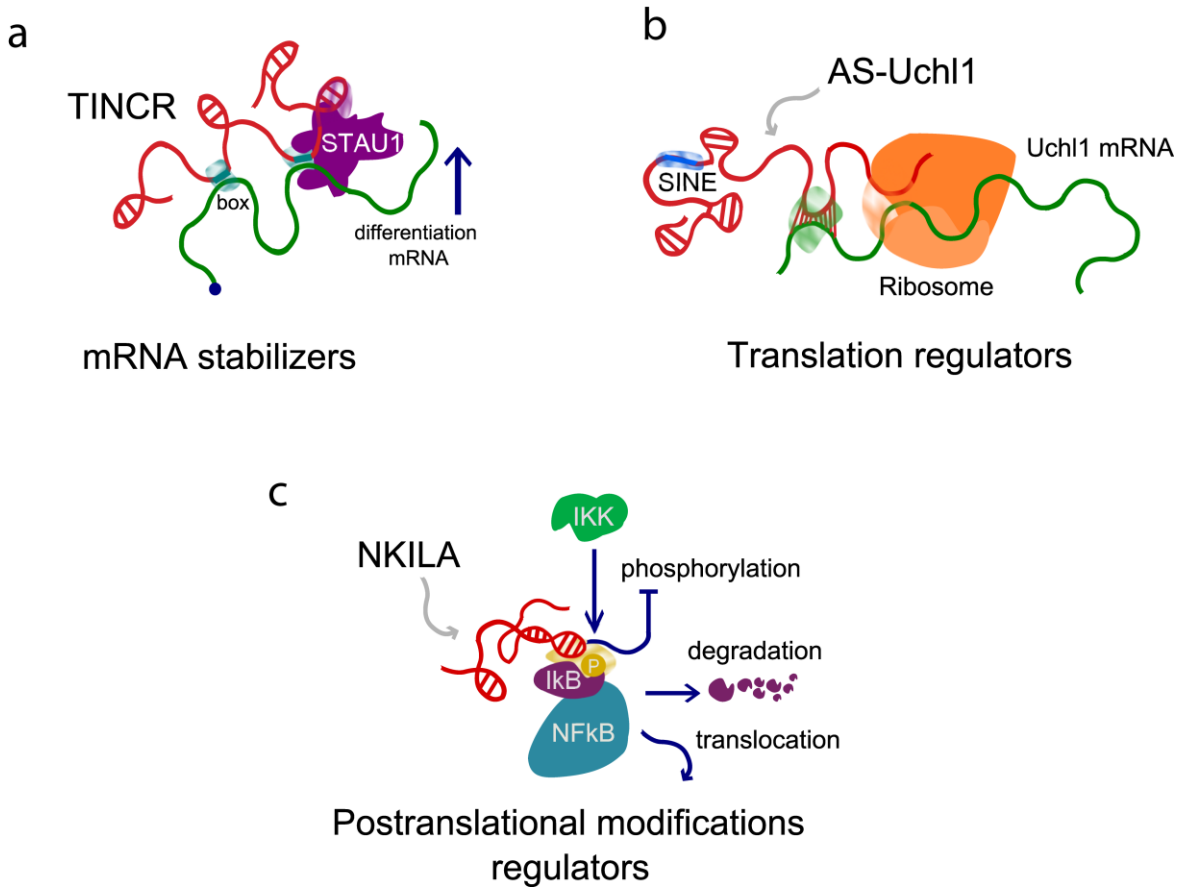

Figure 3. Cytoplasmic lncRNAs acting as mRNA stabilizers, translational regulators and posttranslational modifications regulators. (a) The terminal differentiation-induced ncRNA (*TINCR*) interacts with differentiation mRNAs and binds to the Staufen Double-Stranded RNA Binding Protein 1 (STAU1) protein, promoting the stability of the mRNA targets. (b) The 5' end of the antisense transcript to the ubiquitin carboxy-terminal hydrolase L1 gene (*AS-Uchl1*) recognizes its sense transcript and increases the translation efficiency of Uchl1. *AS-Uchl1* activity depends on the presence an embedded inverted SINEB2 element. (c) The nuclear factor-κB interacting lncRNA (*NKILA*) binds to IκB and masks its phosphorylation inhibiting IκB and preventing the activation of NF-κB.

1.3 LncRNAs and Cancer

Cancer is the name given to a group of heterogeneous diseases characterized by defects in the regulatory circuits that govern normal cell growth [133]. Much work has been done to identify the genetic causes driving cancer, as well as the signaling pathways and protein components altered in the disease.

The advent of new sequencing technologies (DNA and RNA-seq) has generated growing data sets from tumor cohorts (http://cancergenome.nih.gov/) where novel mutations and disease drivers have been discovered. In addition, the differential expression of lncRNAs has also been identified [32], suggesting an important role in cancer progression. A detailed explanation on how the cancer genome and transcriptome helped identified lncRNAs involved in the disease will be covered in the following sections.

Unlike normal cells, the cell proliferation program is constantly turned on in cancer cells. Several lncRNAs have been identified as inducers of this process. One example is the prostate cancer associated transcript 1 (*PCAT-1*), a lncRNA upregulated in prostate cancer that supports cancer cell proliferation by *trans* regulating genes involved in mitosis and cell division. [134]. One of *PCAT-1's* mechanisms of action involves the posttranscriptional repression of the tumor suppressor breast cancer type 2 susceptibility protein (BRCA2) [135]. Besides *PCAT-1*, Prostate-specific transcript 1 (*PCGEM-1*) is another prostate associated lncRNA whose overexpression promotes cell growth [136]. *PCGEM-1* regulates tumor metabolism by acting as a coactivator of the androgen receptor (AR) and c-Myc [137]. In addition to *PCGEM1*, other lncRNAs (*PVT1, CCAT1/CARLo-5*) regulate the oncogene *MYC*. A well-studied example is (plasmacytoma variant translocation 1) *PVT1*; a lncRNA localized to a commonly amplified region (8q24) in cancer that also contains *MYC*. Mouse models demonstrated that *MYC* amplification alone is insufficient to enhance tumor formation. In contrast, *PVT1* and *MYC* co-amplification resulted in an increase of *MYC* protein levels suggesting a *cis* acting mechanism, which could be exploited to target *MYC*-driven cancers [138]. On the other hand *MYC* can also regulate the expression of a group of lncRNAs named *MYClos* [139].

One of the mechanisms for oncogene activation is DNA amplification. Besides *PVT1* there are other examples of oncogenic lncRNA with increased copy number resulting in induced cancer cell proliferation. For example, the focally amplified lncRNA on chromosome 1 gene (*FAL1*) is frequently amplified in epithelial tumors [140]. Experimental data demonstrates that *FAL1* acts as an oncogene by stabilizing the polycomb complex protein B lymphoma Mo-MLV insertion region 1 homolog (BMI1). As a result BMI1 binds to the promoter of the cell-cycle regulator cyclin-dependent kinase inhibitor 1a (*CDKN1A*) repressing its expression. Another amplified lncRNA is the survival-associated mitochondrial melanoma-specific oncogenic ncRNA (*SAMMSON*) [141], whose expression is restricted to human melanoma, specifically to malignant melanomas; moreover *SAMMSON's* inhibition, in patient-derived melanoma xenografts (PDX), reduced tumor proliferation, suggesting its potential use as a biomarker and anti-cancer target. Interestingly *SAMMSON* regulates mitochondrial function by promoting the localization of the protein p32.

Apart from cell proliferation lncRNAs are also implicated in other cancer promoting processes such as angiogenesis and metastasis. The highly up regulated in liver cancer lncRNA (*HULC*), as it name indicates, was identified as the most overexpressed lncRNA in hepatocellular carcinoma [142]. Further studies demonstrate *HULC* overexpression in gastric [143], colon [144], and pancreatic cancer [145]. Its functional characterization revealed a role in cell proliferation and angiogenesis [146]. *HULC* overexpression in HepG2 cells up regulates the tumor angiogenesis-associated factor Sphingosine Kinase 1 (SPHK1).

LncRNAs also participate in cancer invasion and metastasis. *MALAT1* expression is associated with a high risk of metastasis progression in patients with early stage non-small cell lung cancer [147]. Furthermore, *MALAT1* down regulation impairs metastasis in mice [148]. Other metastasis related lncRNAs are *SChLAP1* in prostate cancer [94], *HOTAIR* [66] and *NKILA* [149] in breast cancer.

Other lncRNAs are involved in tumor suppressor pathways such as the p53 pathway. It has been shown that p53 activates the expression of several lncRNAs that contribute to its tumor suppressor response. LincRNA-p21 was the first lncRNA identified to be a p53 direct target induced upon DNA damage. Several publications propose two different

mechanisms (*trans* [150] vs *cis* [151]) to explain the p53-dependent apoptosis and cell cycle arrest mediated by LincRNA-p21. Other p53-regulated lncRNAs include: *PANDAR* [102], *Pr-lncRNA-1/10* [152], *LINC-PINT* [90], and *DINO* [105]. Moreover, genome wide analysis of p53 regulated eRNAs identified *lncRNA LED*, which associates to a strong enhancer within the *CDKN1A* gene and regulates its expression [153]. On the other hand, lncRNAs can indirectly regulate p53, such is the case of the maternally expressed gene (MEG3), which inhibits the expression of mouse double minute 2 homolog (MDM2), a p53 negative regulator [154].

The list of lncRNAs with tumor suppressor function is not exclusively related to p53. An interesting example is the neuroblastoma-associated transcript1 (*NBAT-1*) [155]. Low *NBAT-1* expression is accompanied with highly proliferative and poorly differentiated neuroblastoma tumors. *NBAT-1* low levels are due to its promoter hypermethylation and the presence of a high-risk neuroblastoma-associated SNP in one of its introns.

Other lncRNAs with tumor suppressor function include: *GAS5* [156] which sensitizes cells to apoptosis acting as a competitor for glucocorticoid receptor (GR) binding to DNA; PTEN Pseudogene 1 (*PTENP1*) [157] functions as a decoy for miRNAs targeting the tumor suppressor PTEN, and lncRNA-Low Expression in Tumor (*lncRNA-LET*) which regulates hypoxia signaling by promoting (nuclear factor 90) NF90 degradation [158]. These are only some examples from the long list of lncRNAs deregulated in cancer. Exploring their mechanisms could lead to novel therapies in cancer treatment.

1.4 LncRNAs as readouts: biomarkers and prognosis

It is well recognized that lncRNAs show tissue-specific expression patterns; making them ideal candidates for cancer diagnosis. Indeed the lncRNA *PCA3* (Prostate Cancer Antigen 3) is currently used as a biomarker for prostate cancer, approved by the US Food and Drug Administration (FDA). Because *PCA3* levels are detected in the urine it is considered a non-invasive biomarker that can even outperform the previously used prostate specific antigen (PSA) test [159]. Along the same lines, the lncRNAs *SChLAP*1 has been identified as a biomarker for metastatic progression of prostate cancer [160]. In addition, recent publications show that *HOTAIR* expression can predict responsiveness to cancer therapies

[161], suggesting the potential use of lncRNAs expression profiles in personalized medicine.

Differential expression of lncRNAs has been associated with tumor stage and cancer subtypes enabling its use as prognosis marker. In addition, integration of the expression values of several lncRNAs helps establish signatures to predict patient survival [162]. Despite the multiple studies showing lncRNAs potential use in prognosis, more work needs to be done in order to include them as part of routine clinical-care.

1.5 Genomic Instability, an Enabling Hallmark of Cancer

It is now widely accepted that cancer is a disease of the genome [163], since its origin and progression relies on genome changes/mutations (genetic or epigenetic). A key concept to explain cancer is the model of clonal expansion, which states that cancer development is analogous to Darwinian evolution in the sense that it depends on acquisition of genetic variation, coupled to a process of natural selection [164]. The acquired genetic variation observed in cancer cells compared to its progenitor cell is termed somatic mutation, to distinguish from germline mutation passed on from parents to offspring.

It has been calculated that tumor cells contain from 10^4–10^5 somatic mutations compared to normal cells, including single base changes, structural rearrangements and copy-number changes [165]. However, only few of them confer a growth advantage and are positively selected during cancer progression. These mutations are termed ‘driver’ mutations. The rest of mutations are termed ‘passenger’ because they are present along cancer evolution but do not confer a selective growth advantage [166]. Depending on their functional effect driver mutations can be classified into gain-of-function mutations (neomorphic mutations), which activate cancer-promoting genes called oncogenes, or loss-of-function mutations (hypomorphic mutations), which prevent the proper functioning of the genes called tumor suppressors [165].

An important question is how mutations appear and how they manage to accumulate; genomic instability is indeed part of the answer. Genomic instability enables a cancer cell to continuously modify its genome, promoting the accumulation of mutations. Genomic

instability is a characteristic of all human cancers [167]. It is broadly classified into two types: microsatellite instability (MIN) and chromosomal instability (CIN). MIN is a result of defects in mismatch repair genes (MMR), characterized by modification of the repetitive sequences present in microsatellites. CIN consists on the gain or loss of whole chromosomes or fractions of chromosomes that usually occur during errors in mitosis [168].

Despite its frequent occurrence in all tumors, the underlying mechanisms of CIN are poorly understood. Several models have been proposed to explain the molecular basis of CIN, for example, oncogene-induced replication stress [169] and mitosis defects. The major consequences of CIN are structural variations (SVs) (translocations and inversions) and copy number alterations (CNAs). CNAs include global aneuploidy, deletions (loss of heterozygosity (LOH), and homozygous deletions (HOMDs)), and amplifications. Because of its relation to the results presented in the next section, a detailed explanation of gene amplification follows (Figure 4a).

1.5.1 Gene amplification

Gene amplification refers to an increase in the copy number of a gene derived from redundant replication of a region in the genome. The resulting amplified region is called an amplicon. The first evidence of gene amplification in mammalian cells was the case of the dihydrofolate reductase (*DHFR*) gene in methotrexate (MTX)-resistant cells. MTX treatment resulted in DNA replication arrest and cytotoxicity, except for a group of cells that survived [170]. The surviving cells showed a copy number increase of the *DHFR* gene. Karyotype analysis of these cells showed two abnormalities: extra chromosomal amplifications (double-minute chromosomes, DMs), and intra chromosomal amplifications (homogeneous staining regions, HSRs). DMs are DNA fragments that lack centromeres and telomeres but are able to replicate autonomously [171]. On the other hand, HSRs are segments that form part of chromosomes but they do not show the characteristic discrete bands of trypsin giemsa staining present in metaphase chromosomes [172] (Figure 4b).

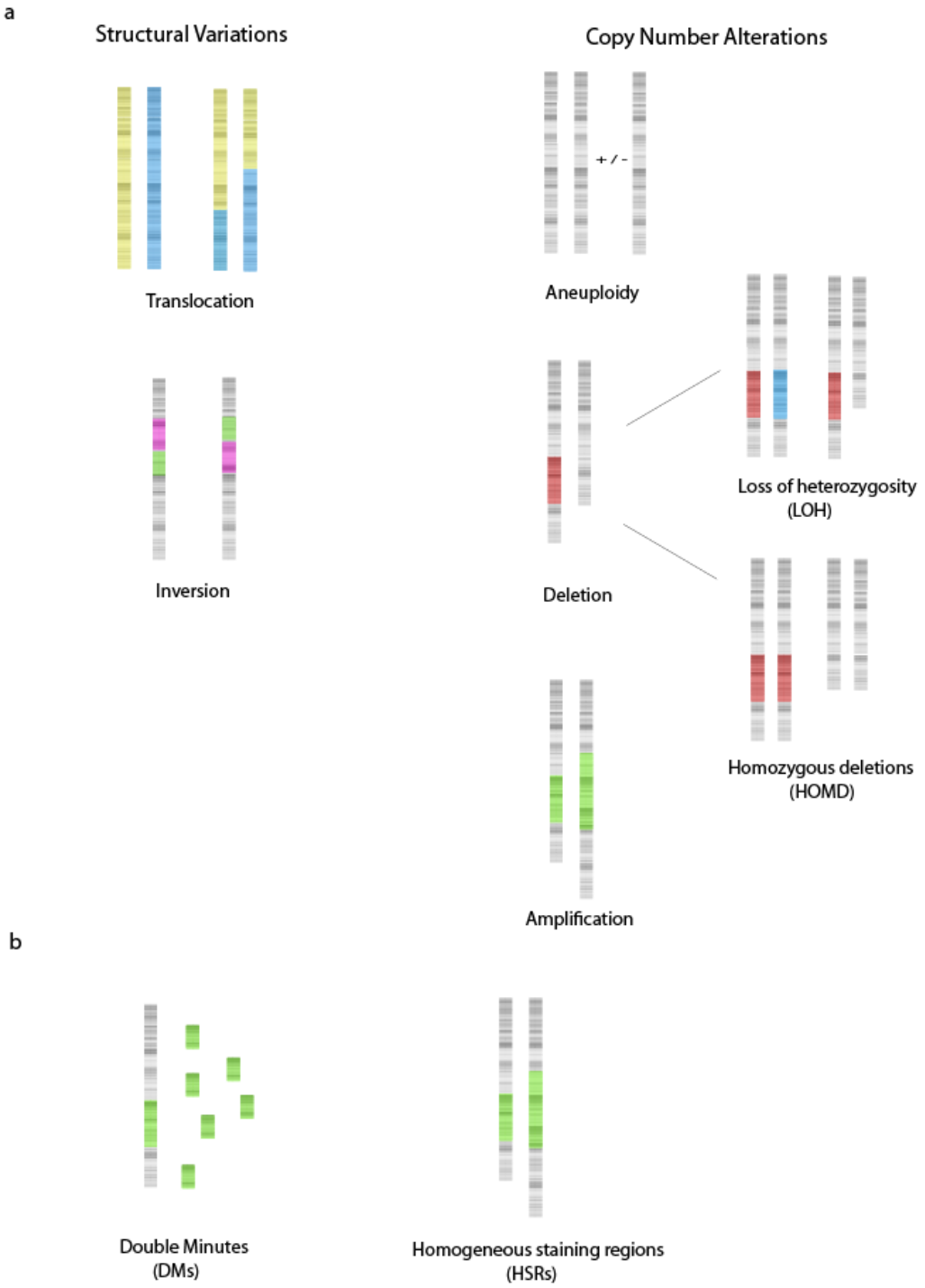

Figure 4. Classes of genomic alterations resulting from chromosomal instability (CIN). (a) Genomic alterations are broadly classified into structural variations, also called balanced rearrangements and copy number alterations. Among the structural variants there are two types: translocations, where fragments from two chromosomes trade places, and inversion, occurring when a genomic region is flipped. Copy number alterations are classified into: global aneuploidy, when a whole chromosome is lost or gained; deletions, when portions of a chromosome are lost. These alterations can be divided into loss of heterozygosity (LOH) and homozygous deletions (HOMD), and amplifications. (b) Amplifications are divided into extrachromosomal double minutes (DMs), which lack centromeres and telomeres and replicate autonomously, and homogeneous staining regions (HSRs).

1.6 The 'omics' Era and Data Integration of Cancer

Back in 1986 Renato Dulbecco stated that 'if we wish to learn more about cancer we must concentrate in the cellular genome' [173]. Few years later the human genome project was launched and a decade after the human genome draft sequence was available. The human genome sequence opened the cancer genomics field enabling the discovery of other genes involved in the disease and recurrent alterations specific to each tumor type. In addition, comprehensive projects such as The Cancer Genome Project [174], The Cancer Genome Atlas (TCGA) [175], and The International Cancer Genome Consortium (ICGC) [176] generated large amounts of high-throughput data from exomes, whole genomes, transcriptomes and epigenomes of cancer cells. All these data revealed the heterogeneity of cancer and raised new challenges for analyzing it.

1.6.1 Copy Number Alterations, Identification of Cancer Driver Genes

Somatic copy number alterations (SCNAs) refer to the copy number change of a chromosomal segment that is of somatic origin meaning that it is identified when germline DNA is compared to the DNA of other cell from the same individual. SCNA should be distinguished from copy number variations (CNVs) which are defined as a DNA sequence with different copy number identified by comparing the germline of two different individuals [177]. Growing interest in studying SCNAs comes from the observation that some oncogenes and tumor suppressors reside in amplifications and deletions respectively [177], [178].

1.6.2 Detection of SCNAs

1.6.2.1 Fluorescence in situ hybridization (FISH)

FISH is a cytogenetic technique that uses fluorescent DNA probes to identify specific sequences within the nucleus, resulting in colored signals of particular chromosomal regions that can be visualized using fluorescence microscopy. It is commonly used for diagnostic of genetic diseases and for guiding targeted therapies in cancer. For example human epidermal growth factor receptor-2 (Her-2) amplification in breast cancer [179] and epidermal growth factor receptor (EGFR) amplification in lung cancer [180]. Because of its simplicity and reliability FISH remains as part of the routine clinical practice [181]. Nevertheless, FISH is only useful for detecting known alterations. For the identification of novel chromosomal alterations the following methods are used (Figure 5a).

1.6.2.2 Comparative genome hybridization (CGH)

Comparative genome hybridization (CGH) consists of contrasting the DNA content of two genomes. A test and a reference genome are differentially labeled and used as probes that cohybridize competitively either to metaphase chromosomes [182] or to DNA microarrays (aCGH) [183]. The hybridization intensity ratio (test/reference) is then measured for each DNA fragment; based on these measurements the copy number is calculated. In contrast to FISH that identifies alterations of specific regions, with CGH all the genome is examined.

When CGH was developed metaphase chromosome preparations were used for hybridization, rendering a detection resolution of 5 to 10 Mb. Further application of DNA spotted arrays for hybridization improved the resolution up to 30 Kb, by reducing the size of the sequence target (BACs > cDNA > oligonucleotide microarrays) and by increasing the density of coverage throughout the genome [184]. Furthermore targeted arrays have been developed to interrogate in detail the copy number of regions with known clinical significance. Some drawbacks in aCGH use are 1) the fact that it is limited to detect copy number changes; chromosomal rearrangement like translocations and inversions are missed by this method, and 2) it does not reveal the location of the SCNA (DM or HSR, for the case of the amplifications) Figure 5b.

1.6.2.3 Single nucleotide polymorphism (SNP) arrays

SNP arrays are widely used in genetic association studies. However, copy number alterations can also be inferred from them. Similar to aCGH, SNP arrays are based on DNA hybridization between complementary sequences. The difference relies on the sequence-specific oligonucleotides, which are able to discriminate between the different SNP alleles. The hybridization of DNA sample depends on the genotype and is measured based on the fluorescent signal.

Two types of data are generated from SNP arrays: 1) intensity data, used to infer the copy number of the genomic region containing the SNP and 2) genotype data of the different SNPs alleles [185]. The combination of these two data types improves the identification of copy number and most importantly enables the detection of loss-of-heterozygosity (LOH) events. In addition to the genotype data, SNP arrays have a high-density coverage what increases the detection resolution to < 0.7 kb [186] (Figure 5c).

1.6.2.4 Next generation sequencing (NGS)

The advance of NGS technologies opened up opportunities to use exome and whole genome sequencing approaches to identify CNAs. Advantage of NGS include high coverage and single base resolution, leading to more accurate estimation of copy number, precise detection of the boundaries of the copy number altered segments (breakpoints) and identification of novel structural rearrangements (inversions, translocations) [187] (Figure 5d).

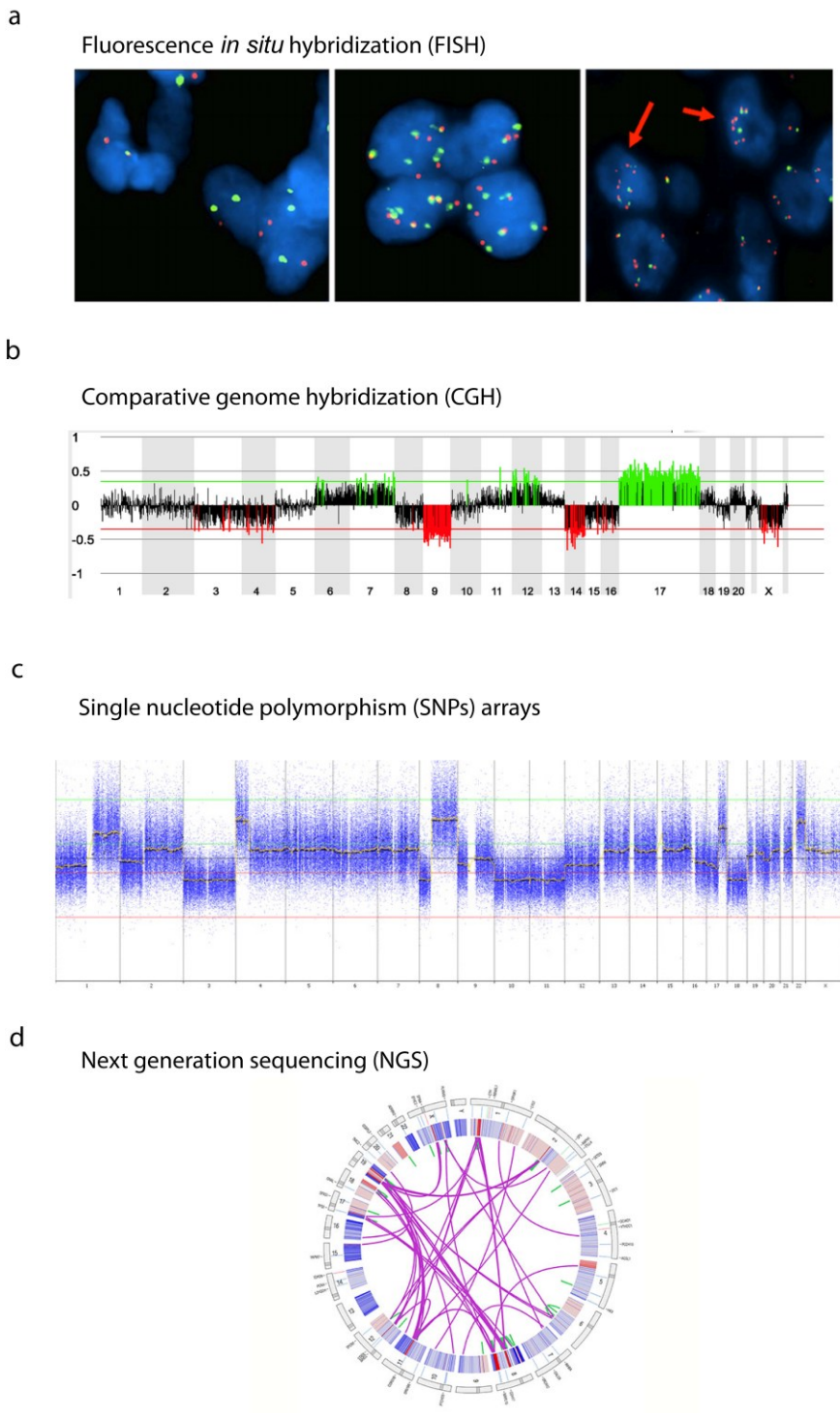

Figure 5. The different methods used for copy number alteration (CNA) detection. (a) Example of FISH analysis of chromosome 7 (green) and EGFR CNA (red), left disomy of chr7, middle polysomy of chr7 and right EGFR copy number alterations. (b) Array CGH copy number profile showing numerical imbalances (chr9). The X-axis represents the chromosomes and the Y-axis represents the normalized log2 Cy3/Cy5 fluorescence intensity. (c) Copy number alteration plot obtained by SNP array analysis showing the losses of chromosome 1p, 2p, 3, 8p, 10, 11, 18 and the gains of chromosome 4p, 8q, 17 and 22. (d) Circus plot showing the copy number alterations obtained by whole genome sequencing (WGS) amplified (red) and deleted (blue) regions are shown. In addition, the detected translocations are represented as purple lines.

1.6.3 Identification of Driver Genes

Current technologies enabled the characterization of SCNAs in thousands of cancer genomes; showing that SCNAs affect a large fraction of the genome, targeting multiple genes simultaneously. Furthermore, integration of SCNAs with gene expression has led to the discovery of novel cancer driver genes [188], demonstrating the potential use of this approach. However, the SCNA analysis to discover cancer drivers brings two challenges [178], [189]. The first challenge is to distinguish the SCNAs acting as driver events that contribute to cancer development from those that are just passengers and have been acquired during cancer evolution. Analysis of large collections of tumor samples helped identify SCNAs acting as drivers [177]; based on the idea that driver events confer a growth advantage and are under positive selection, driver SCNAs appear at higher frequencies compared to passenger events. However, it has been shown that some recurrent SCNAs can also be passenger events, arising from other mechanism in the absence of positive selection. For example, deletions in common fragile sites due to increased rates of chromosome breakage present in these regions [190]. The second challenge is to identify which gene inside the SCNA is a cancer driver. Considering that a SCNA typically affects multiple genes, computational algorithms have been developed [189] [191] to delineate peak regions likely to harbor driver cancer genes, for example the algorithm called Genomic Identification of Significant Targets in Cancer (GISTIC).

A comprehensive analysis of SCNAs in human cancers using these algorithms identifies several of their features. Based on their size, SCNAs can be classified into arm-level or focal alterations. It has been calculated that in a cancer sample, 25% of the genome is affected by arm-level SCNAs and 10% by focal SCNA [177]. Recent analyses estimate a median of 24 focal SCNAs (11 amplifications and 12 deletions) and 5 arm-level alterations (3 amplifications and 5 deletions) per cancer sample [178]. Focal SCNAs have a median length of 1.8 Mb ranging from 0.5 kb to 85 Mb, and interestingly, there is an inverse correlation between their size and their frequency. In contrast to arm-level alterations, that are exclusively present in low amplitude (one copy gain or loss), focal alterations have higher amplitude (multiple copies). Using the GISTIC algorithm, focal alterations are delineated into peak regions showing a median of 4 or 3 genes per amplified or deleted

peaks, respectively. Classical oncogenes like *CCND1*, *EGFR, MYC*, *ERBB2*, *CCNE1*, *MCL1* and *MDM2* are present inside recurrently amplified peaks. On the other hand, tumor suppressor genes including *PTEN, CDKN2A/B*, *STK11, PARK2, and QKI* are identified as frequent targets of homozygous deletions. The most recent pan-cancer analysis identified 140 SCNA peak regions. However, only 35 of them contained a known cancer driver [178]. The rest of the peaks remain unexplored what opens up an opportunity to identify novel oncogenes or tumor suppressors.

1.7 Lung Cancer

Lung cancer is still the most common cause of cancer-related deaths worldwide accounting for 1.6 million deaths in 2012 [192]. Moreover, the number of lung cancer deaths is expected to grow up to 3 million by 2035 (from 1.1 million to 2.2 in men and 0.5 million to 0.9 million in women) [193].

1.7.1 Lung cancer classification

Lung cancer is classified into two major types: small cell lung cancer (SCLC), and non-small cell lung cancer (NSCLC) [194] (Figure 6a).

SCLC represents around 10-15% of all the diagnosed lung tumors [195] and is highly associated with smoking. This class of tumors arises from the neuroendocrine cells, which have a round and small size with minimal cytoplasm, granular nuclear chromatin and lack nucleoli (Figure 6b). It is the most rapidly growing type of lung cancer, and tends to spread quickly. Usually at time of diagnosis the patients already present metastasis. It has the worst prognosis with a 5-year survival rate of only 5% [196].

NSCLC is the most common type of lung cancer comprising 80-85% of the cases. Abnormal growth of the lung epithelial cells gives rise to this type of tumor. Based on the type of lung cells where the tumor arises, NSCLC is divided into three subtypes: adenocarcinoma (ADC), squamous cell carcinoma (SCC), and large-cell carcinoma (LCC). ADC accounts for 40% of all lung cancer, and is the most common type of lung cancer present in non-smokers. It is located in the mucus making gland cells of the outer parts of the lungs (periphery). Compared to other types of lung cancer ADC tends to be slow growing (Figure

6c). SCC also called epidermoid carcinoma accounts for 30% of all lung cancer [197], it arises from bronchial epithelial cells located inside of airways in the lung. Particular features of SCC cells include keratinization and presence of intercellular bridges between adjacent cells (Figure 6d). LCC accounts for 10% of lung cancers, it often starts in the central part of the lung. Large and abnormal cells characterize LCC. In contrast to ADC, LCC grows quickly and tends to spread to nearby lymph nodes. It is usually discovered at later stages making it harder to treat (Figure 6e).

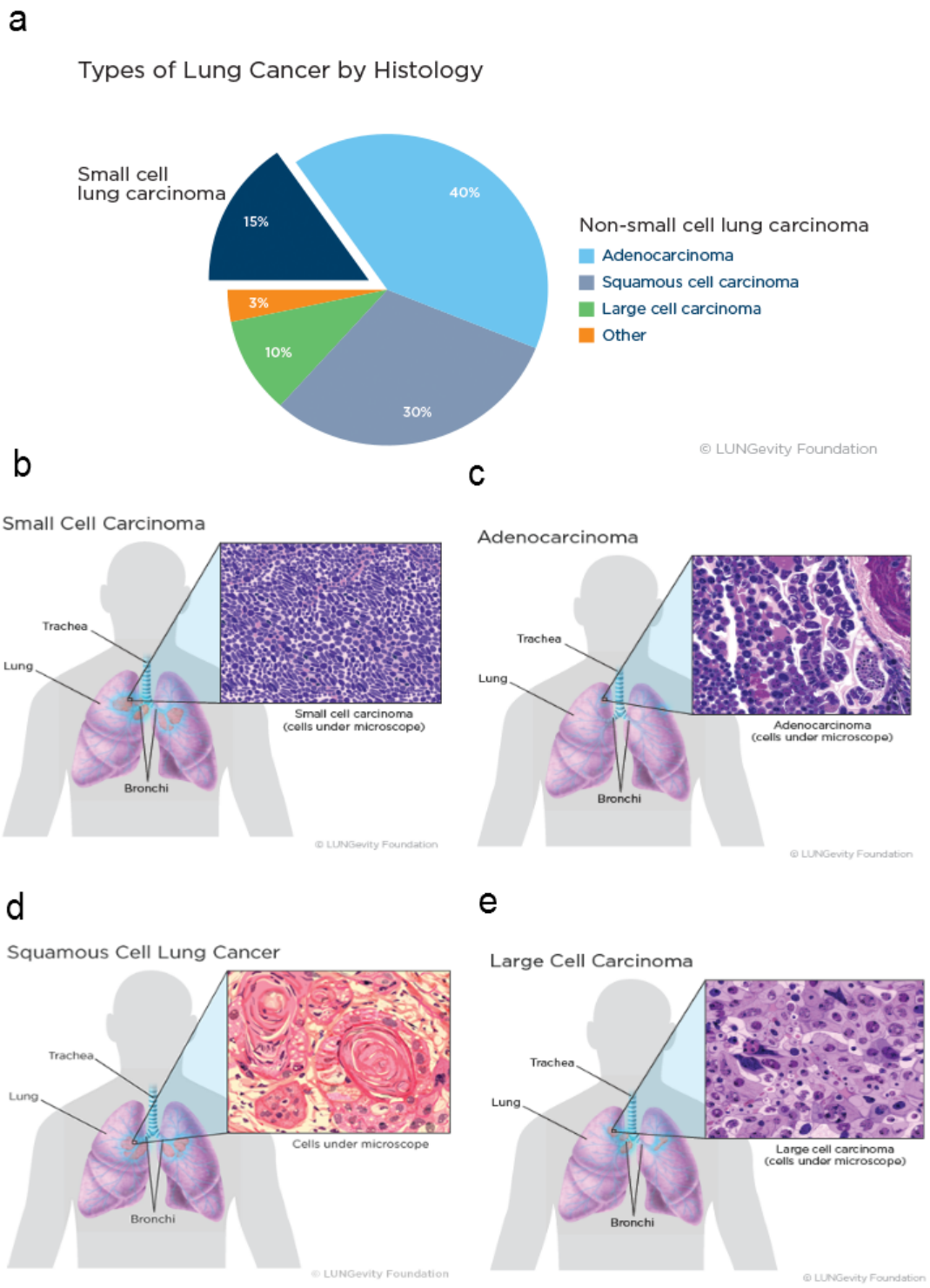

Figure 6. Lung cancer classification by histology. (a) Different types of lung cancer classified histologically based on cells that form the tumor. (b) Small cell carcinoma (SCLC); (c) adenocarcinoma (ADC); (d) squamous cell carcinoma (SCC); (e) large cell carcinoma (LCC). Figures taken from: https://www.lungevity.org/.

1.7.2 Lung Cancer Genomics

Large-scale projects to comprehensively characterize lung cancer have identified the most recurrent genomic alterations present in this tumor type. The identification of these alterations has opened new opportunities for molecularly targeted therapies changing patient treatment, for example the gefitinib therapy for patients with activating mutation in the epidermal growth factor receptor (*EGFR*) [198] and the crizotinib treatment for the gene fusion comprising EMAP-like protein 4 *(EML4)* and anaplastic lymphoma kinase *(ALK)*. Many other recurrent alterations in lung cancer such as mutations in *AKT1*, *ERBB2*, and *PIK3CA* emerge as promising therapeutic targets. In the following section the most common genomic alterations identified in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are summarized.

Analysis of 230 LUAD tumors from the TCGA revealed eighteen genes significantly mutated [199]. The most commonly mutated gene was *TP53* (46%) followed by the oncogene *KRAS* (33%) whose mutations result in abnormal protein activation. *EGFR* was mutated in 14% of the cases while *BRAF* (10%), *PIK3CA* (7%). Next in the list were the tumor suppressor genes *STK11* (17%), *KEAP1* (17%), *NF1* (11%) and *RB1* (4%). In addition mutations in RNA splicing genes such as *RBM10* (8%) and *U2AF1* (3%) and in the chromatin modifying genes *SETD2* (9%), *ARID1A* (7%), *SMARCA4* (6%), were also identified. Significant copy number alterations comprised amplifications of *EGFR*, *CCNE1*, *KRAS*, *NKX2-1*, *MET*, *MDM2*, *TERT* while the most significant deletion mapped to *CDKN2.*

Aberrant RNA transcripts were also present in lung adenocarcinoma. The identified gene fusions involved the genes *ALK*, *ROS1* and *RET2.* Mutations in *U2AF1* were associated with the alternative splicing of the *CTNNB1* mRNA. Key pathways altered in lung adenocarcinoma include the *RTK/RAS/RAF* (76%), p53 (63%) and *PI3K/AKT/mTOR* (25%) pathways.

Additional characterization of LUSC using a set of 178 tumor samples revealed the most frequent alterations found in this cancer subtype [200]. The somatic mutation rate calculated from LUSC tumors was of 8.1 mutations per Mb, higher than the mutation rate

observed for other tumor types (BRCA 1 per Mb, OV 2.1 per Mb, and COAD 3.2 per Mb). *TP53* was the most commonly mutated gene (81%), followed by *MLL2* (20%).

Significant deletions include the genes *FOXP1*, *NF1*, and *PTEN*. Additional studies report frequent deletion of chromosome 3p in early stages of the disease. This deletion maps to *RASSF1A, FUS1, VHL*, and *FHIT* [201]. The identified amplifications harbor *BCL2L1*, *CDK6*, *NFE2L2/NRF2*, *MDM2* and *MYC.* Interestingly amplification of 3q is exclusive of LUSC, moreover high-grade lesions are associated with chromosome 3q amplification and invasive cancer. Among the genes in 3q are *SOX2*, *TP63*, *PIK3CA* and *EPHB3*.

Oxidative stress response and cell differentiation were the two pathways where most of the mutated genes are involved. The oxidative stress response pathway is altered by the mutations and copy number alteration of *NFE2L2/NRF2 (*19%*)*, *KEAP1 (*12%*)* and *CUL3 (*7%*)*. On the other hand overexpression and amplification of *SOX2 (*21%*)* and *TP63 (*16%*)* results in deregulation of squamous cell differentiation.

Many of the driver alterations present in LUAD for example the activating mutations of *EGFR* and the *ALK* fusions are rarely found in LUSC, because of this current targeted therapy to treat this type of tumors are often ineffective. The need to identify novel targets led to the discovery of potentially druggable genes altered in the disease. Many of these genes belong to the tyrosine kinase families (*ERBBs, FGFRs* and *JAKs*). These findings open up new avenues for treatment.

1.8 The NF-κB signaling pathway

The NF-κB signaling pathway is involved in the regulation of diverse processes including adaptive and innate immunity, inflammation, proliferation and apoptosis. Moreover the NF-κB pathway has provided a link between inflammation and cancer. Central players of this pathway include the NF-κB transcription factors and its inhibitors (IκBs).

A variety of signals leading to NF-κB activation are transduced through two main pathways referred to as the 'canonical' or the 'alternative' pathway, which differ in the input signals, the involved proteins and the gene activation outcomes.

The canonical pathway is turned on by proinflammatory cytokines (tumor necrosis factor alpha TNFα, interleukin 1 beta IL-1β), pathogen associated molecules (lipopolysaccharide LPS) and antigen receptors. The signaling cascade results in the activation of IκB kinase complex (IKK) formed by IKK1/IKKα, IKK2/IKKβ and NEMO/IKKγ. IKK is considered the core element of the pathway because almost all the upstream signals converge on it.

Once activated, IKK phosphorylates IκB. Phosphorylated IκB is then recognized by the Skp, Cullin, F-box containing complex (SCF), a E3 ubiquitin ligase that catalyzes the lysine 48 ubiquitination of IκB. This modification targets IκB for subsequent proteosomal degradation. The absence of IκB releases NF-κB, which translocates into the nucleus, binds to the κB sites and promotes the transcription of its target genes. The target genes induced by this pathway include proinflammatory cytokines, chemokines, adhesion molecules and others.

On the other hand, the alternative pathway is activated by the B Cell-activating factor (BAFFR), CD40 ligand, lymphotoxin-β and others. This pathway uses other members of the NF-κB and IκB family (e.g. RelB and p100), and it relies exclusively on IKKα, whose activation is mediated by NF-κB inducing kinase (NIK). Once activated, IKKα phosphorylates p100 which is processed into p52. This proteolysis results in the formation of the RelB/p52 dimers that translocate into the nucleus and regulates the expression of genes involved in adaptive immunity and B cell maturation (Figure 7)

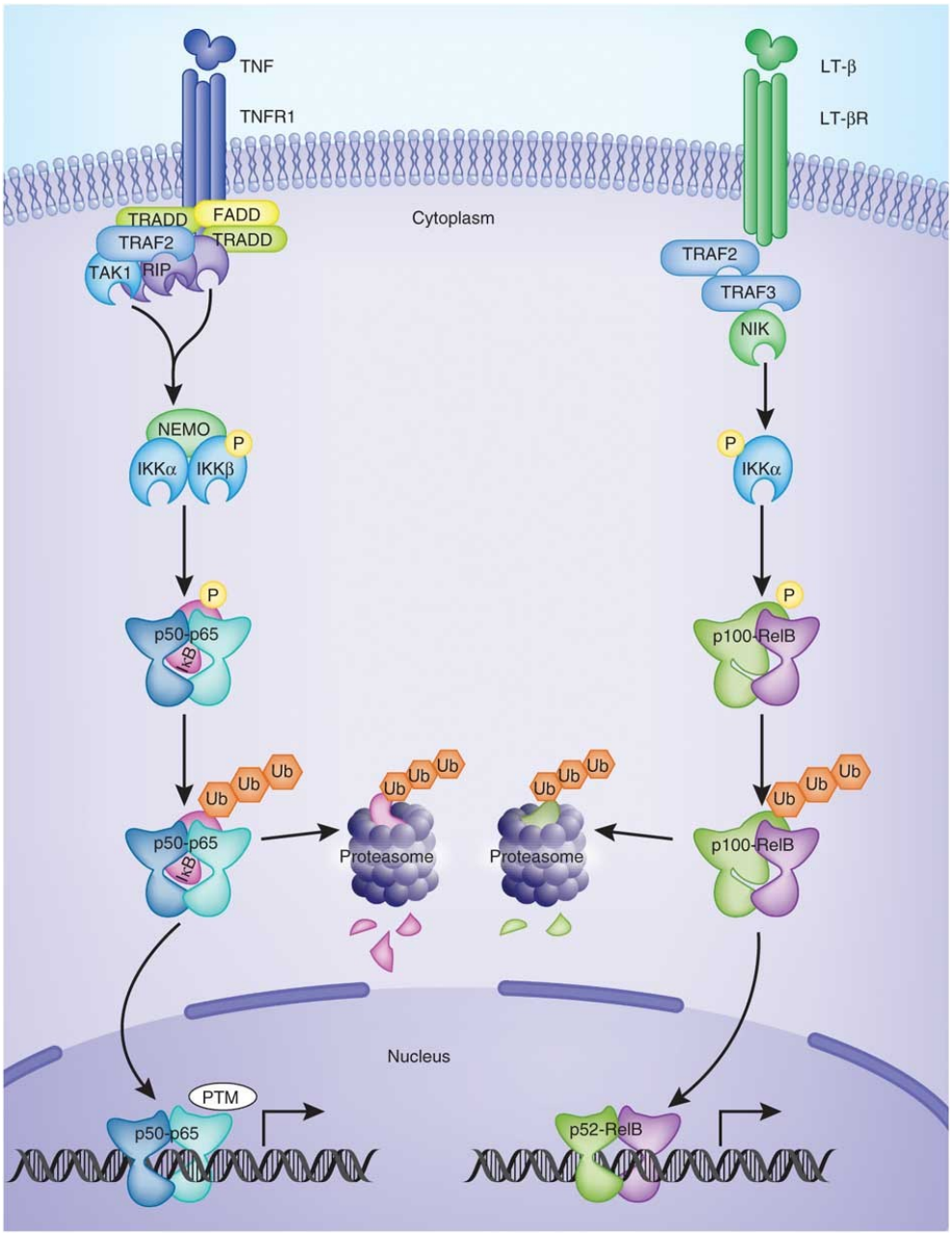

Figure 7. Canonical and alternative pathway of NF-κB. Activation of NF-κB can be achieved by two different pathways. The canonical pathway consists in the activation by diverse stimuli such as TNFα, IL-1β, LPS via several receptors (TNFR1, TLR, RANK, CD30/40). For example, TNF binding stimulates TNFR1 leading to the binding of TRADD, followed by the recruitment of TRAF2. TRAF2 associates with RIP; RIP1 via Lys63-linked polyubiquitin chains recruits TAK1 kinase and IKK complex, resulting in the phosphorylation and activation of IKK. Activated IKK phosphorylates IκB in two serine residues what promotes IκB ubiquitination and consequently its proteosomal degradation. NF-κB is then released and translocate into the nucleus activating the transcription of its target genes. The alternative or non-canonical pathway is activated by CD40L, BAFF and lymphotoxin-β (LT-β). For example when LT-β binds it receptor (LT-βR) the NIK kinase is recruited by TRAF2/3. NIK then phosphorylates IκBα; this pathway relies exclusively on IκBα. Once activated by NIK, IκBα phosphorylates p100, resulting in p52 production. Since p52 lacks the ankyrin repeats (A) it is able to dimerize with RelB and translocates into the nucleus. Figure taken from [2].

1.8.1 The NF-κB family of transcription factors

NF-κB was discovered more than 30 years ago as a nuclear factor that binds to the enhancer element of the immunoglobulin kappa light-chain of activated B cells, from where it got its name [202]. Later studies revealed that the mammalian family of NF-κB transcription factors consist of five members: p105/p50 (NF-κB1), p100/p52 (NF-κB2), p65 (RelA), c-Rel and RelB. All these proteins share a 300 aminoacid long Rel homology (RH) domain which is essential for their dimerization, interaction with the inhibitory proteins (IκB), DNA binding and nuclear translocation. NF-κB regulates the expression of numerous target genes controlling several cellular processes such as cell growth, immune response, inflammation, tissue invasiveness and apoptosis.

NF-κB family members can be divided in two classes based on the C-terminal sequence of their RH domain [203]. The first class is composed of p105/p50 and p100/p52. Upon maturation, p105 and p100 are cleaved into p50 and p52 respectively and become active. The second class comprises p65, c-Rel and RelB. These proteins have a transactivation domain (TAD), which positively regulates the transcription of their target genes (Figure 8).

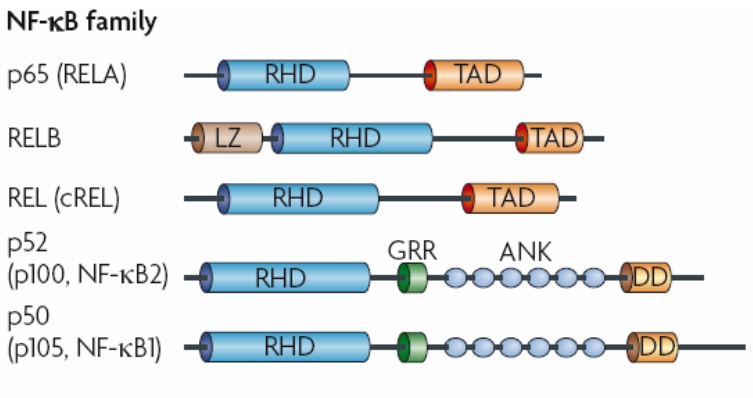

Figure 8. NF-κB transcription factor family. NF-κB transcription factor family is comprised by five members: RelA/p65, RelB/p50, c-Rel, p105/p50 and p100/p52. All these proteins contain a Rel homology domain (RHD) which mediates dimerization, nuclear translocation, DNA binding, and interaction with IκBs. Additionally RelA, RelB, c-Rel have a transactivation domain (TAD) that mediates the initiation of transcription of the NF-κB target sites. The p100 and p105 precursor proteins contain seven ankyrin repeats (A); these two proteins are then processed to p50 and p52 respectively, rendering them active for nuclear translocation. Figure from [1]

To bind DNA, NF-κBs associate in dimers. The combinatorial diversity of all the family members results in the formation of distinct NF-κBs, which induce or inhibit specific set of genes. The p50/p65 heterodimer is the most abundant and regulates the expression of more genes compared to the other ones. Commonly the term NF-κB refers to this dimer. NF-κB binds to specific sequences (called κB sites) in promoters and enhancer regions. These consensus κB sites consist of 9-10 nucleotides with the following sequence: 5'-GGGRNNYYCC-3' (R-purine; Y-pyrimidine, and N-nucleotide) [204]. Because of its potential to control broad gene expression programs, NF-κB activity is tightly regulated at multiple levels through upstream signaling molecules to downstream feedback loops

1.8.2 Regulation of the NF-κB pathway

1.8.2.1 Inhibitors of NF-κB (IκBs)

The main regulatory mechanism of NF-κB activity is mediated by a family of proteins known as inhibitors of NF-κB. IκB protein family is classified in three groups [205]: The typical IκBs (IκBα, IκBβ, IκBε) sequester NF-κB in the cytoplasm; the precursor proteins (p100 and p105) inhibit NF-κB prior to get processed; and the atypical IκB proteins (IκBζ, Bcl-3 and IκBNS) regulate NF-κB function in the nucleus (Figure 9).

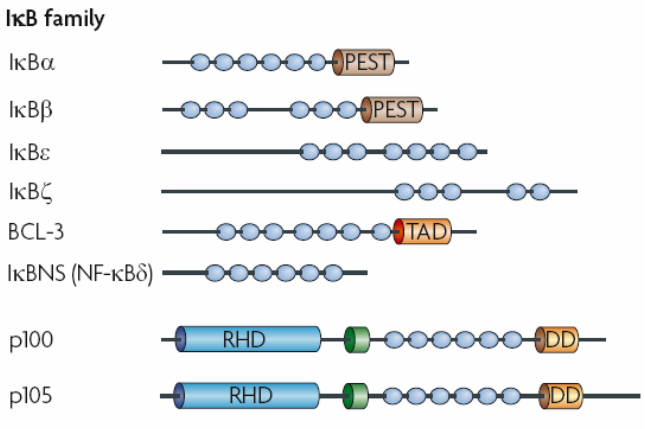

Figure 9. Inhibitors of NF-κB (IκBs). Eight IκBs have been described (IκBα, IκBβ, IκBε, IκBζ, BCL-3, IκBNS, p100, p105). These proteins contain 6 or 7 ankyrin repeats (A), which bind to the nuclear localization sequence of NF-κB, retaining it in the cytoplasm. In IκBα and IκBβ there is a region rich with proline, glutamate, serine, and threonine residues named the PEST domain which is involved in protein degradation. Other domains present in these proteins include the transactivation domain (TAD) and the death domain (DD). Figure from [1].

The established model for IκB function states that IκB sequesters NF-κB in the cytoplasm directly binding to it and masking its nuclear localization sequence (NLS). Upon activation of the pathway IκBα is degraded releasing NF-κB, which translocates into the nucleus.

1.8.2.2 Activation of the IKK complex, a crucial step in the cascade

IKK is a core element of the pathway, where a great variety of stimuli converge, resulting in the activation of the complex. The IKK complex consists of two catalytically active kinases IKK1/IKKα, IKK2/IKKβ and a regulatory subunit NEMO/IKKγ (Figure 10).

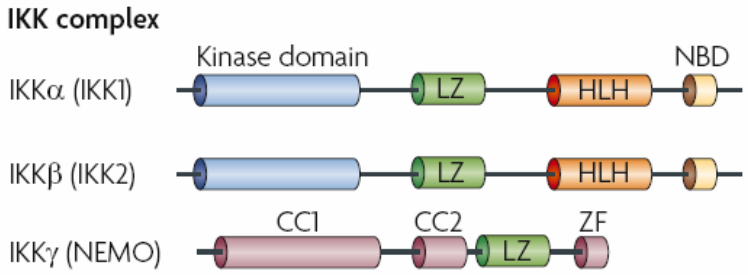

Figure 10. IKK complex. The IKK complex is formed by three subunits: IKKα, IKKβ and IKKγ; the first two (α, β) contain a kinase domain responsible for phosphorylating IκB, a helix-loop-helix (HLH), and a leucine zipper (LZ) for dimerization, they both bind via the nemo binding domain (NBD) to the third subunit: IKKγ. IKKγ/NEMO acts as a regulatory subunit, which contains a coil–coil (CC) and a zinc finger ZF domain where K63-linked polyubiquitin chains attach. Figure taken from [1].

IKK becomes activated by the phosphorylation of some of its serine residues; a reaction catalyzed by upstream kinases like NIK, mitogen-activated protein Kinase/ERK kinase kinase 3 (MEK3), and the TGF-β activating kinase 1 (TAK1). Moreover the IKK complex is the site of crosstalk with other signaling pathways. For example, some publications show that the serine-threonine kinase Akt can phosphorylate IKK in response to TNFα [206]. Additionally to phosphorylation it has been shown that ubiquitination also plays an important role in IKK activation [207]. Further explanation of the role of ubiquitination in the NF-κB pathway will be presented in the next sections.

1.8.2.3 Terminating NF-κB response, negative regulators of the pathway

Termination of NF-κB transcriptional activity is mediated by the upregulation of inhibitor proteins such as A20 and cylindromatosis (CYLD), which are induced by NF-κB itself [208], [209], establishing an autoregulatory negative feedback loop that controls the overactivation of the pathway.

Inhibitors such as p100 and p105 are also induced by NF-κB, and the synthesis of these proteins alters the pool of IκB dimers modulating NF-κB activity. Moreover, among the induced targets of NF-κB is IκBα. After protein synthesis, IκBα enters the nucleus and removes NF-κB from the DNA, and shuttles it to the cytoplasm [210]. If the activating signals are still present, IKK phosphorylates IκBs, which results in repeated cycles of reactivation and an oscillatory NF-κB activity [211].

Additional mechanisms are required to turn off the pathway completely. Among all, deubiquitinases (DUBs) play a central role in this process. All these examples demonstrate the high level of regulation of the NF-κB pathway; therefore it is reasonable to think that its dysregulation results in a myriad of diseases including cancer.

1.9 NF-κB in cancer

The role of NF-κB in cancer was suggested since its discovery as a homolog of the protein v-Rel, a transforming gene coming from an avian retrovirus highly oncogenic. [212]. It is now clear that oncogenic mutations of p65 and other NF-κB proteins are present in cancer, however they are rare and occur exclusively in lymphoid malignancies. Examples include the chromosomal rearrangements or deletions affecting the NF-κB2 locus, observed in lymphoid malignancies [213]. This rearrangement results in a truncated version of p100 that lacks the ankyrin repeats generating a constitutively active p52 protein. Another examples is the translocation (14;19)(q32;q13), observed in B-cell leukemias [214]. The result of this translocation is the overexpression of BCL-3, which acts as co-activator of p50 and p52 dimers, leading to an increase NF-κB activity [215]. On the other hand, in solid tumors NF-κB is constitutively active due to either the dysregulation of the mechanism controlling its activity (e.g. defective IκBα activity, constitutive IKK

activity) or to the persistent proinflammatory stimuli present in the tumor microenvironment, constantly inducing the pathway. Growing evidences demonstrate that NF-κB plays a central role connecting inflammation and cancer [216].

Alterations of several component of NF-κB signaling have been identified in epithelial cancer. For example somatic mutations and translocations of IκBs (IKBKA, IKBKE and IKBKB) are present in breast and prostate tumors [217], [218]. In addition TNFR-associated factor 2 (TRAF2*)*, a receptor involved in activation of the NF-κB signaling cascade is amplified in human epithelial cells [219]. Moreover, amplification of TRAF6 in lung cancer cells with activated RAS is associated with increased NF-κB activity [220].

Many of the genes regulated by NF-κB are broadly implicated in the hallmarks of cancer. For example NF-κB promotes cell proliferation by regulating the expression of cell cycle genes such as cyclin D1 (CCND1) [221], CDK2 kinase [222], and c-myc [223]. NF-κB is also involved in migration, invasion and angiogenesis by regulating the expression of matrix metalloproteinases (MMPs), adhesion molecules (ICAM-1, VCAM-1), chemokine receptors (CXCR4) and vascular endothelial growth factor [224]. The anti-apoptotic effect of NF-κB is mediated by the upregulation of target genes such as the inhibitors of apoptosis (IAPs): c-IAP1, c-IAP2, and XIAP, which directly inhibit caspase-8 [225], [226]. Moreover NF-κB regulates the expression of the FLICE-like inhibitory protein (FLIP), which competes with caspase-8 interfering with the dead receptor apoptotic pathway [227]. In contrast, a pro-apoptotic role has also been described for NF-κB which induces the death receptors: DR4, DR5 and Fas [228]. Moreover, activation of NF-κB using UV, daunorubicin or doxorubicin, results in the repression of anti-apoptotic genes [229], suggesting that the pro or anti apoptotic function of NF-κB could be determined by the nature of the activating stimulus.

1.10 TNFα, an activator of the NF-κB pathway

A potent activator of the NF-κB pathway is the tumor necrosis factor (TNFα), a multifunctional cytokine involved in inflammation, cell survival and apoptosis. Initial observations of the effect of TNFα showed that it could suppress tumor proliferation in animal models and induce cytotoxicity of cancer cells [230], [231].

TNFα is initially synthesized as transmembrane protein of 26 kDa (pro-TNFα), which is processed by the TNFα-converting-enzyme (TACE) [232] generating the soluble form (sTNFα) a 17-kDA protein that forms homotrimers when in solution. TNFα acts via two distinct receptors, Type 1 and 2 (TNFR1/TNFRSF1A and TNFR2/TNFRSF1B) [233]; the major difference between both is the presence of the death domain (DD) exclusively in TNFR1, which has a huge implication on TNFα biological activities. TNFR1 has been widely studied because of its role as a dual receptor. In addition to the induction of apoptosis, TNFR1 also transduces cell survival signals. How these opposite signals are regulated still remains obscure.

The pathway mediated by TNFR1 starts with the binding of TNFα, which induces the trimerization of the receptor and the recruitment of TNFR1 associated death domain protein (TRADD). TRADD serves as a docking platform for key downstream adaptor proteins, which activate distinct signaling pathways. These downstream adaptors include the receptor interacting protein (RIP), TRAF2, and Fas-associated death domain (FADD), which further recruits other proteins to activate the NF-κB pathway, the mitogen-activated protein kinases (MAPKs), and cell death respectively.

The dual role of TNFα becomes evident when analyzing the signaling pathways that it activates. Apoptosis signaling is mediated by FADD that binds pro-caspase-8 and activates the caspases cascade leading to apoptosis. In contrast, signaling mediated by TRAF2 inhibits apoptosis by inducing MAPKs (extracellular signal-regulated protein kinases (ERK), cJun N-terminal kinase (JNK) and p38), which results in the activation of the transcription factors cFos/cJun [234]. Additionally, the major pathway activated by TNFR1 is the one leading to NF-κB activation resulting in the induction of pro-survival genes. NF-κB activation is dependent on the adaptor proteins RIP1 and TRAF2, which are involved in the activation of the core element of the pathway, the IKK complex. TRAF2 recruits IKK to TNFR-1, while RIP1 mediates IKK activation together with the kinase TAK1 [235]. Upon activation IKK phosphorylates IκB, subsequently triggering its ubiquitination and proteosomal degradation. IκB degradation results in NF-κB release and translocation to the nucleus where it activates transcription of its targets.

1.11 Ubiquitination in the NF-κB signaling pathway

Ubiquitin is a 76 aminoacid long protein that is attached to target proteins, either promoting their proteosomal degradation, regulating their trafficking, or altering their function. Ubiquitination plays a central role in the NF-κB signaling pathway by mediating critical steps in the signaling cascade, for example degradation of IκB, processing of NF-κB precursors (p100 and p105), and activation of IKK.

Ubiquitination is a reversible covalent modification catalyzed by a three step cascade (activation, conjugation, ligation) involving three different types enzymes: E1, E2 and E3. The inverse process is catalyzed by proteases named deubiquitylation enzymes (DUBs) [236], which include more than 100 DUBs classified in five families: ubiquitin carboxyl-terminal hydrolase (UCH), ubiquitin specific protease (USP), ovarian tumor domain (OTU), Machado-Joseph disease protease (MJD) and JAMM.

Several rounds of ubiquitination with the same protein as substrate results in the formation of polyubiquitin chains. Two types of polyubiquitin chains can be generated: linear chains where the carboxyl group of one ubiquitin is bound to the amino group of another [237], and non-linear chains which use internal lysines as the substrate for the attachment of other ubiquitin monomer (Figure 11).

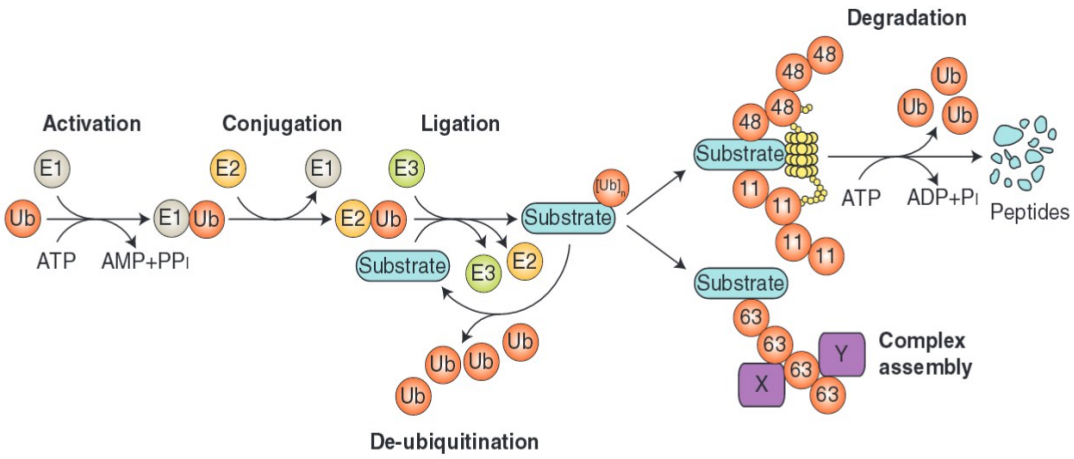

Figure 11. The ubiquitination system. The ubiquitination reaction occurs in three sequential steps: 1) activation, ubiquitin is activated by an ubiquitin-activating enzyme (E1), which uses ATP to first acyl-adenylate the ubiquitin molecule and then attach it to a cysteine residue of E1 itself. 2) Conjugation, this step is done by an ubiquitin-conjugating enzyme (E2) which transfers the ubiquitin to its active site. 3) Ligation, mediated by E3 ubiquitin ligases, which is in charge of the bond formation between a lysine of the substrate protein and the C-terminal glycine of ubiquitin. Multiple rounds of ubiquitination using ubiquitin as substrate generate polyubiquitin chains. Ubiquitin contains seven lysine residues where the chain can be built, resulting in a variety of polyubiquitin chains. The topology of the polyubiquitin chain dictates the outcome of the substrate protein. Lys-48- and Lys-11-linked polyubiquitination targets substrates for degradation, while Lys-63-linked polyubiquitin chains function as scaffolds of signaling complexes.

Seven possible lysines can participate in this linkage (K6, K11, K27, K29, K33, K48 and K63). Depending on the lysine residue used for polyubiquitination a different chain topology is obtained. Furthermore, the type of polyubiquitin chain dictates the outcome on its protein target. For example, K11 and K48-linked polyubiquitination results in proteosomal degradation of the targets (degradative ubiquitination). In contrast K63-linked polyubiquitin chains serve as a scaffold to promote the assembly of protein complexes involved in signal transduction (non-degradative ubiquitination).

In the context of NF-κB pathway, the K48-linked polyubiquitination of IκB results in its proteosomal degradation and subsequent NF-κB release. Moreover, K63 ubiquitination of upstream proteins in the pathway promotes the activation of the IKK complex. One of the K63-linked polyubiquitinated proteins is RIP1. Upon TNFα binding to TNFR1, associated RIP1 gets K63-linked polyubiquitinated. These K63 chains serve as a scaffold for the recruitment of the IKK complex. In addition, the TAK1 kinase [238] recognizes and binds

the K63-polyubiquitin chain of RIP, activating the recruited IKK. This way polyubiquitination orchestrates protein interactions to achieve signal transduction.

As previously mentioned, NF-κB activity is tightly regulated at different levels making it reasonable to think that, if ubiquitination is involved in its activation, there should also be a mechanism to control it. For example two well-studied targets of NF-κB, which act as negative regulators of the pathway, are: cylindromatosis (CYLD) and A20 both of them are deubiquitinases, which regulates upstream proteins of the NF-κB pathway. CYLD directly interacts with NEMO and removes its K63-linked polyubiquitin chains resulting in the inactivation of the IKK complex. Other targets of CYLD related with the NF-κB pathway include TAK1 and RIP1 [239], [240]. Similar to CYLD, A20 also contributes to the negative regulation of the NF-κB pathway, by targeting upstream signaling proteins. A target of A20 is RIP1, which gets K48 polyubiquitinated after K63 deubiquitination. Ubiquitinated RIP1 is then targeted for degradation by the proteosome, reducing the activation of IKK and terminating NF-κB activation [208]. Many other DUBs such as USP2, USP4, USP11, USP15 and USP21 have components of the NF-κB pathway as targets [241]. The interplay between ubiquitination and deubiquitination plays a central role in the regulation of NF-κB activity.

2. RESEARCH AIMS AND OBJECTIVES

The aim of this thesis is to identify lncRNAs involved in cancer by analyzing the somatic copy number alterations (SCNAs) present in tumors.

In order to achieve this aim we propose the following objectives:

- To classify the SCNAs in cancer and identify those harboring lncRNAs.
- To select one of the identified candidates (LUAD-amp-1) and experimentally assess its function in cancer.
- To explore the mechanisms by which LUAD-amp-1 exerts its function.

3. RESULTS

3.1 Classification of the somatic copy number alterations (SCNAs) found in cancer genomes

In order to investigate the regions with frequent copy number alterations in cancer, we retrieved the SCNA data available from The Cancer Genome Atlas (TCGA) of 25 different tumor types comprising a total of 7448 tumors (Figure 1a). To detect the SCNAs we used the GISTIC 2.0 algorithm [189], which assigns a score to each alteration based on its amplitude (copy number changes) and frequency across all samples (G-score=Frequency x Amplitude). Q-values were then calculated using G-scores and a threshold of q-value < 0.25 was established to select significant alterations. The results were then reported at three different levels: regions, enlarged peaks and peaks (Figure 1b). With GISTIC 2.0 algorithm 1377 SCNAs were identified, (540 amplifications and 837 deletions). In addition, for each alteration GISTIC was able to identify focal peaks, defined as the part of the alteration with the greatest amplitude and frequency, likely to harbor a cancer driver gene. Because we aimed to pinpoint genes that function as cancer drivers, we focused the rest of our work on analyzing the focal peaks (Figure 1c).

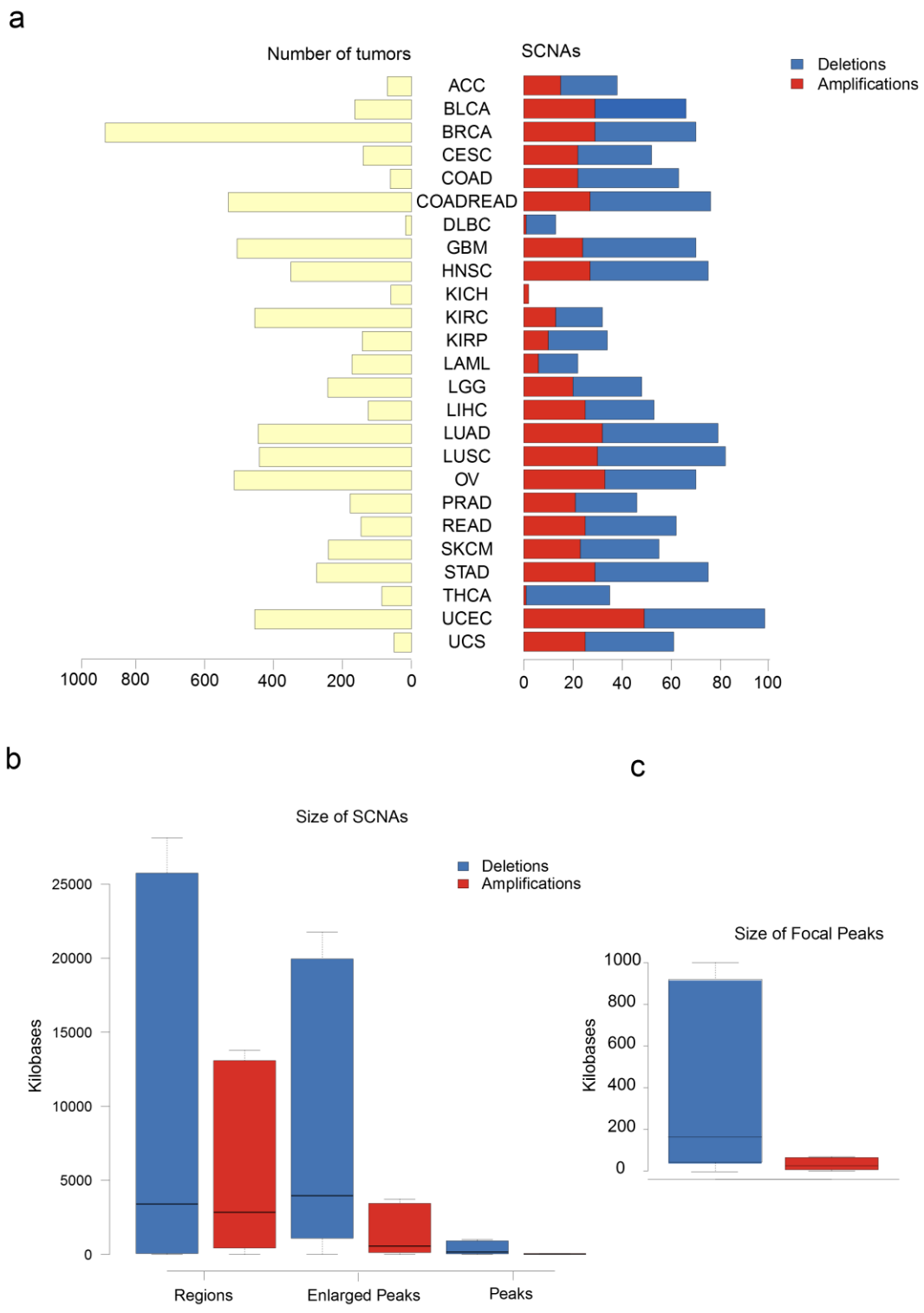

Figure 1. Tumor samples used and SCNA identified. (a) The 25 TCGA tumor types and the number of samples corresponding to each of them. Additionally, the SCNAs (deletions and amplifications) identified by the GISTIC algorithm (q-value < 0.25), used as the starting point of our analysis are represented. (b) Size ranges of deletions and amplifications at three levels: regions, enlarged peaks and peaks. (c) Size range of the focal peaks reported by GISTIC, defined as the part of the alteration with the greatest amplitude and frequency. Color code shows deletions in blue and amplifications in red.

The merging of all the focal SCNAs resulted in 1076 unique copy number altered regions. 956 of them were specific to a single tumor type while the rest (120) were recurrent alterations, present in more than one tumor type (Figure 2a). We then annotated the genes (coding and non-coding) inside the SCNAs. For this the Gencode v19 annotation of the human genome was used. This annotation includes 20,345 protein-coding and 37089 noncoding genes (13870 lncRNAs, 9013 small noncoding RNAs, 14206 pseudogenes).

Classification of the SCNAs based on the biotypes of the harbored genes resulted in the following categories: 543 SCNAs included protein-coding and noncoding genes, 312 have only protein-coding genes, while 103 only contained ncRNAs (miRNA, pseudogenes, lincRNAs, and antisense lncRNA). 97 regions did not contain any annotated gene; however, we cannot exclude that they harbor regulatory elements such as (enhancers [242] or insulators, etc.), or lncRNAs non annotated by Gencode v19 implicated in the disease. Interestingly all the identified 'gene deserts' correspond to amplifications (Figure 2b).

Importantly, 103 regions only contained non-coding genes, adding to a total of 180 genes. We annotated them with the different ncRNAs biotypes, analysis that pinpointed 101 (86 lincRNAs and 15 antisense) copy-number-altered lncRNAs (Figure 2c).

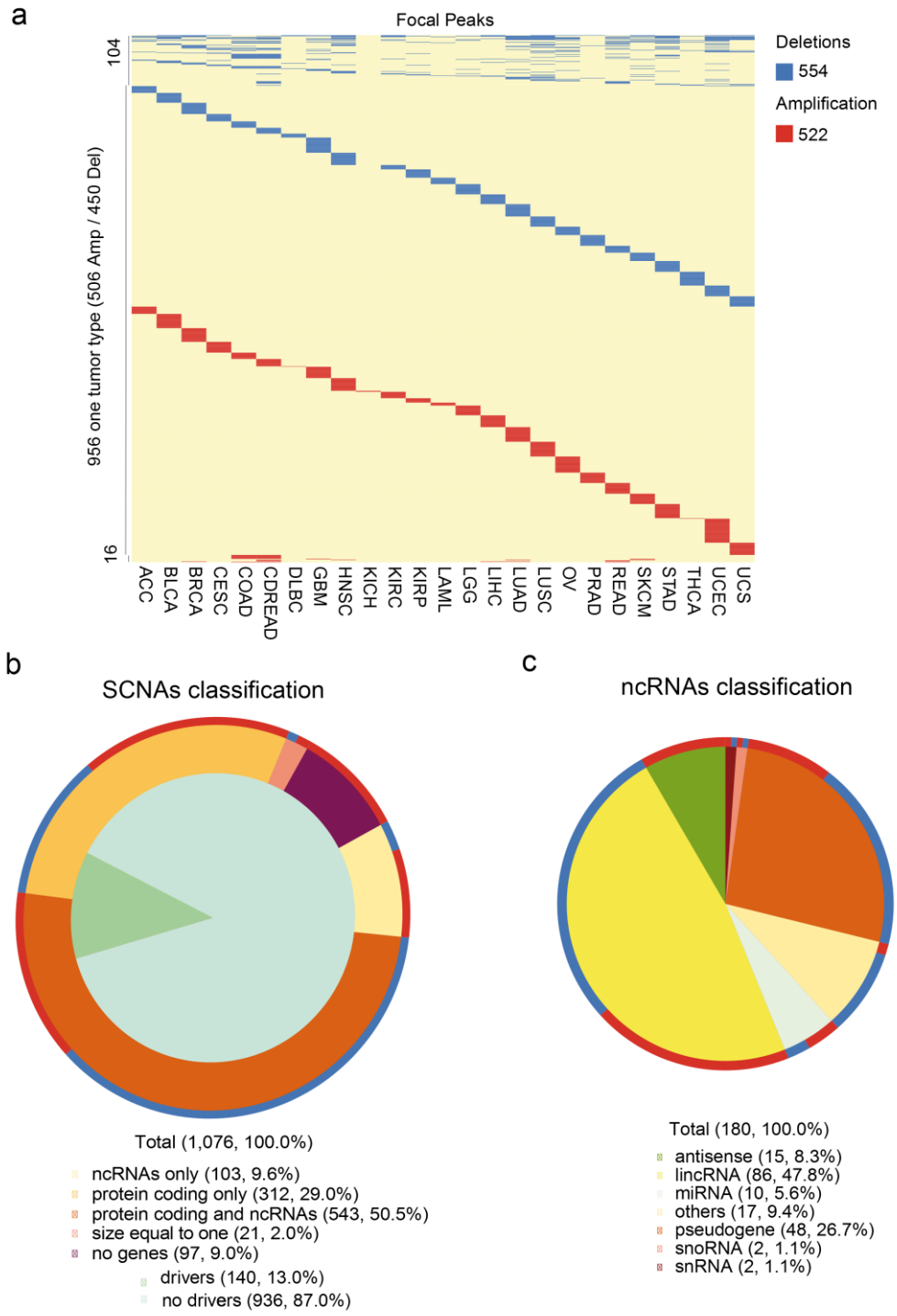

Figure 2. Overview of SCNA distribution along the different tumor types and SCNA gene annotation. (a) Distribution of the 1076 SCNAs identified; each row represents an alteration and each column a tumor type. Color code shows deletions in blue and amplifications in red. 120 of the SCNAs were recurrent alterations (104 deletions / 16 amplifications), while the other 956 were identified only in one tumor type (450 deletions / 506 amplifications). (b) Classification of the SCNAs based on the biotypes of the genes inside them. The annotation corresponds to Gencode v19. The outer circle follows the previously described color code. (c) Distribution of the ncRNA genes based on the different biotypes.

To further investigate the nature of the genes contained in the 1076 focal SCNAs, we annotated and classified them based on their oncogenic features. For this we compared the protein-coding genes contained in the SCNAs with a gene list of cancer drivers generated by Tamborero et al. [243]. In their study, Tamborero et al. define 291 protein-coding genes as high confidence drivers by integrating the results of four methods, i.e. (MuSiC, OncodriveFM, OncodriveCLUST and ActiveDriver). Each method analyzes mutation data considering different signals of positive selection (e.g. mutation frequency, functional mutation bias, clustered mutations, and phosphorylation associated mutations). Therefore we considered that the reported list is a reliable and comprehensive resource to determine if the SCNAs identified indeed contain high confidence cancer driver protein-coding genes. Using this list we found that out of the 1076 SCNAs, 140 contained a known cancer driver gene, confirming the reliability of our SCNA analysis. In agreement with this, we found tumor suppressor genes such as *PTEN* and *RB1* mapped to frequently lost regions while oncogenes such as *MYC* and *ERBB2* were found inside SCNA frequently amplified (Figure 3). On the other hand, the remaining 936 SCNA did not contain a previously defined cancer driver protein-coding gene, leaving open the possibility to discover new cancer drivers inside them.

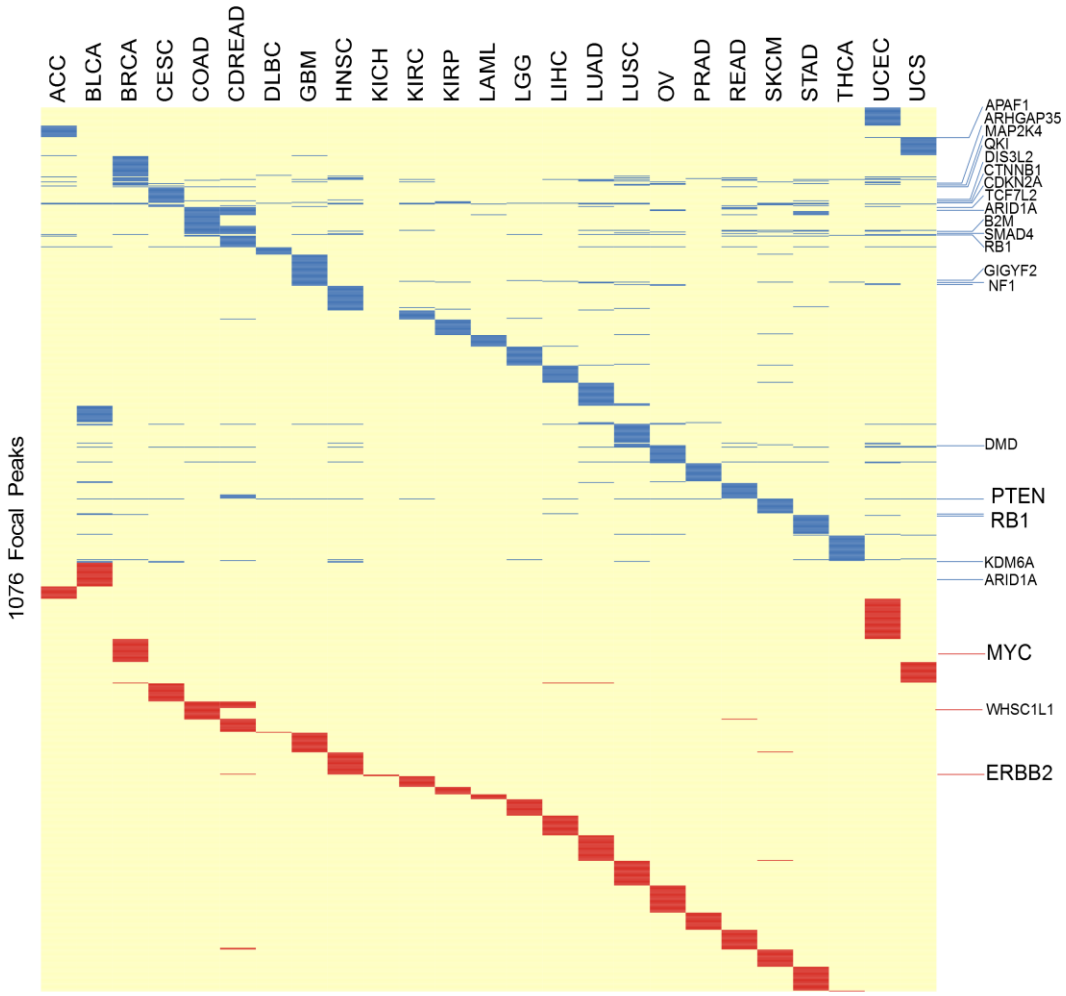

Figure 3. SCNAs harboring cancer drivers. Distribution of all the SCNAs in the different cancer types, highlighting the identified cancer driver gene (shown on the right) that maps to each alteration. Color code shows deletions in blue and amplifications in red.

In summary, the analysis of the focal SCNAs present in several tumor types identified genomic regions that do not contain any known cancer driver protein-coding gene. Moreover, some of them only contain lncRNAs that are frequently lost or amplified, and therefore are potentially involved in cancer.

3.1.1 LncRNAs inside SCNAs are regulated by cancer-related transcription factors (TF)

To increase our insight into the potential functional implication in cancer of the lncRNAs within frequent SCNAs, we analyzed their regulation by relevant transcription factors. To that end we retrieved the transcription start sites (TSS) of the copy number altered lncRNAs and arbitrarily defined their regulatory region -1 Kb upstream and 1 Kb downstream from them. The resulting genomic coordinates were then intersected with the binding sites of 161 transcription factors (TFs) obtained from Chromatin Immunoprecipitation (ChIP-seq) experiments reported by ENCODE. Interestingly, amplified lncRNAs were enriched for oncogenic factors such as MYC (p-value = $6.93e^{-7}$), MAX (p-value = $6.97e^{-6}$) and JUND (p-value = 0.0005201) among others (Figure 4a). On the other hand, deleted lncRNAs were enriched in binding sites for POUF51/OCT4, (p-value = 0.0080) and NANOG (p-value = 0.0384), pointing towards a regulation by cancer related transcription factors (Figure 4b). The complete list of the 161 TFs from ENCODE and their associated p-values obtained in this analysis are presented in the 'TFBS.counts' table of the appendix. The list is organized from the lowest to the highest p-value.

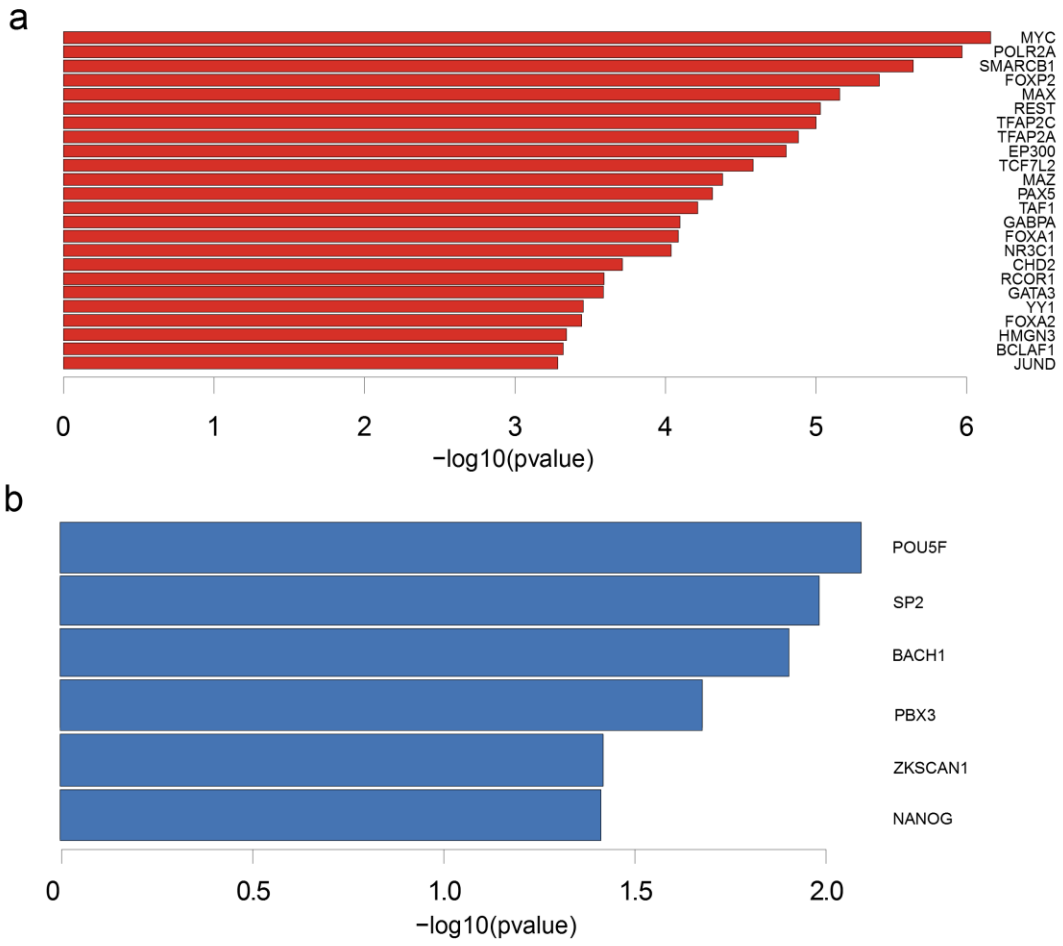

Figure 4. Transcription factors associated to copy number altered lncRNAs. (a) Transcription factors significantly enriched near amplified lncRNAs and (b) deleted lncRNAs, the -log10 (p-value) is represented.

3.1.2 Chromatin states associated with copy number altered lncRNAs

To further investigate the specific characteristic of the identified copy numbered altered regions, we analyzed them using the chromatin state segmentation track generated by ENCODE. This track displays chromatin state segments of the genome (active promoter, weak promoter, poised promoter, strong enhancer, weak enhancer, insulator, transcriptional transition, transcriptional elongation, weak transcribed, Polycomb-repressed, heterochromatin, repetitive), which were inferred by using ChIP-seq data from histone marks and chromatin binders in multiple cell types (H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K9ac, H4k20me1, CTCTF).

By analyzing the chromatin states associated with the copy number altered lncRNAs we found out that the amplified lncRNAs were enriched in the poised promoter chromatin state in 7 of the 9 cell lines analyzed. In contrast, amplified protein coding genes where enriched for this state in only 2 of the 9 cell lines. Similarly, for the deleted lncRNAs an enrichment of poised promoters was observed, that was no present in the deleted protein coding genes. The copy number altered lncRNAs were enriched in chromatin regions tagged as enhancers in all the cell types included in the anlaysis. However, no differences were observed in the enhancers chromatin states (strong or weak), between copy number altered protein coding and lncRNAs. Interestingly, the chromatin state defining insulator regions was enriched exclusively in the deleted lncRNAs; neither the deleted proteins nor the amplified lncRNAs showed this enrichment. Insulator proteins bind to DNA in specific locations called insulator regions. These regions act as boundary elements for separating different chromatin domains [244]. CTCF (CCCTC binding Factor) is the main insulator protein in vertebrates. Recent studies have also implicated lncRNA in the regulation of CTCF function in chromatin looping. Our results suggest that frequently deleted lncRNAs could be involved in a *cis*-acting mechanism regulating CTCF in insulator regions. (Figure 5)

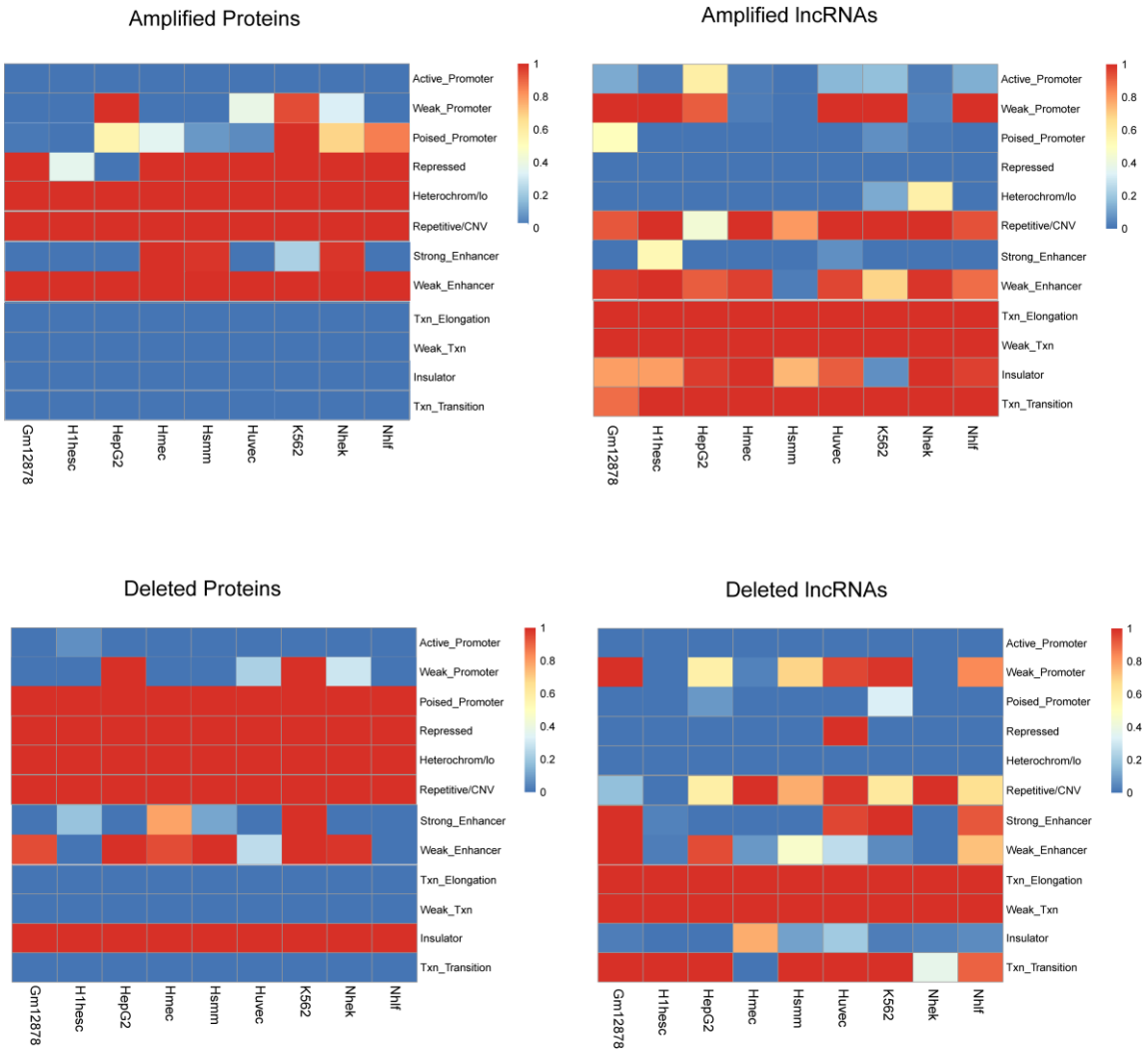

Figure 5. Chromatin states associated with copy number altered genes. Heatmaps representing the p-values associated to the enrichment of several chromatin states in 9 different cell lines for amplified and deleted protein coding genes and lncRNAs. p-values are represented using the color scale shown.

3.1.3 Several lncRNA inside SCNAs harbor cancer-associated SNPs

Thousands of SNPs have been shown to significantly associate with human cancer risk and progression. However, the remaining challenge is to identify those SNPs that are functional and link the effect of the variant to the disease. We sought to detect functional lncRNA-SNPs associations [155]. In order to do this, a list of curated cancer-associated SNPs (719, p-value < 10^{-5}) was obtained from the Genome Wide Association Study catalog [245]. By mapping these SNPs to the coordinates of copy-number altered lncRNAs we obtained the list presented as 'SNPs.lncRNAs' of the appendix. Among the SNPs mapping to copy numbered altered lncRNAs we found nine of them mapping to the *CASC8* locus. The *CASC8* locus encodes for various transcripts involved in cancer, including *CARLo-5* (also known as *CCAT1*) [246]. It has been shown that *CARLo-5* expression correlates with the rs6983267 genotype associated to increased cancer susceptibility (the rs6983267 SNP and long non-coding RNA CARLo-5 are associated with endometrial carcinoma), suggesting that other identified SNPs could have a similar role. In addition, in this analysis we identified several loci of cancer related lncRNAs, which harbor cancer-associated SNPs. For example: *LINC-PINT* (rs2048672), *MIR99AHG* (rs2823779), and *CASC11* (rs10094872, rs9642880).

3.1.4 Differential expression of lncRNAs contained within focal SCNAs

At the time we were conducting the SCNA analysis a publication from Iyer et al. [32] presented an *ab initio* transcriptome assembly using RNA-seq libraries from TCGA, ENCODE and other public datasets. The *ab initio* reconstruction does not rely on existing annotations, enabling the discovery of previously unannotated transcripts. Moreover, this approach applied to a large number of samples (n > 7,000) helped delineate new transcript structures and expression in specific cancer types and cell lineages, resulting in the most comprehensive annotation available so far. We therefore used this new annotation named *MiTranscriptome* to reannotate the previously identified SCNAs. The analysis pipeline we followed is summarized in Figure 6. Out of the 1076 unique SCNA coming from GISTIC, 140 contained a previously known cancer driver. Further gene annotation using Gencode classified the alterations into four categories: 97 with 'no genes', 473 with 'coding and noncoding', 242 with 'only coding', and 103 with 'only

noncoding' genes. Subclassification of the 'noncoding only' category identified 67 SCNA containing lncRNAs plus other noncoding RNAs (miRNAs, snRNAs, etc.) and 36 SCNAs with 'only lncRNAs'. The combination of the 36 SCNA with 'only lncRNAs' and RNA-seq data from *MiTranscriptome* pinpointed 20 putatively functional lncRNAs. In addition we found that 13 SCNA previously classified as having 'no genes', with *MiTranscriptome* annotation came out as having an associated gene.

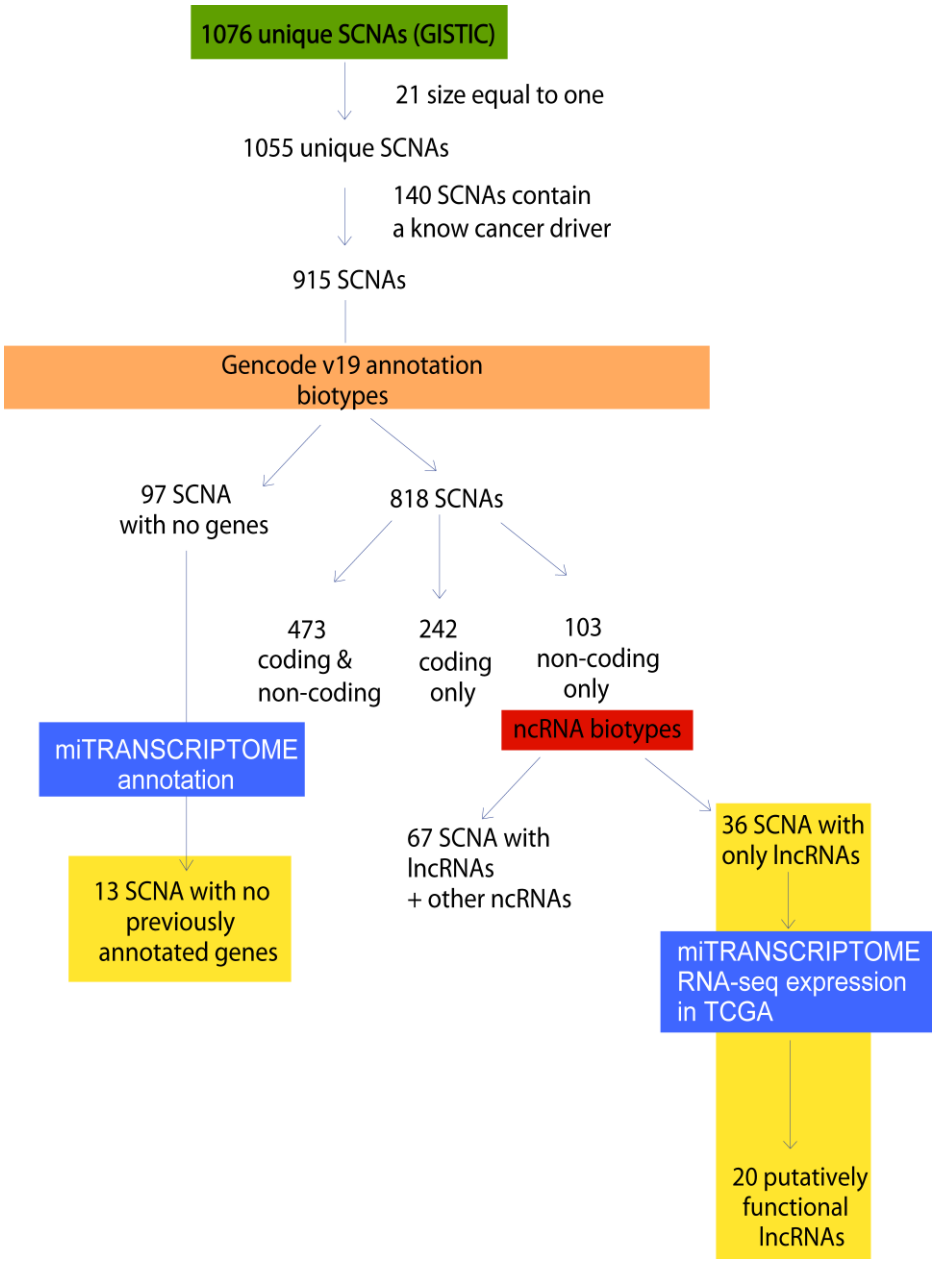

Figure 6. Pipeline for selection of SCNAs harboring lncRNAs. The 1076 significant SCNAs (q-value < 0.25) used as the starting point of the analysis were annotated and classified based on the biotypes of the genes contained in them. Subsequent filtering and integration of expression from *MiTranscriptome* resulted in a list of 20 putatively functional SCNAs.

MiTranscriptome not only helped us with the annotation of new genes, but it also enabled us to assess the expression of lncRNAs across different tumor types. Based on the idea that a functional SCNA should result in expression changes of the containing genes, we cross-compared our list of copy number altered lncRNAs with the RNA expression profiling from *MiTranscriptome*. The comparison identified a total of 20 putative functional alterations of lncRNAs (14 amplifications and 6 deletions) where expression changes were consistent with copy number status i.e., the lncRNA showed a significant difference in the expression when comparing normal and tumor samples in which the amplified lncRNAs were more highly expressed and the deleted lncRNAs less expressed. The list of candidates is presented in the following table:

	SCNA_ID	SCNA	Expression	Tumor type	lncRNA_ID
1	CNA_36	Amplification	Tumor > Normal	COADREAD	
2	CNA_88	Amplification	Tumor > Normal	UCEC	
3	CNA_200	Amplification	Tumor > Normal	READ	*CARLo-1/CASC8*
4	CNA_359	Amplification	Tumor > Normal	COADREAD	
5	CNA_623	Amplification	Tumor > Normal	LUAD	
6	CNA_19	Amplification	Tumor > Normal	LIHC	
7	CNA_201	Amplification	Tumor > Normal	HNSC	
8	CNA_202	Amplification	Tumor > Normal	LUSC	*CARLo-1/CASC8*
9	CNA_875	Amplification	Tumor > Normal	HNSC	
10	CNA_203	Amplification	Tumor > Normal	BLCA	
11	CNA_469	Amplification	Tumor > Normal	STAD	
12	CNA_526	Amplification	Tumor > Normal	UCS	
13	CNA_181	Amplification	Tumor > Normal	GBM	
14	CNA_190	Amplification	Tumor > Normal	UCS	*PVT1*
15	CNA_244	Deletion	Tumor < Normal	STAD	
16	CNA_182	Deletion	Tumor < Normal	ACC	
17	CNA_922	Deletion	Tumor < Normal	SKCM	
18	CNA_605	Deletion	Tumor < Normal	KIRCH	
19	CNA_887	Deletion	Tumor < Normal	THCA	
20	CNA_793	Deletion	Tumor < Normal	LUAD	

Table 2. Candidate lncRNAs

Among the amplified regions selected by this method we found some previously characterized oncogenic lncRNAs such as *PVT1*, localized downstream from the *MYC* locus [138], or *CARLo-5* [246] located in an amplified region upstream of *MYC*. However, most of the identified lncRNAs remain functionally uncharacterized.

3.2 LUAD-amp-1, a lncRNA targeted by amplification in lung cancer

In order to further explore the possible role of some of the lncRNAs identified by our SCNA analysis, we decided to focus on those identified in lung cancer (LUAD and LUSC). We selected this tumor type based on the fact that lung cancer is still the leading cause of cancer death worldwide.

Three candidate alterations were identified in lung cancer: CNA_202 (LUSC), CNA_623 (LUAD) and CNA_793 (LUAD). The analysis of the lncRNAs mapping to them showed that CNA_202 contained the *CASC8* lncRNA, which has been related with prostate and bladder cancer [247], [248]. On the other hand the CNA_793 mapped to the frequently deleted Prader Willis/Angelman region containing seven noncoding RNAs (*PWRN2, RP11-580I1.1 RP11-580I1.2, RP11-350A1.2, RP11-107D24.2, PWRN3* and *PWRN1*) [249]. Since some research has been done in the lncRNAs within the CNA_202 and CNA_793 we decided to focus on the CNA_623, which maps to the previously uncharacterized lncRNA RP11-231D202.2.

First we set to test whether the amplification detected in TCGA cohort was also found in independent cohorts of lung cancer patients. Indeed, while the analysis of the TCGA LUAD cohort showed that 43 / 493 (8.7%) samples contain this alteration. The amplification of this region was consistently validated in additional tumor cohorts provided by the Solid Tumor laboratory (cohorts from Pamplona, Sweden, Wistuba with amplification in 8.24, 5-94 and 5.56% of the tumors respectively). These data suggest that the results derived by our analysis of the TCGA data can be extrapolated to other samples independently obtained.

Cohort	Amplified	Total	% Amplified
TCGA.LUAD	43	493	8.72
Pamplona	7	85	8.24
Sweden	6	101	5.94
Wistuba	9	162	5.56

The CNA_623 focal peak (chr8: 42093819-42107104) uniquely maps to the lncRNA RP11-231D20.2 without containing any protein-coding gene. RP11-231D20.2 is located between plasminogen activator tissue type (*PLAT*) and inhibitor of nuclear factor kappa B kinase subunit beta (*IKK*β*/IKBKB*) protein-coding genes, being a divergent antisense transcript of *IKBKB*. Hereon we will to refer to it as LUAD-amp -1 (Figure 7).

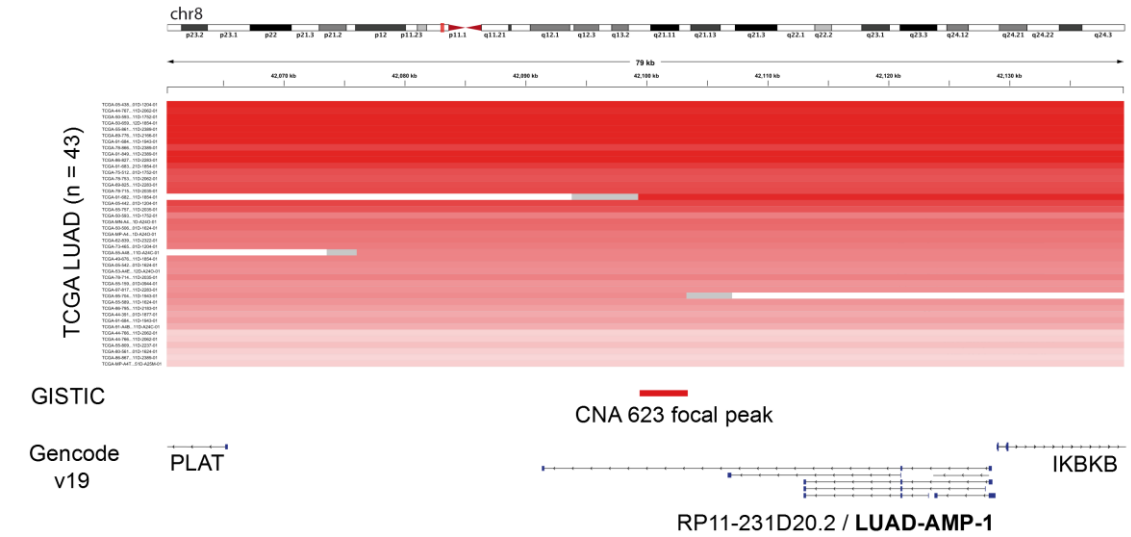

Figure 7. Genomic location of the focal peak pinpointing LUAD-amp-1. CNA_623 identified in the 43 tumor samples of TCGA-LUAD 43/493 (8.7%). In addition, the focal peak (chr8: 42093819-42107104) and the lncRNA RP11-231D20.2/LUAD-AMP-1 mapping to it is presented.

3.2.1 LUAD-amp-1 gene structure and detection

Gencode v19 annotation shows six different transcripts associated to LUAD-amp-1 (*ENSG0000023408*) gene: ENST00000523459, ENST00000521470, ENST00000521802, ENST00000520890, ENST00000518994, ENST00000518213 (Figure 8a). In order to detect the average expression levels of LUAD-amp-1 we used both qRT-PCR and RNA-seq analyses. First, qRT-PCR primers were designed mapping to exon A shared between three of the six isoforms (3/6), exon C (5/6), exon D (3/6), and intron-spanning primers (A-C and C-D). Total RNA from lung cancer cell lines, was extracted and analyzed by qRT-PCR with the primers. The expression levels of exon C, exon D and intron-spanning C-D were similar, and no signal coming from intron-spanning junction A-C was obtained (Ct = undetermined)) (Figure 8b). RNA-seq data from ENCODE support these results, where only a consistent signal is present between exons, C and D (Figure 8c). In addition, to assess which of the isoforms of LUAD-amp-1 is present in the tumor samples, RNA-seq data from TCGA-LUAD was analyzed. A tumor sample with a mean expression of LUAD-amp-1 was selected, and the RNA-seq reads were mapped to the locus. The RNA-seq data supports the expression of the isoform ENST00000521802 (shown in red) (Figure 8d), isoform composed by three exons (B, C, and D), and with a total transcript length of 415 nucleotides.

Based on these analyses, for subsequent experiments we focused on the ENST00000521802 form of LUAD-amp-1 and the qRT-PCR primers mapping on exon C, D, and junction C-D were used to check its expression changes.

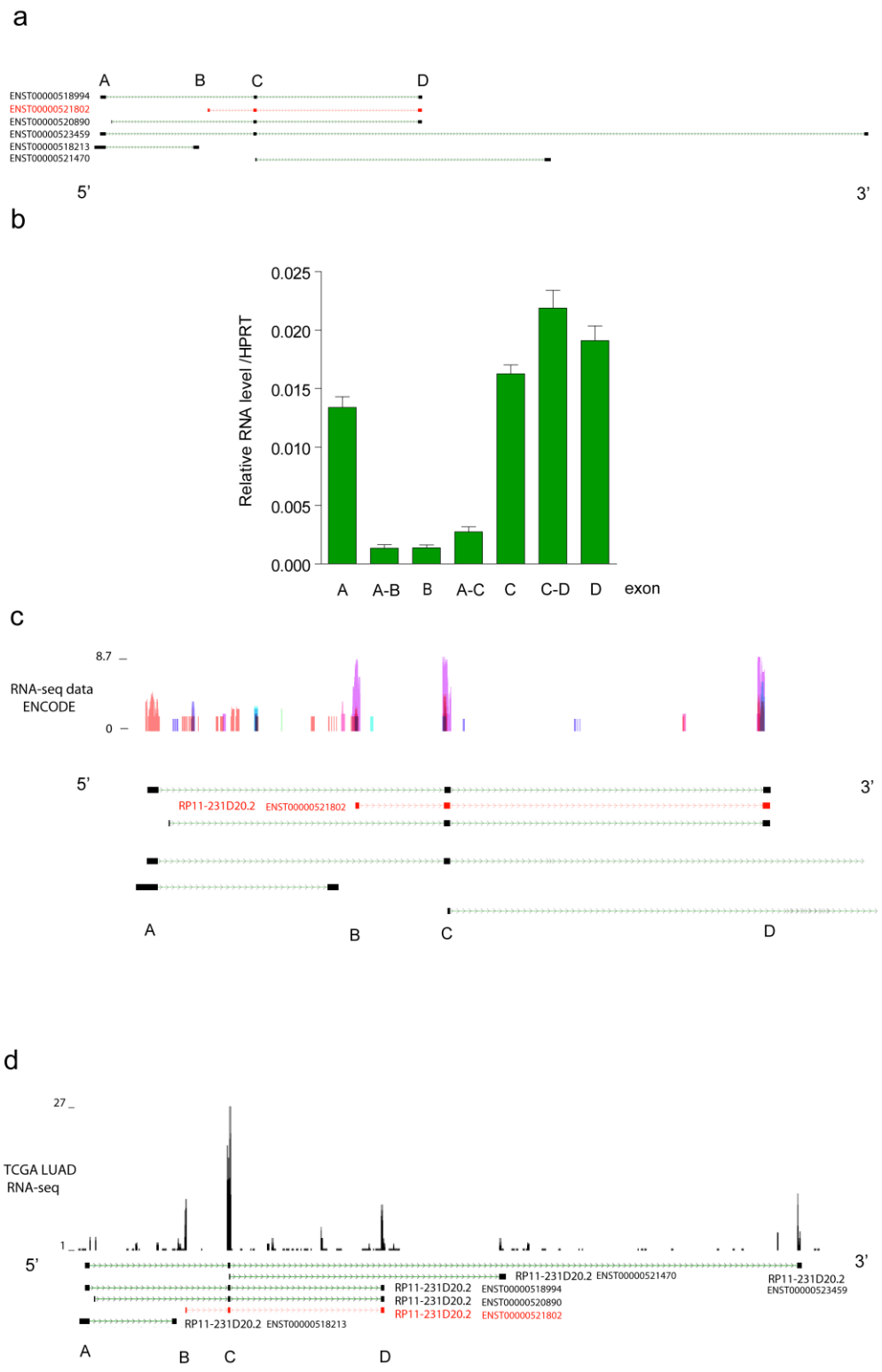

Figure 8. **LUAD-amp-1 gene structure and isoform expression**. (a) The six different isoforms of LUAD-amp-1 (RP11-231D02.2 Gencode v19) and the Ensembl Transcript ID for each of them are presented. Isoforms are shown in the 5' to 3' direction and exons are identified with letters (A, B, C, D. (b) RNA expression of LUAD-amp-1 quantified with several primer sets mapping to the different exons (qRT-PCR values are represented normalized to HPRT). (c) RNA-seq track showing the expression of LUAD-amp-1 in the 9 ENCODE cell lines. (d) RNA-seq data from TCG-LUAD supports the isoform ENST00000521802 (red), which was used for the rest of the experiments.

3.2.2 LUAD-amp-1 is overexpressed in several tumor types

To evaluate the possible role of LUAD-amp-1 in lung cancer, we wanted to test whether LUAD-amp-1 was indeed overexpressed in the samples where the amplification was present. To do this, we classified the samples in two groups: with (43) and without amplification (322). We observed a significantly higher expression of LUAD-amp-1 in the amplified group (Figure 9a), indicating that the amplification of this genomic region leads to higher expression levels of LUAD-amp-1. In addition, the expression analysis of all the TCGA-LUAD samples (comparing normal vs tumor) also showed a significant difference on LUAD-amp-1 expression levels (Figure 9b). Further analysis of the TCGA-LUAD cohort revealed that around 66% of the tumor samples having a higher expression of LUAD-amp-1 compared to the normal samples lack the amplification of LUAD-amp-1 locus (Figure 9c). Moreover, LUAD-amp-1 was also identified as overexpressed in other tumor types: lung squamous carcinoma (LUSC) and head and neck carcinoma (HNSC) (Figure 9d). Additionally, in three other cohorts (GSE19188, GSE18842, GSE19804) the expression of LUAD-amp-1 was found higher in tumors compared to nontumoral samples (Figure 9e). The observed overexpression LUAD samples lacking amplification as well as in the additional tumor types suggests that other mechanisms apart from amplification could be involved in LUAD-amp-1 overexpression. Overall, we identified LUAD-amp-1 as a lncRNA targeted by amplification in lung cancer which is expression is up regulated in tumor samples.

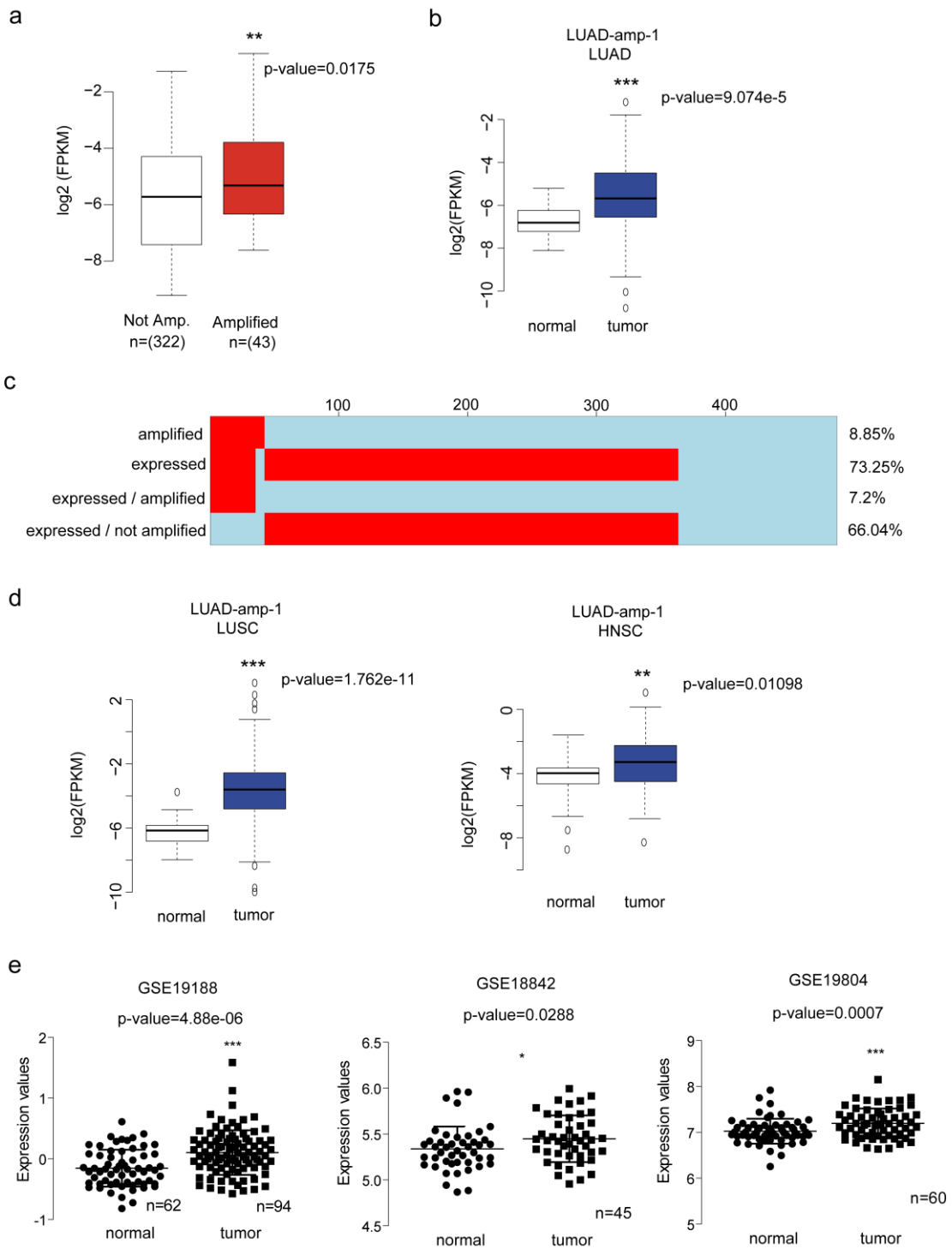

Figure 9. LUAD-amp-1 is overexpressed in several lung cancer tumor cohorts. (a) LUAD-amp-1 expression quantified with RNA-seq data from TCGA, samples were grouped based on the presence or absence of the LUAD-amp-1 amplification. (b) Separating samples into normal and tumor reveals an overexpression of LUAD-amp-1. (c) TCGA-LUAD cohort showing the percentage of samples with amplification and/or overexpression. (d) Overexpression of LUAD-amp-1 in additional tumor types (LUSC, HNSC). (e) LUAD-amp-1 expression from microarray data of additional series of patients, using the probe 231378_at. GSE19188, 91 NSCLC tumor and 65 adjacent normal lung tissues samples; GSE18842, 46 tumor samples and 45 paired nontumoral samples of NSCLC; GSE19804, 60 pairs of tumor and adjacent normal lung tissue from nonsmoking female cancer patients. Statistical significance is represented as (*) p-value ≤ 0.05, (**) p-value ≤ 0.01, (***) p-value ≤ 0.001.

3.2.3 The promoter of LUAD-Amp-1 is hypomethylated in lung squamous carcinoma (LUSC)

Interestingly, the highest difference in LUAD-amp-1 expression comparing tumor versus normal samples was observed in the LUSC cohort (p-value=$1.762e^{-11}$) where no amplification of LUAD-amp -1 was identified. To investigate if other mechanisms could explain the observed overexpression we analyzed the DNA methylation of LUAD-amp-1 loci. Two differentially methylated CpG (cg26394282, cg16230352) mapping to the 5' end of LUAD-amp-1 were identified (p-value=$1.89e^{-24}$, p-value=$3.05e^{-19}$) in the LUSC tumor cohort (Figure 10a), suggesting that hypomethylation could explain the observed overexpression of LUAD-amp-1 in this tumor type. These results were confirmed in a different tumor cohort from IIS La Fe (cg26394282 p-value=$7.09e^{-19}$, cg16230352 p-value= $6.48e^{-09}$) (Figure 10b). The two hypomethylated probes (cg26394282, cg16230352) mapping close to the 5' end of LUAD-amp-1 are represented in Figure 12a. Together all these data indicate that LUAD-amp-1 as a lncRNA targeted by genetic (amplification LUAD), and epigenetic mechanisms (hypomethylation LUSC); suggesting that it could have a role as an oncogene in lung cancer pathogenesis.

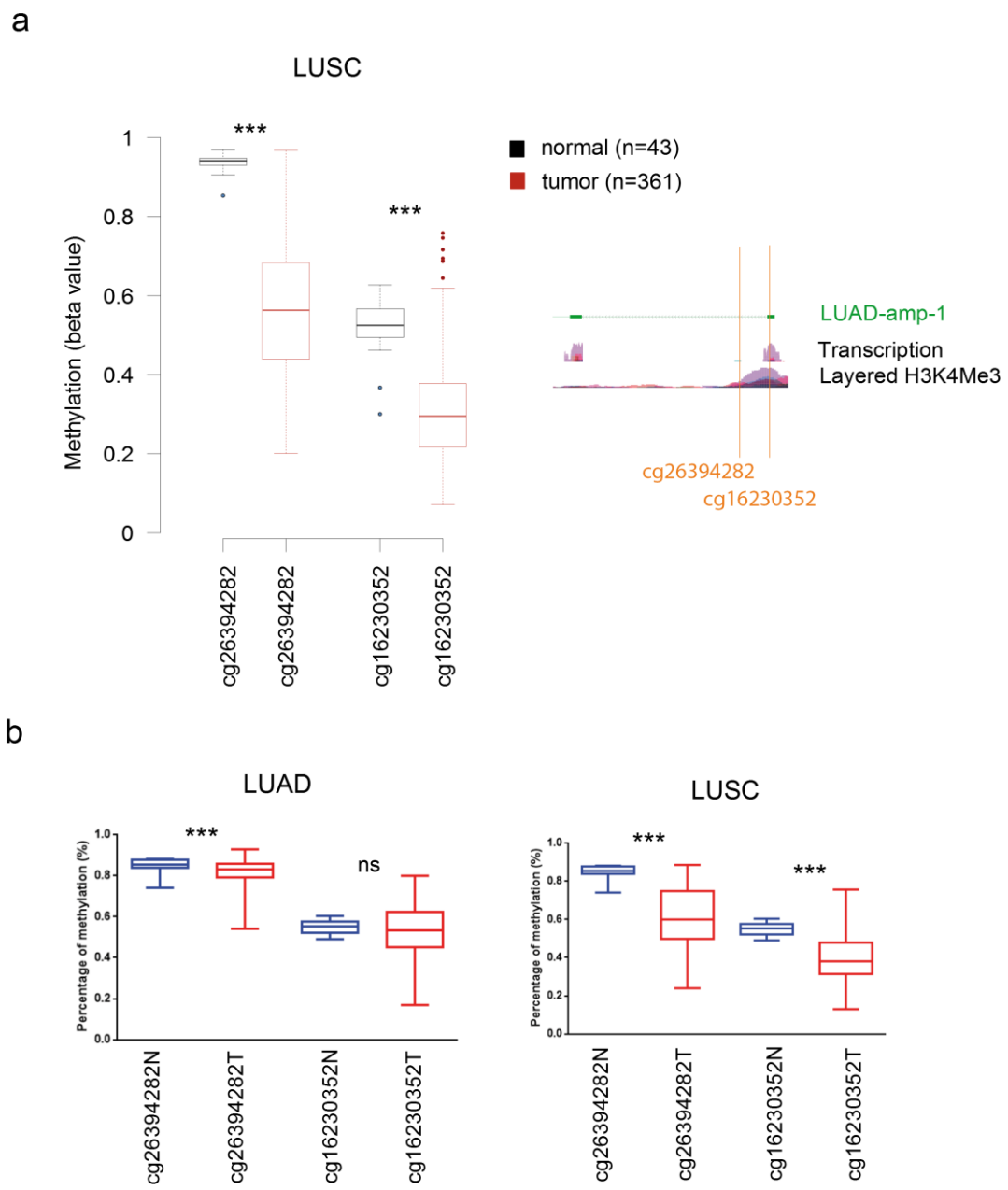

Figure 10. Differential methylation of LUAD-amp-1. (a) The two differentially methylated CpGs (cg26394282, cg160230352) from LUSC samples, mapping near one of the exons of LUAD-amp-1. Quantification is reported using the Beta-value, the ratio of the methylated probe intensity and the overall intensity (sum of methylated and non-methylated probe intensities). (b) Additional NSCLC cohorts (LUAD and LUSC) from IIS La Fe , showing the percentage of methylation values for the CpGs mapping to LUAD-amp-1. Statistical significance is represented as (*) p-value ≤ 0.05, (**) p-value ≤ 0.01, (***) p-value ≤ 0.001.

3.3 Functional characterization of LUAD-amp-1

In order to experimentally test the potential role of LUAD-amp-1 as an oncogene we first set to find lung cancer cell lines with a genetic background similar to the one present in the tumor samples (amplification of LUAD-amp-1). For this, we interrogated The Cancer Cell Line Encyclopedia (CCLE). The HCC95 cells showed the highest level of amplification of the LUAD-amp-1 region, followed by the H1648 cell line (Figure 11a).

To experimentally confirm the presence of the amplification we extracted genomic DNA from HCC95 and H1648, and quantified by qPCR the number of copies. Values obtained from BJ fibroblast cells, where LUAD-amp-1 locus is not amplified, were used as a reference of two copies. Using this quantification we observed and increase in the number of copies of LUAD-amp-1 region in HCC95 and H1648 cells (~8 copies) (Figure 11b).

We then checked the expression of LUAD-amp-1 in the same cell lines, confirming a high expression in HCC95 and H1648 compared to other lung cancer cells BJ fibroblasts, A549 and H2170. Interestingly, A549 and H2170 cells do not have copy number changes, suggesting that the detected amplification in the HCC95 and H1648 could be responsible for the expression of LUAD-amp-1, similar to what we observed in the tumor samples (Figure 11c). We thus identified two cell lines that have high expression and amplification of LUAD-amp-1 locus, making them valuable tools to carry out experiments and investigate the role of LUAD-amp-1 in lung cancer.

Further analyses of LUAD-amp-1 amplification and expression in a lung cancer cell line panel from CCLE, revealed that around 12% of the cells analyzed have amplification of LUAD-amp-1. In this same cell panel 10.97% of the cells show both expression and amplification of LUAD-amp-1 (Figure 11d).

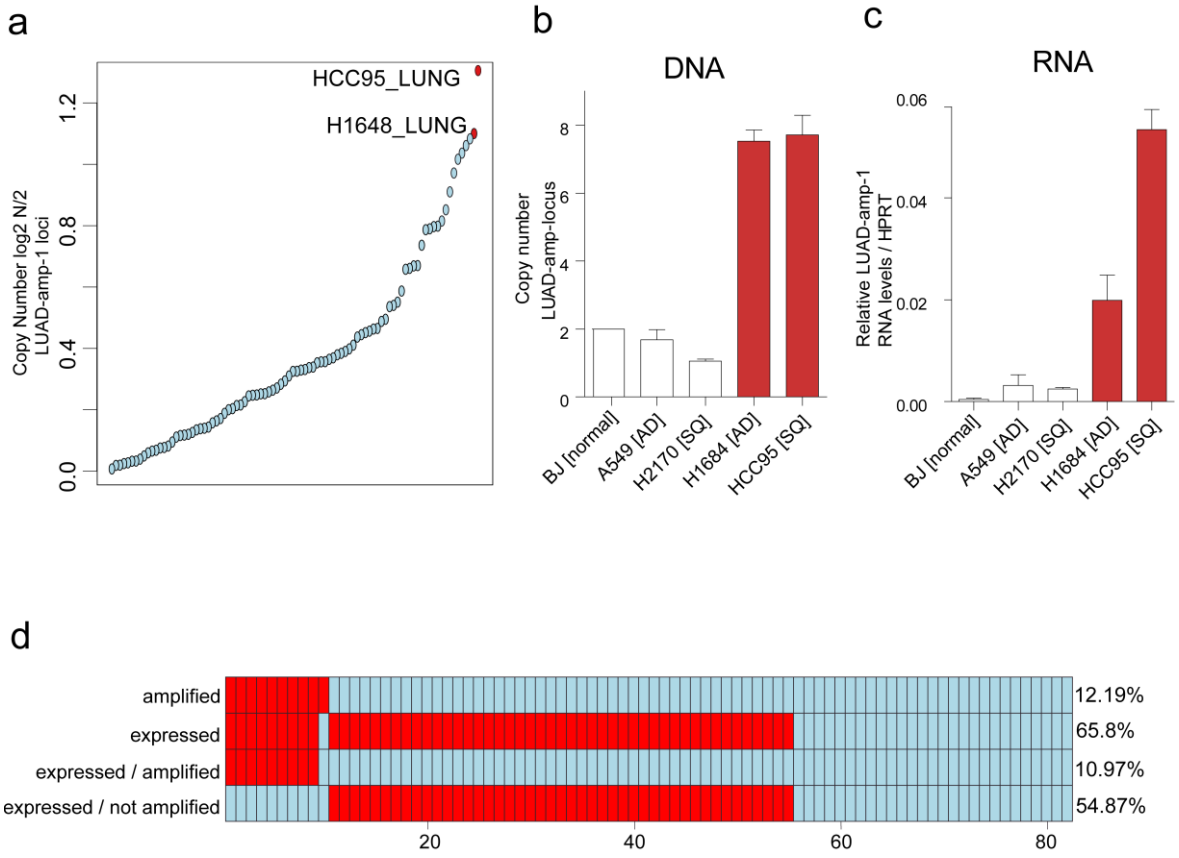

Figure 11. LUAD-amp-1 amplification in lung cancer cell lines. (a) Copy Number of the LUAD-amp-1 loci in different lung cancer cell lines, data retrieved from CCLE. (b) Experimental calculation of the copy number of LUAD-amp-1 loci using qRT-PCR from genomic DNA. The obtained qRT-PCR values are represented normalized to the PEX19 genomic regions, which has no significant aneuploidy in these cell lines and relative to the BJ cells (normal lung fibroblasts) having two copies. (c) RNA expression of LUAD-amp-1 in the same cell lines where copy number was measured, [AD] adenocarcinoma, [SQ] squamous cell carcinoma. qRT-PCR RNA expression values were normalized to HPRT. (d) Amplification and expression of LUAD-amp-1 (marked in red) in the lung cancer cell panel from CCLE.

3.3.1 CRISPR/Cas9-mediated deletion of LUAD-amp-1 limits cell proliferation

As a first approach to test experimentally the effect of LUAD-amp-1 amplification we decided to revert it by applying the genome editing technology, CRISPR/Cas9. For this we used the HCC95 cell line that presented both, amplification and expression of LUAD-amp-1. Two single guide RNAs (sgRNAs) flanking exon 3 of LUAD-amp-1 were designed to obtain HCC95 clones with a genomic deletion of around 500 bp (Figure 12a). HCC95 cells were transfected with the CRISPR/Cas9 and sgRNAs-expressing plasmids, and after one day single cells were separated in 96 wells plates using Fluorescence-Activated Cell Sorting (FACS). The obtained clones were genotyped using a set of external oligos mapping to the flanking regions where the sgRNAs were designed (*out* primers), and also a set of internal primers mapping to the intended deletion (*in* primers) (Figure 12b). PCR and gel electrophoresis showed that the obtained clones were heterozygous and contained the intended deletion (confirmed by Sanger sequencing, Appendix). Additionally, qRT-PCR results using gDNA demonstrated a reduction in the copy number of LUAD-amp-1 exon 3 (Figure 12c). Because of the multiple copies of this region in the genome of HCC95 cells and the low frequency of two editing events occurring in the same cell, probably several rounds of editing are needed to obtain clones with homozygous deletion. However, the obtained heterozygous clones had a reduction of LUAD-amp-1 RNA levels (Figure 12d).

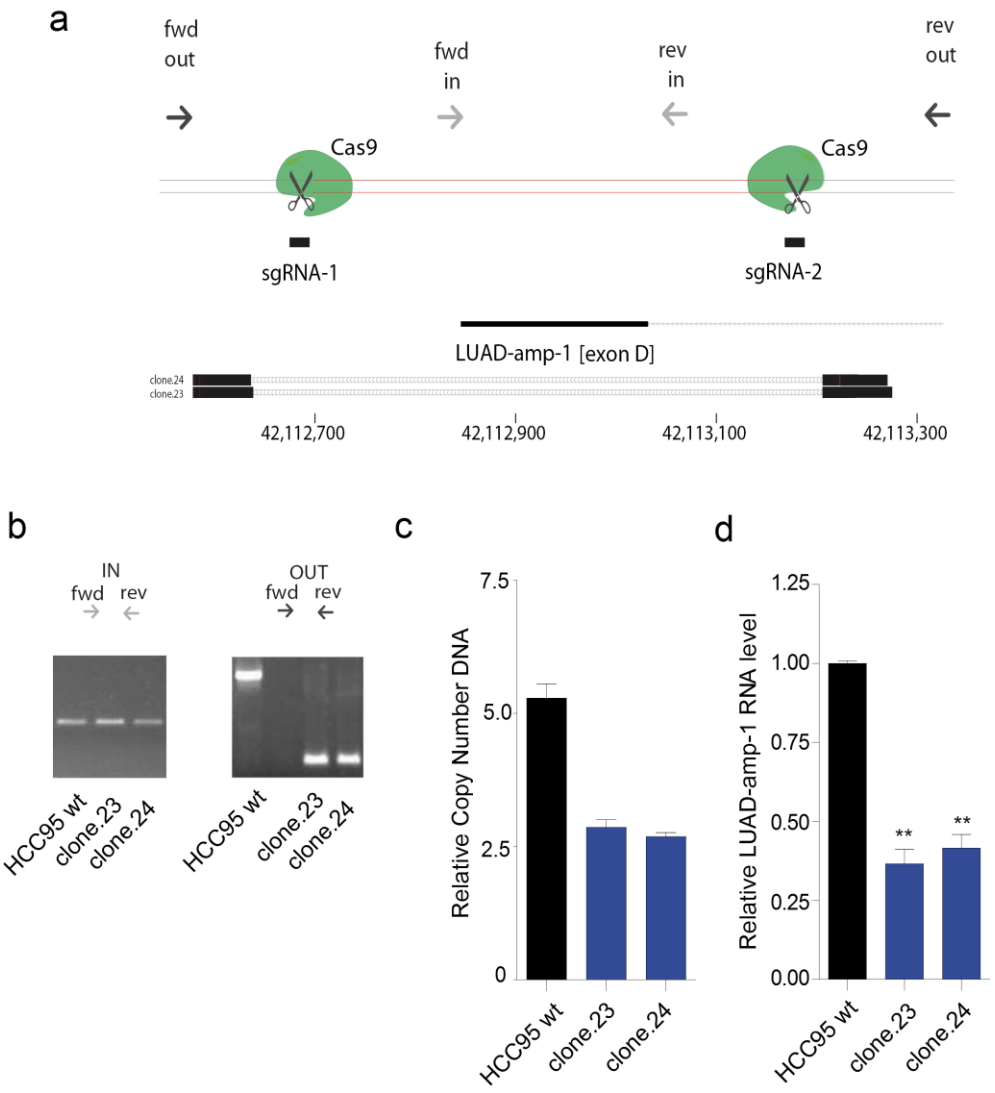

Figure 12. CRISPR/Cas9 targeting of LUAD-amp-1 loci. (a) Schematic representation of the CRISPR/Cas9 strategy used to delete exon D of LUAD-amp-1. Single guide RNAs (sgRNAs) mapping to the flanks of exon D where the CRISPR/Cas9 is targeted to cut. The annealing sites for the outside primers used (fwd, rev) to screen for the deletion are indicated. BLAT results from the obtained sequences of the PCR products, showing the boundaries of the obtained deletion (b) Gel images showing the PCR products obtained with the fwd/rev in and out primers using gDNA from the parental HCC95 cells and the clones (23 and 24). (c) Relative DNA copy number of LUAD-amp-1 exon D quantified by qRT-PCR using gDNA from the CRISPR/Cas9 positive clones. (d) RNA expression of LUAD-amp-1 in the CRISPR/Cas9 positive clones, values represented normalized to HPRT and relative to the expression of LUAD-amp-1 in wt cells. Graph shows mean ± SD of three independent experiments. Statistical significance is represented as (*) p-value ≤ 0.05, (**) p-value ≤ 0.01, (***) p-value ≤ 0.001.

To test the resulting phenotype of copy number reduction of LUAD-amp-1 we performed cell proliferation assays with the CRISPR/Cas9 clones. These experiments revealed a decrease in cell growth when LUAD-amp-1 copy number was reduced. In addition, colony formation experiments showed a decrease in the number of colonies (Figure 13 a, b). These results suggest that amplification of LUAD-amp-1 contributes to the highly proliferative phenotype of cancer cells.

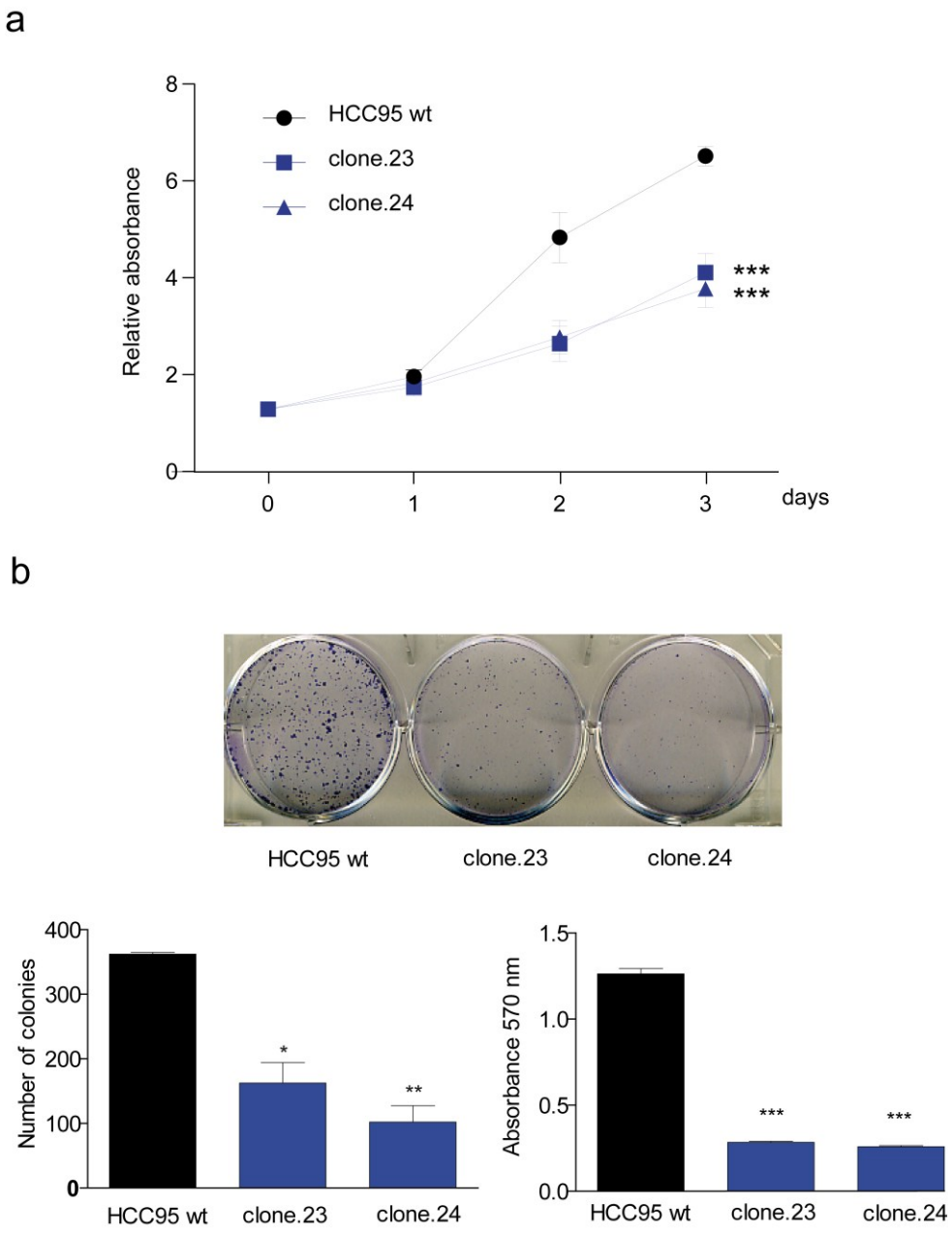

Figure 13. Copy number reduction of LUAD-amp-1 impairs cell proliferation. (a) Cell proliferation rate, measured by MTS, of HCC95 cells engineered with CRISPR/Cas9 technology. A population of parental cells was used as a control (b) Colony formation capacity was measured by clonogenic assays in cells with (clone 23, 24) and without (wt) LUAD-amp-1 deletion. The number of colonies and the crystal violet (CV) absorbance values are shown. Graphs show mean ± SD of three independent experiments. Statistical significance is represented as (*) p-value ≤ 0.05, (**) p-value ≤ 0.01, (***) p-value ≤ 0.001.

3.3.2 LUAD-amp-1 inhibition reduces cell proliferation of lung cancer cells *in vitro*

The approach of using genomic editing with CRISPR/Cas9 to assess the effect of LUAD-amp-1 inhibition did not allow us to distinguish whether the observed effect in cell proliferation was due to copy number reduction of the LUAD-amp-1 locus or the decreased expression of LUAD-amp-1 RNA. Thus, in order to differentiate between both possibilities we carried out RNAi experiments using siRNAs where only the RNA is targeted, leaving the LUAD-amp-1 amplified loci intact. We used three different siRNAs targeting LUAD-amp-1 mapping to exons A, C as presented in (Figure 14a). siRNAs 1 and 2 reduced the levels of LUAD-amp-1 more than 50%. However, siRNA 3 mapping to exon A was not able to reduce the expression of the other exons. These results are consistent with the previous evidence where no signal from the junction A-C was detected possibly due to a lack of connection between these two exons among the isoforms (Figure 14b).

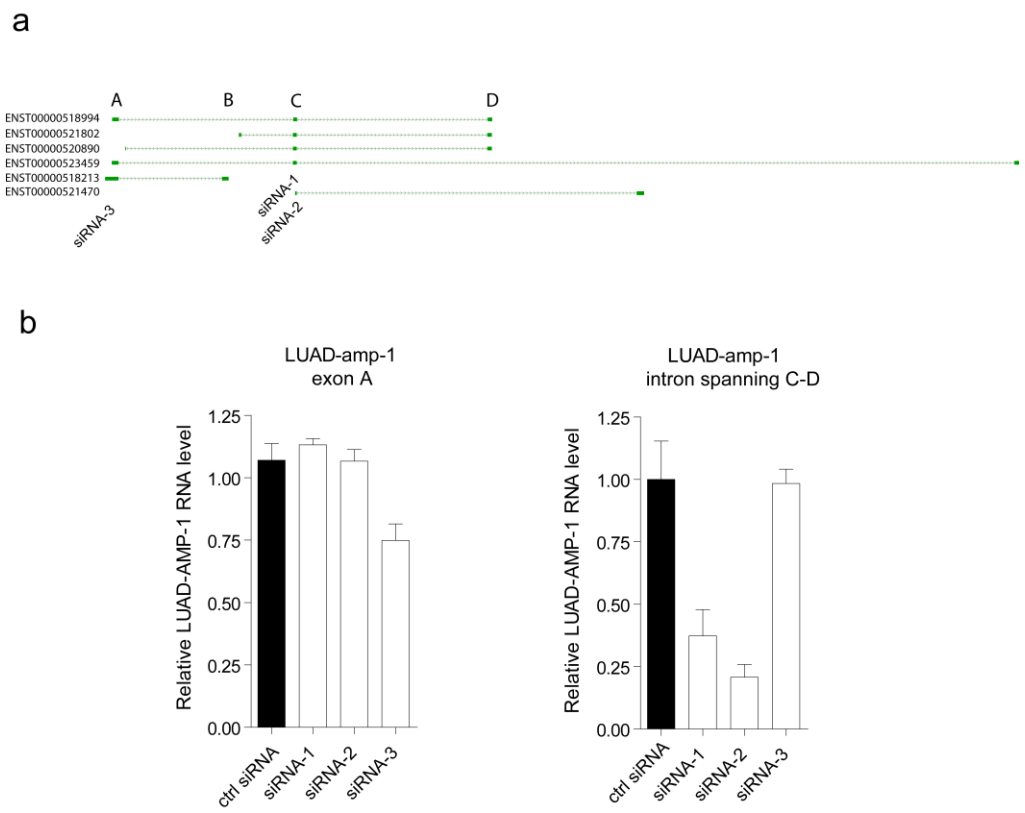

Figure 14. LUAD-amp-1 inhibition using RNAi strategy. (a) Schematic representation of the location where siRNAs (1, 2 and 3) target LUAD-amp-1. (b) RNA levels of LUAD-amp-1 after siRNA transfection. LUAD-amp-1 levels detected using primers recognizing exon A and the junction between exon C and D. LUAD-amp-1 expression is normalized to HPRT and represented relative to a scramble siRNA used as control (ctrl siRNA). Graphs show mean ± SD of two independent experiments.

Next, to investigate the effects of LUAD-amp-1 inhibition by siRNA, we performed cell proliferation assays (MTS) in H1648 and HCC95 cells. Proliferation was monitored starting twenty-four hours post-transfection up until six days. LUAD-AMP-1 inhibition significantly reduced cell proliferation compared to control siRNA. Similar results were obtained in both cell lines (Figure 15 a, b).

Next, clonogenic assays were carried out to test the capacity of individual cancer cells to proliferate forming colonies, a characteristic associated with tumorigenicity. As shown in Figure 15c colony formation was hampered after LUAD-amp-1 inhibition. Taken together these results indicate that LUAD-amp-1 downregulation inhibits the oncogenic phenotype (excessive cell proliferation) of lung cancer cell lines *in vitro.*

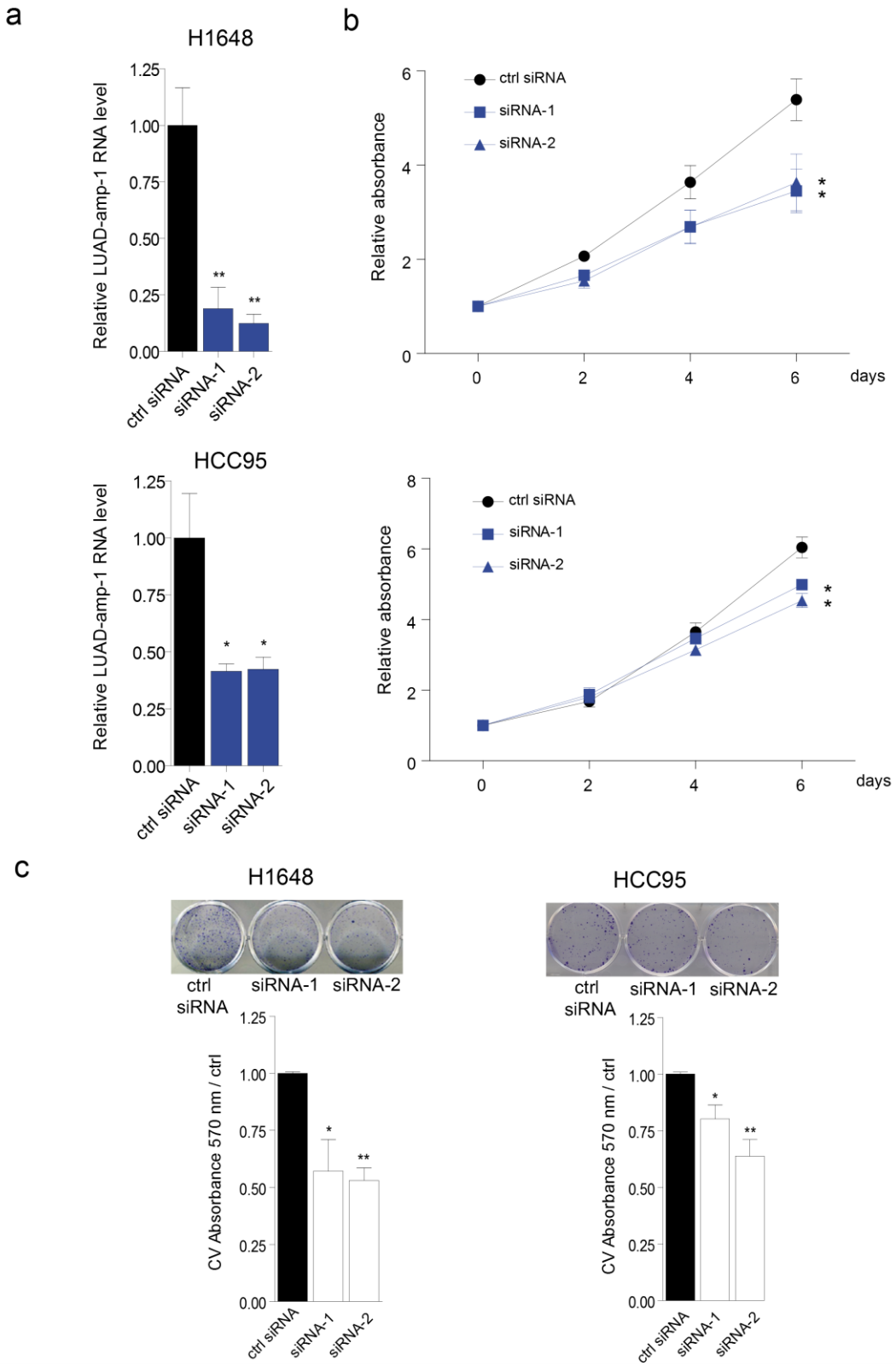

Figure 15. LUAD-amp-1 inhibition impairs cell proliferation and colony formation. (a) Inhibition levels of LUAD-amp-1 obtained with siRNA 1 and 2 in H1648 and HCC95 cells; data is normalized to HPRT and relative to control siRNA (ctrl). (b) Cell proliferation of HCC95 and H1648 cells after LUAD-amp-1 inhibition; data is represented as the relative absorbance obtained with the MTS reagent. (c) Clonogenic assay of HCC95 and H1648 cells where LUAD-amp-1 was inhibited; CV absorbance values are represented relative to the ctrl. Graphs show mean ± SD of three independent experiments. Statistical significance is represented as (*) p-value ≤ 0.05, (**) p-value ≤ 0.01.

3.3.3 LUAD-amp-1 inhibition promotes cellular apoptosis

In order to explore additional cancer phenotypes where LUAD-amp-1 could be involved we performed Annexin-V/7-AAD detection assays to quantify the levels of cellular apoptosis. Cells where LUAD-amp-1 was inhibited with siRNA showed a higher percentage of apoptotic cells compared to the cells transfected with the nontargeting siRNA. To confirm this observation we performed additional Annexin-V/7-AAD experiments using the CRISPR/Cas9 clones, where LUAD-amp-1 loci were targeted and its expression was reduced. Likewise, clones 23 and 24 with low RNA levels of LUAD-amp-1 showed a significantly higher percentage of apoptotic cells (Figure 16). These results show that LUAD-amp-1 is involved in the control of apoptosis in lung cancer cell lines.

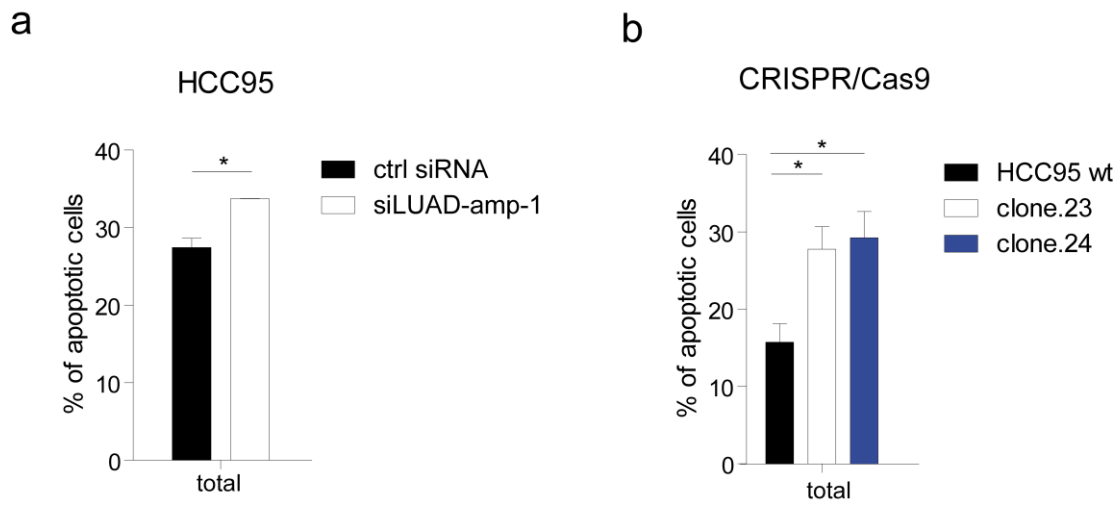

Figure 16. LUAD-amp-1 inhibition promotes cellular apoptosis. (a) Apoptosis of HCC95 and transfected with siRNAs against LUAD-amp-1 or with the control siRNA. Apoptosis was assessed with Annexin V/7-AAD. The percentage of Annexin V and 7-AAD positive cells is represented. (b) Apoptosis levels of the CRISPR/Cas9 clones where LUAD-amp-1 expression was reduced. Graphs show mean ± SD of two independent experiments. Statistical significance is represented as (*) p-value ≤ 0.05.

3.3.4 LUAD-amp-1 promotes the oncogenic phenotype of lung cancer cells *in vitro*

To further evaluate the oncogenic features of the lncRNA, we performed the reciprocal experiments by overexpressing LUAD-amp-1 using the isoform ENST00000521802 conformed by 3 exons (B, C, D) of a total length of 415 nt shown in Figure 17a. The cDNA was cloned in a plasmid and transiently transfected into cells. Even though high overexpression levels of LUAD-amp-1 were achieved with transient transfection in H1648 and HCC95 cells (8000 fold and 30 fold respectively, Figure 17b), no significant effects on cell proliferation were observed in the MTS experiments (Figure 17c). However, LUAD-amp-1 overexpression significantly increased the clonogenic capacity of both cell lines (Figure 17d). The results of these experiments demonstrate the capacity of LUAD-amp-1 to promote colony formation. However, we reasoned that the limited effect of LUAD-amp-1 overexpression observed in MTS could be due to the already saturating levels of LUAD-amp-1 present in these cells. To test this hypothesis we used a different strategy.

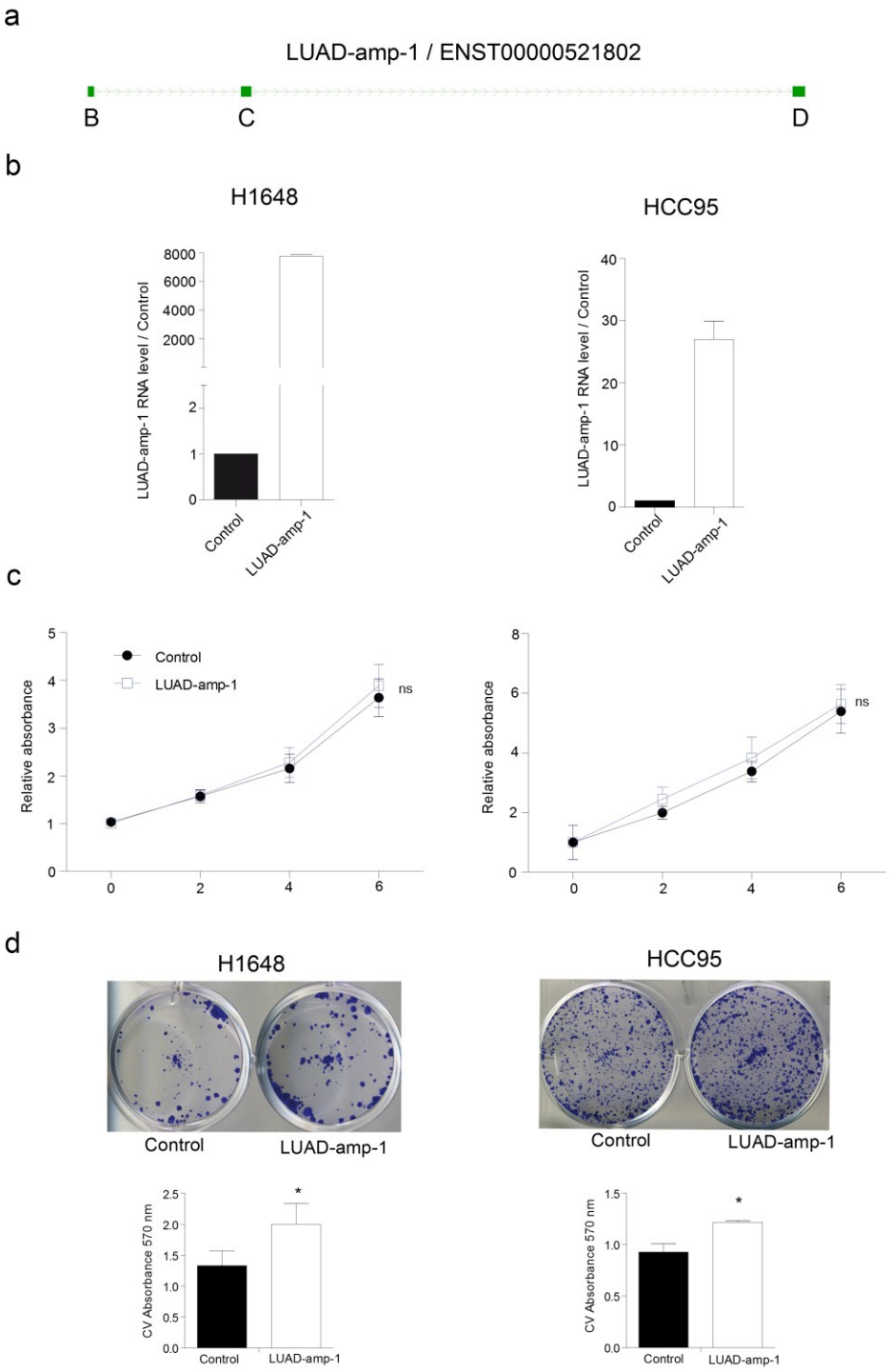

Figure 17. LUAD-amp-1 overexpression promotes colony formation. (a) LUAD-amp-1 isoform used for overexpression experiments. (b) LUAD-amp-1 levels in H1648 and HCC95 cells transfected with the pcDNA3-LUAD-amp-1 plasmid; the empty vector pcDNA3 was used as control. LUAD-amp-1 expression is normalized to HPRT and represented relative to the empty vector. (c) Cell proliferation after overexpressing LUAD-amp-1, measured by MTS (d) Clonogenic assay of H1648 and HCC95 cells transiently overexpressing LUAD-amp-1. Graphs show mean ± SD of three independent experiments. Statistical significance is represented as (*) p-value ≤ 0.05.

We used A549 cells, where LUAD-amp-1 basal expression is very low (Figure 18a). We then stably overexpressed LUAD-amp-1 (600 fold) in A549 cells using retroviral transduction, resulting in a significant increase in cell proliferation and colony formation (Figure 18 b, c). Altogether, these experiments demonstrate that LUAD-amp-1 increases the *in vitro* oncogenic phenotype of cancer cells.

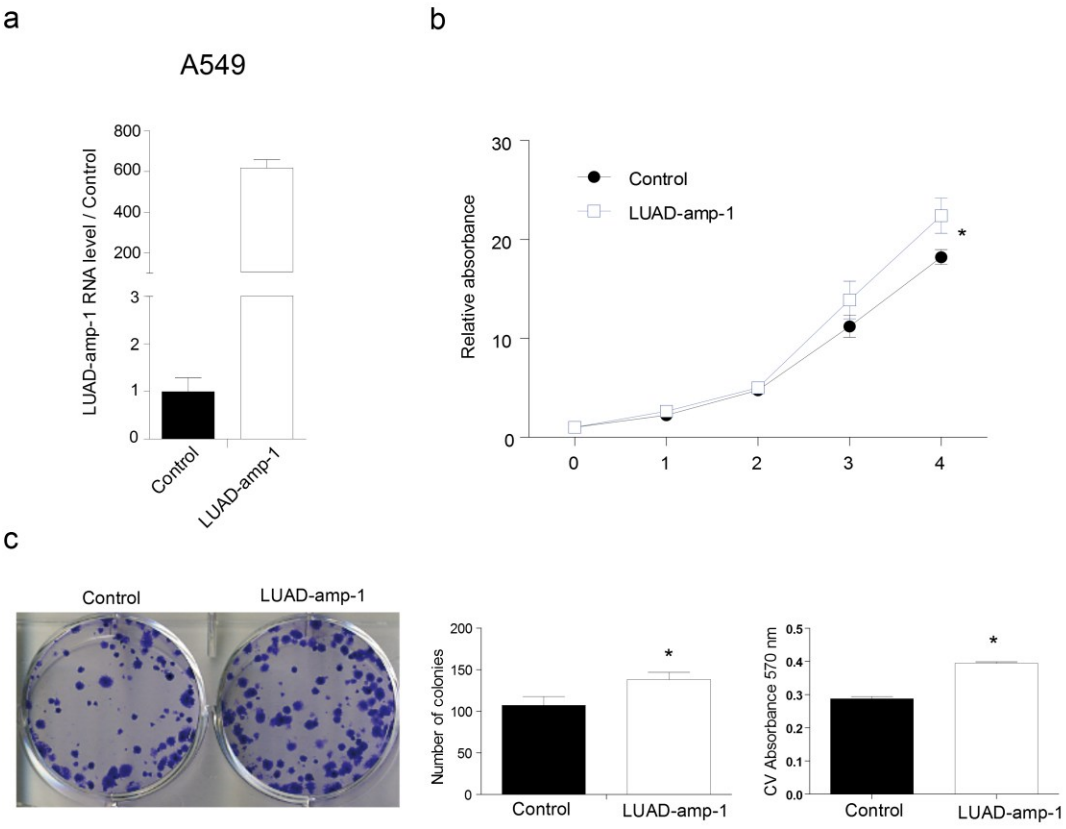

Figure 18. LUAD-amp-1 stable overexpression promotes cell proliferation and colony formation. (a) RNA levels of A549 stable cells overexpressing LUAD-amp-1, data is represented relative to the levels of the control cell line, transduced with virus carrying the empty vector. (b) Cell proliferation of A549 overexpressing LUAD-amp-1, measured by MTS. (c) Colony formation experiment of A549 cells with LUAD-amp-1 overexpression, data is represented as number of colonies formed and as the crystal violet absorbance obtained after distaining. Graphs show mean ± SD of three independent experiments. Statistical significance is represented as (*) p-value ≤ 0.05.

3.3.5 LUAD-amp-1 inhibition reduces *in vivo* tumor formation

We next investigated the effects of LUAD-amp-1 inhibition *in vivo* by using subcutaneous xenograft mouse models engrafted with HCC95 cells transfected either with siRNAs targeting LUAD-amp-1 or control siRNA (Figure 19a). Measurements of tumor size for a period of 70 days after injection demonstrated that LUAD-amp-1 has a strong impact on *in vivo* tumor growth. Consistent with the previous *in vitro* results, LUAD-amp-1 inhibition resulted in a smaller tumor size compared to the control (Figure 19b, c).

As a second strategy to confirm LUAD-amp-1 function *in vivo* we used the HCC95 cell lines previously engineered with the CRISPR/Cas9 system where LUAD-amp-1 expression was reduced (Figure 19d) and which also showed a decrease in cell proliferation in *in vitro* experiments. The two CRISPR/Cas9 clones (clone 23 and 24) with less copy number and expression of LUAD-amp-1 were injected subcutaneously in the flank of BALB/c Rag2-/- mice, as an additional control a third group of mice was injected with HCC95 parental cells where LUAD-amp-1 locus was not modified. Results of this experiment demonstrate that down regulation of LUAD-amp-1 reduces tumor growth (Figure 19e). The use of these two different but complementary inhibition strategies (siRNAs and CRISPR/Cas9) helped us narrow down to LUAD-amp-1 as a functional lncRNA, which has protumoral effects *in vitro* and *in vivo*.

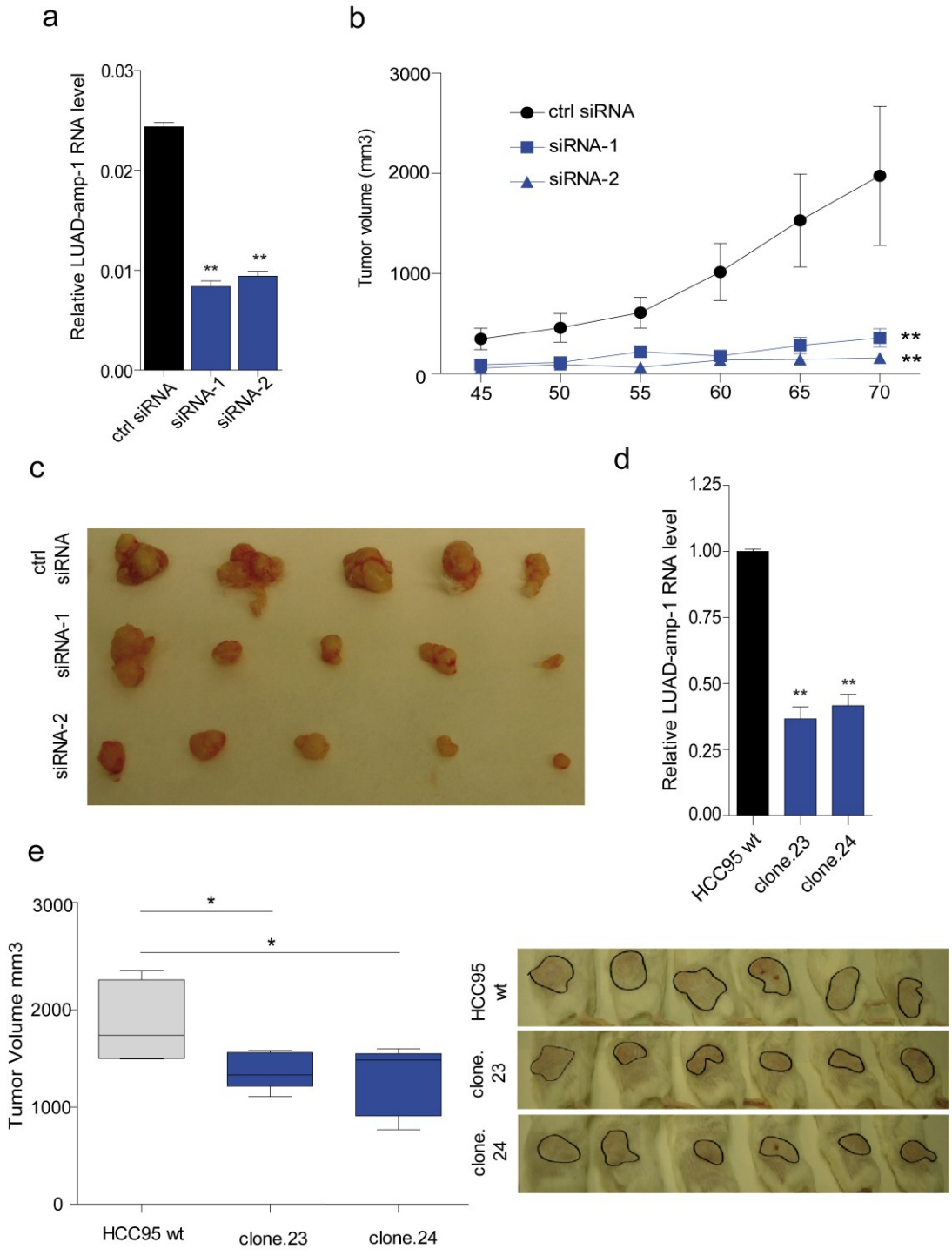

Figure 19. LUAD-amp-1 inhibition reduces tumor formation *in vivo*. (a) LUAD-amp-1 levels in HCC95 cells after inhibition with siRNAs. RNA levels are represented normalized to HPRT. (b) Size of the tumors formed by subcutaneously injecting HCC95 cells with LUAD-amp-1 inhibitions in eight-week-old female BALB/c Rag2-/- mice, tumor sizes were measured every five days. Graph shows mean ± SEM of n=5 per experimental condition. Statistical significance was calculated using one way ANOVA and Bonferroni Multiple Comparison Test comparing to siRNA-Ctrl. (*) p-value ≤ 0.05, (**) p-value ≤ 0.01. (c) Tumors formed by the HCC95 cells with LUAD-amp-1 inhibition. (d) RNA expression levels of LUAD-amp-1 in the CRISPR/Cas9 clones, mean ± SD of three experiments. (e) Size of the tumors generated by injecting HCC95 cells engineered with CRISPR/Cas9 technology. Boxplots shows tumor volumes of n=6 mice per experimental condition at day 45. For statistical analysis t-test was used to compare tumor size values between the HCC95 wt group and the clones.

3.3.6 LUAD-amp-1 affects the expression of genes related to the NF-κB pathway

As a first approximation to characterize the mechanism by which LUAD-amp-1 exerts its oncogenic role, we performed a correlation analysis of gene expression using the lung cancer cell lines available data sets at CCLE (LUAD and LUSC). A list of significantly correlating genes (r >|0.5|) was used to perform a gene annotation enrichment analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) [250]. Gene ontologies such as: regulation of cell proliferation, regulation of response to external stimulus, NF-κB cascade, and cytokine activity among others appeared significantly enriched, suggesting a role of LUAD-amp-1 in these biological processes (BP) and molecular functions (MF) (Figure 20a).

Additionally, to obtain some understanding of the cellular pathways affected upon LUAD-amp-1 loss-of-function, a microarray analysis was performed comparing gene expression in HCC95 cells transfected either with siRNAs specific for LUAD-amp-1 or with siRNA control. LUAD-amp-1 downregulation altered the expression of 236 genes (74 up- and 162 down-regulated (p-value <= 0.01). This list of genes was enriched for gene sets such as inflammatory response, epithelial–mesenchymal transition (EMT), TNF-α signaling via NF-κB, and others (Figure 20b). Interestingly, the NF-κB gene set came out as enriched in both the correlation analysis and the microarray data. Furthermore, a set of genes, all regulated by NF-κB, was also altered upon LUAD-amp-1 inhibition (Figure 20c), and their changes in expression were independently validated by qRT-PCR (Figure 20d). However, LUAD-amp-1 neighbor gene *IKBKB* is not among the genes affected by LUAD-amp-1 inhibition.

The results of these analyses suggest that LUAD-amp-1 could exert its oncogenic function via a mechanism involving the NF-κB pathway.

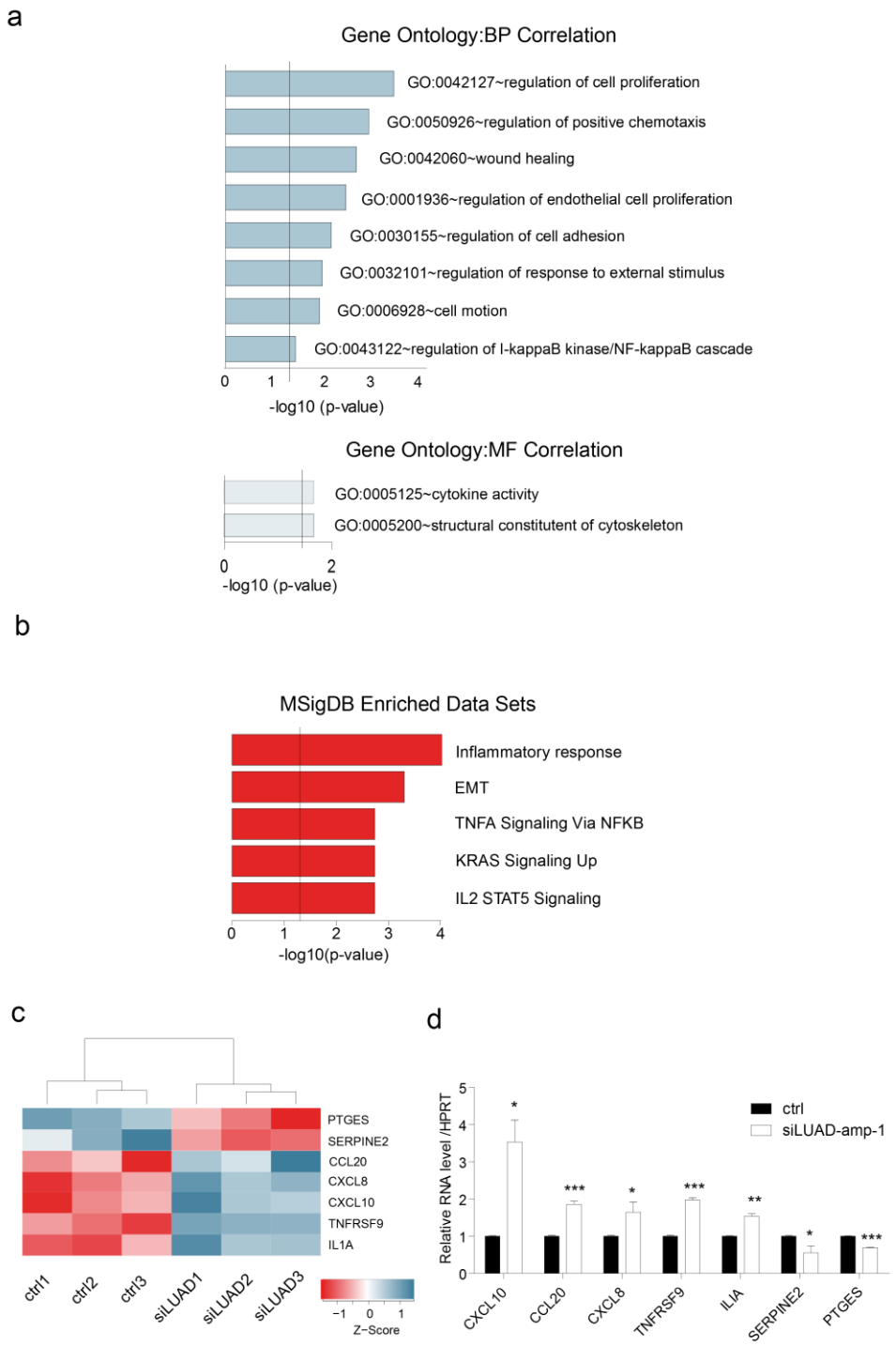

Figure 20. Gene ontologies (GO) of genes correlating with LUAD-amp-1 expression. (a) GO of the genes correlating with LUAD-amp-expression, data was obtained from the CCLE; barplots represent the p-values associated with the enrichment of the identified GOs, grouped by biological process (BP) or molecular function (MF). Threshold line indicates a p-value of 0.05. (b) Enriched data sets obtained with the gene list of differential expressed genes from LUAD-amp-1 inhibition microarray. (c) Heatmap showing the differential expressed genes related to NF-κB in the triplicate samples (3 ctrl and 3 siLUAD1&2). Color code represents upregulated genes in blue, and downregulated genes in red, relative to the control. (d) Validation of the differentially expressed genes after LUAD-amp-1 inhibition using qPCR. Graph show mean ± SD of three independent experiments. Statistical significance is represented as (*) p-value ≤ 0.05, (**) p-value ≤ 0.01, (***) p-value ≤ 0.001.

3.3.7 LUAD-amp-1 is a direct target of p65 and is induced with TNFα

Based on the previous observation that LUAD-amp-1 expression correlates with NF-kB regulated genes, we decided to explore if LUAD-amp-1 was transcriptionally regulated by NF-κB. For this, publicly available ChIP-seq data of the p65/RelA subunit of NF-κB in different cell types (HUVEC GSM1305210 [251], A549 GSE34329 [252], IMR90 GSE43070 [253]) were retrieved. Analysis of the binding sites of NF-κB showed an association of p65/RelA to the LUAD-amp-1 locus (Figure 21a). Moreover, this association was only detected upon TNFα treatment (Figure 21b), similar to what happens with other NF-κB target genes (*IL8, BCL3*) (Figure 21c).

In addition, the analysis of the genomic sequences underlying the ChIP-seq peaks of LUAD-amp-1 region revealed the presence of the consensus binding sequence 'GGGAATTTCC' of p65 (Figure 21d) confirming that these regions are *bona-fide* NF-κB binding sites.

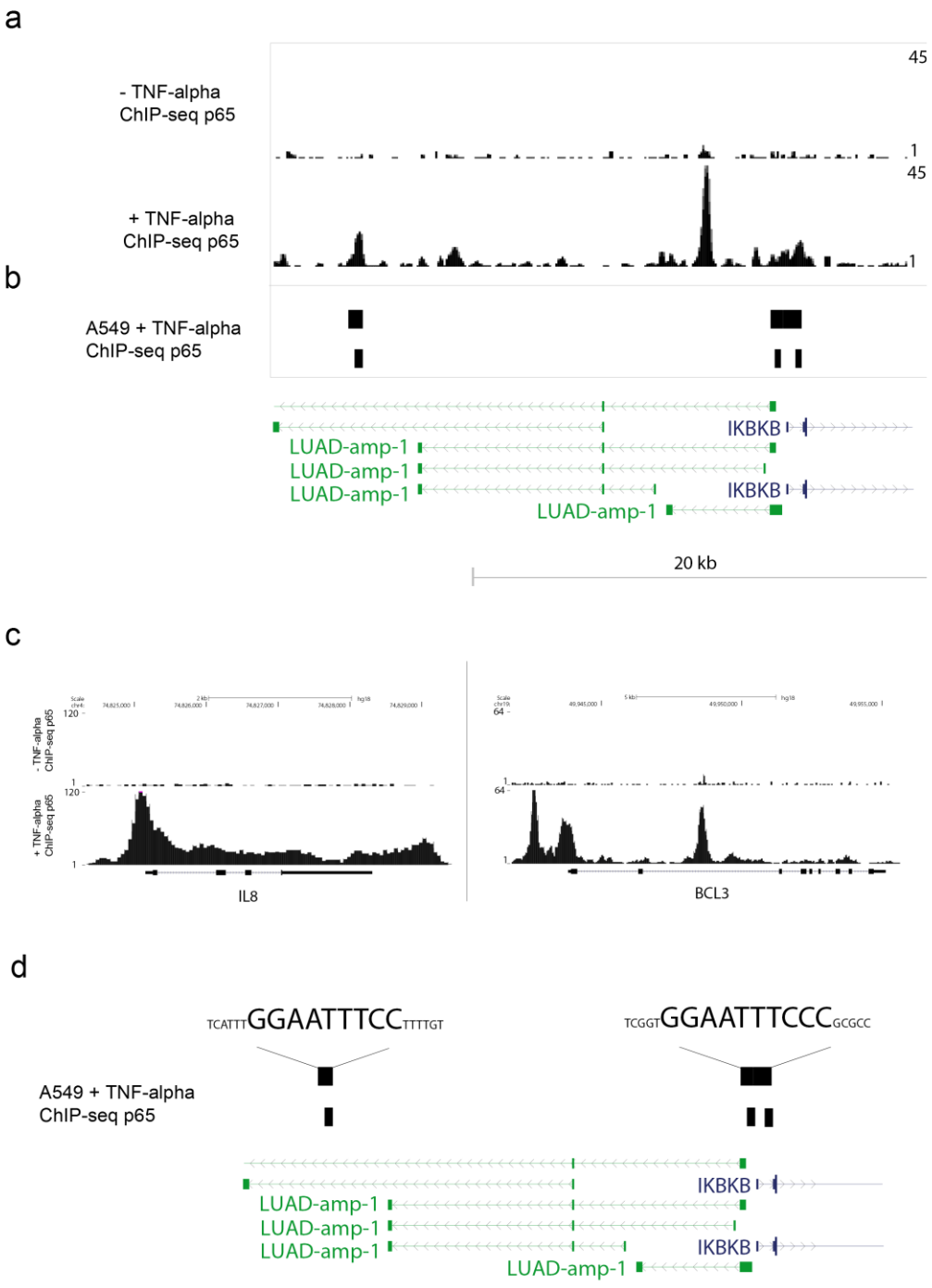

Figure 21. NF-κB/p65 associates to LUAD-amp-1 loci upon TNFα treatment. (a) ChIP-seq peaks of p65 (antibody sc372x), identified in HUVEC cells with and without TNFα treatment (GSM1305210, mapping to LUAD-amp-1 locus (b) Regions enriched for p65 binding in A549 treated with TNFα. The figure was generated loading the available bed files of the GSE34329 data set. (c) ChIP-seq signal showing the binding of p65 to two of its targets (*IL8/CXCL8* and *BCL3*). (d) Consensus sequence corresponding to p65 binding site identified near LUAD-amp-1 region.

The TNFα-dependent binding of p65/RelA to the LUAD-amp-1 loci suggests that LUAD-amp-1 expression could be regulated by NF-κB. Based on this observation, we examined the transcriptional regulation of LUAD-amp-1 mediated by NF-κB.

To activate the NF-κB pathway above the basal levels, HCC95 cells were treated for 0, 0.5, 1, 2, 4, and 6 hours with TNFα (10 ng/ml) followed by RNA extraction for each time point. Under this experimental setting, LUAD-amp-1 expression was induced, similar to other NF-κB target genes (e.g. *IL8/CXCL8* and *BCL3*) (Figure 22a).

Three different induction patterns have been described for p65 targets: 1) early up regulation after 30 minutes followed by a decrease at 4 hours, 2) slow increase of gene expression up to 4 hours, and 3) rapid induction and high expression is maintained [254]. We observed that LUAD-amp-1 follows the second pattern; it is progressively induced reaching the highest peak of expression at 4 hours (Figure 22b). Similar kinetics have been reported for other p65 targets such as *NK4*, prostaglandin E synthase (*PGES*), and A20 [255]. Moreover, inhibition of p65 reduced the expression of LUAD-amp-1 (Figure 22c). Collectively these results demonstrate that LUAD-amp-1 is a direct target of p65 and it is induced after TNFα treatment.

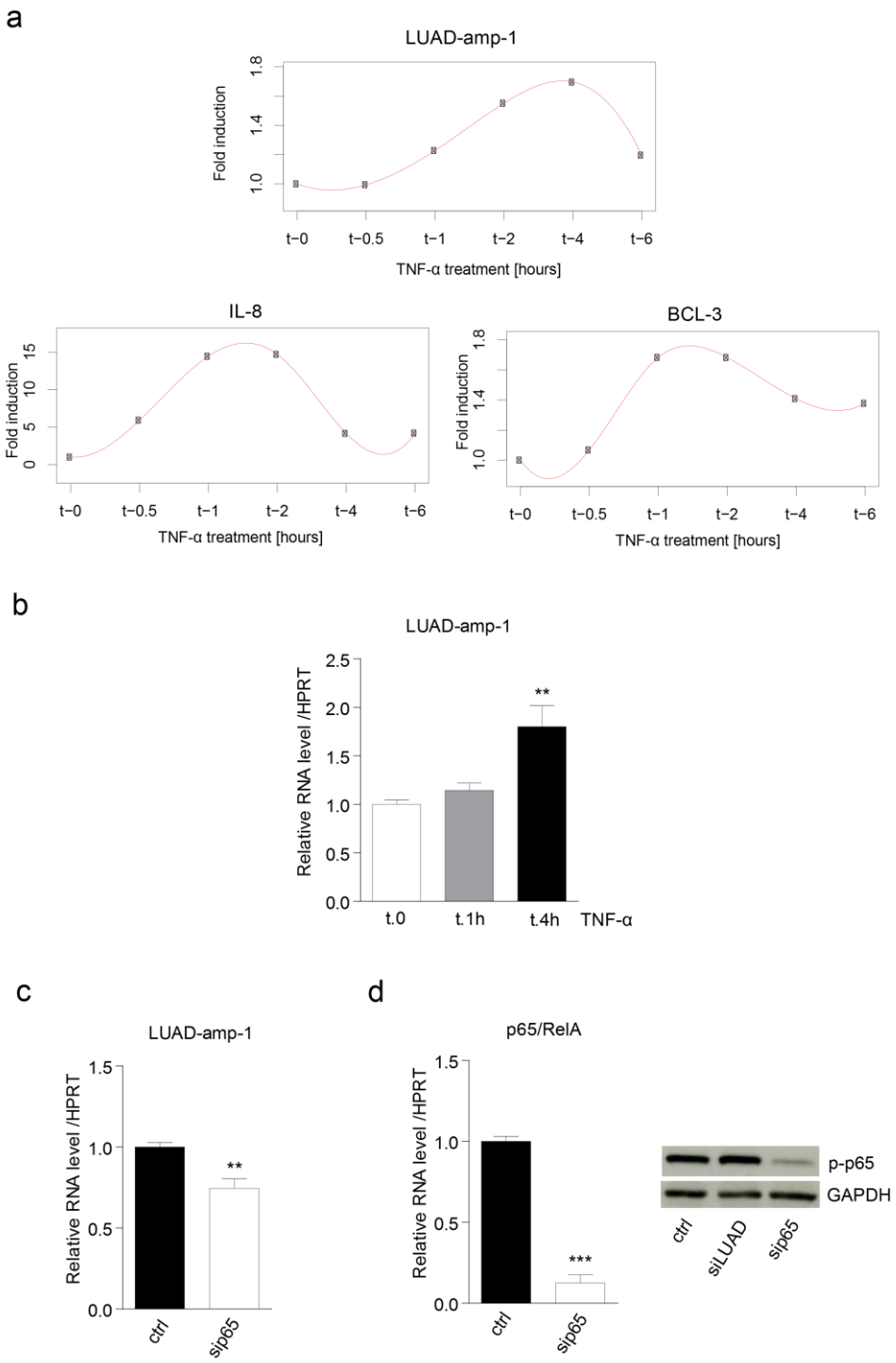

Figure 22. LUAD-amp-1 levels are induced upon TNFα treatment. (a) Time course experiment showing the induction of LUAD-amp-1 and other NF-κB direct targets in HCC95 cells treated with TNFα (10 ng/ml) for the indicated time points. Data was normalized with HPRT and is represented relative to time zero. (b) LUAD-amp-1 expression peaks at 4 hours of TNFα treatment. The plotted values represent the mean of three biological replicates +/-SD. (c) LUAD-amp-1 levels after p65 inhibition in HCC95 cells; expression values are represented relative to the scramble siRNA, used as control. (d) Downregulation of p65 was confirmed at the RNA and protein levels by western blot. Statistical significance is represented as (**) p-value ≤ 0.01, (***) p-value ≤ 0.001.

3.3.8 LUAD-amp-1 regulates the levels of TNFα and other cytokines upon TNFα treatment

Knowing the direct regulation of LUAD-amp-1 by NF-κB, the next question we tried to address was whether LUAD-amp-1 could act as a modulator of the NF-κB pathway. Interestingly, our microarray data showed that LUAD-amp-1 knockdown significantly induced the expression of some cytokines (*CXCL10/IP10*, *IL-8/CXCL8*, and *CCL20)*, proteases (*SERPINE2*) and receptors (*TNFRSF9*), all of them regulated by p65 (Figure 23); so we tested if their induction upon TNFα treatment was modulated by LUAD-amp-1. HCC95 cells where LUAD-amp-1 was inhibited by siRNA transfection were treated for 4 hours with TNFα; time point when LUAD-amp-1 expression peaks. Upon TNFα treatment the expression of this set of genes was induced; moreover this induction was enhanced when LUAD-amp-1 was inhibited indicating that LUAD-amp-1 negatively affects their expression.

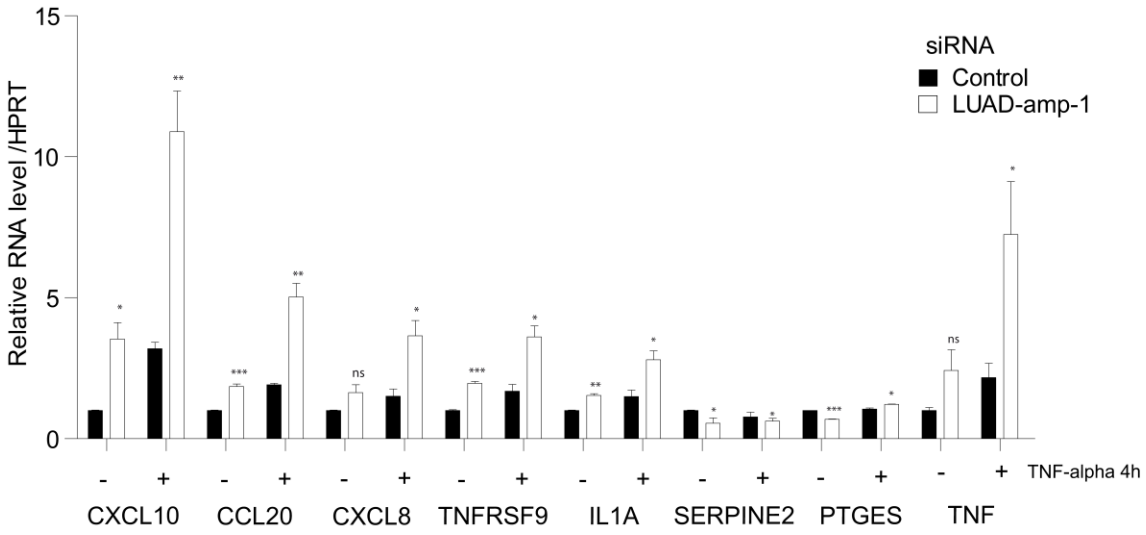

Figure 23. LUAD-amp-1 regulates a set of p65 related genes also induced upon TNF-α treatment. RNA levels of 8 genes altered by LUAD-amp-1 inhibition. The represented conditions include HCC95 cells with/without TNFα treatment and with/without LUAD-amp-1 inhibition. The obtained RNA levels are normalized to HPRT and are represented relative to the control condition (scramble siRNA) with no treatment. Graph shows the mean ± SD of three independent experiments. Statistical significance is represented as (*) p-value ≤ 0.05, (**) p-value ≤ 0.01, (***) p-value ≤ 0.001.

An extra piece of evidence supporting the negative regulation of the NF-κB pathway mediated by LUAD-amp-1 was the observation that TNFα was among the genes induced after LUAD-amp-1 inhibition, suggesting that LUAD-amp-1 could regulate the pathway by controlling the levels of TNFα itself. Even though the differences were not significant in basal conditions; upon TNFα treatment a significant increase was observed. Time course experiments after LUAD-amp-1 inhibition showed an increase of TNFα at all the time points analyzed (Figure 24).

Together, all these data suggest that LUAD-amp-1 besides being a direct target of p65 itself, is also involved in the regulation of a set of p65 target genes, which are induced by TNFα suggesting that LUAD-amp-1 could be acting as a negative regulator of the TNFα / NF-κB signaling pathway.

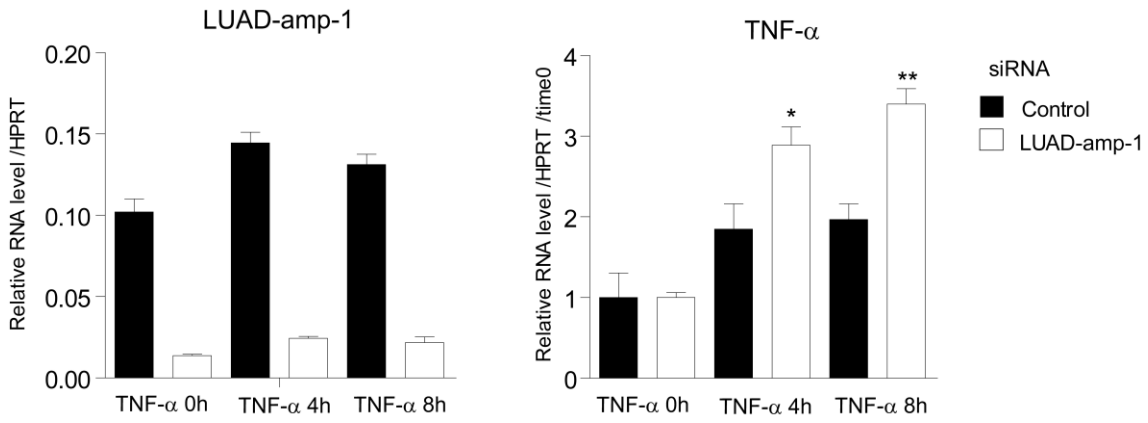

Figure 24. **LUAD-amp-1 regulates TNFα levels**. TNFα induction after LUAD-amp-1 inhibition is increased upon TNFα treatment. Time course experiment of HCC95 cells treated with TNF-α (4 ng/ml) for the indicated time points (0, 4, 8 hours). LUAD-amp-1 expression values are represented normalized to HPRT. Expression levels of TNFα are represented relative to the control siRNA and time 0. Graph shows mean ± SD of two independent experiments. Statistical significance is represented as (*) p-value ≤ 0.05, (**) p-value ≤ 0.01, (***) p-value ≤ 0.001.

3.4 LUAD-amp-1 molecular mechanism

3.4.1 LUAD-amp-1 is a cytoplasmic lncRNA that interacts with SART3

We next set to investigate the mechanism by which LUAD-amp-1 mediates its oncogenic role. First we performed a cell fractionation experiments to identify the subcellular localization of LUAD-amp-1. The results of these experiments revealed that LUAD-amp-1 is a lncRNA mainly localized to the cytoplasm of the cells, similarly or even more cytoplasmic than the protein-coding mRNA HPRT (Figure 25).

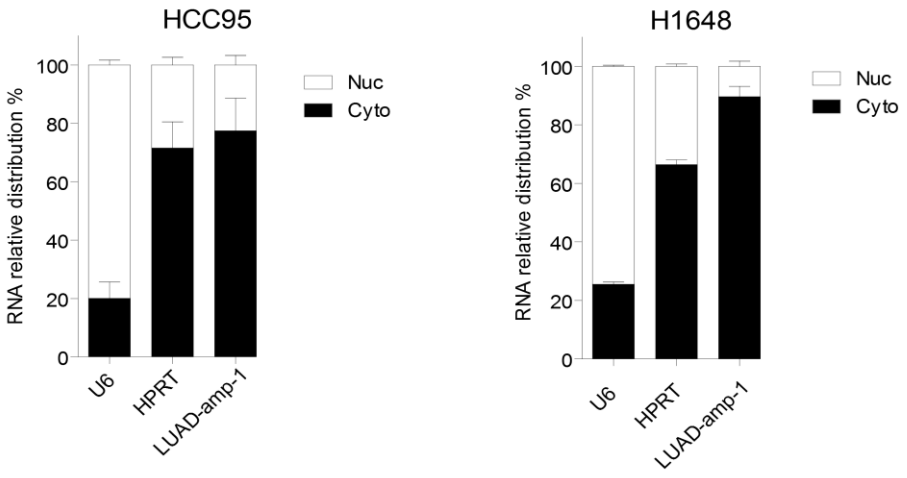

Figure 25. Subcellular localization of LUAD-amp-1. (a) Relative distribution of LUAD-amp-1 in the nucleus and cytoplasm of HCC95 and H1648 cell lines, quantified by qRT-PCR. HPRT and U6 are presented as cytoplasmic and nuclear controls respectively.

To get further insight into LUAD-amp-1 molecular mechanism, we set to identify proteins that interact specifically with the lncRNA and that could function together with it. To that end, we performed an RNA pull-down experiment with *in vitro* transcribed biotinylated RNA using as negative control the antisense sequence of LUAD-amp-1, which has the same length but unrelated sequence. Each of the RNAs was incubated with a fraction of the same protein extracts of HCC95 cells and the proteins retained by the RNAs were run in a polyacrylamide gel.

The differentially observed bands were cut and subjected to mass spectrometry (MS) (Figure 26a). Q15020 (SART3_HUMAN) Squamous cell carcinoma antigen recognized by T cells 3, also referred to as Tip110 (HIV-Tat interacting protein), came out as the protein with the highest number of peptides (21 unique peptides) present only in the LUAD-amp-1 pull-down, which was confirmed in an experimental replicate (11 SART3 unique peptides) (Figure 26b). We then independently validated the results by western blot with an antibody recognizing SART3 (abcam, ab155765). The interaction was confirmed with the endogenous SART3 and also with the myc-tagged version of SART3 (Figure 26c).

We then set to validate the interaction detected between LUAD-amp-1 and SART3 by using a different technique. We used a specific antibody anti-SART3 for RNA immunoprecipitation (RIP) experiment in A549 cells, confirming the interaction of stably expressed LUAD-amp-1 and endogenous SART3 (Figure 26d). We observed an enrichment of LUAD-amp-1 after SART3 immunoprecipitation, compared to the RNA levels obtained with the control IgG. We chose U6, a known binder of SART3, as a positive control and MALAT1 as negative control. Together all these data identifies SART3 as a protein that interacts specifically with LUAD-amp-1.

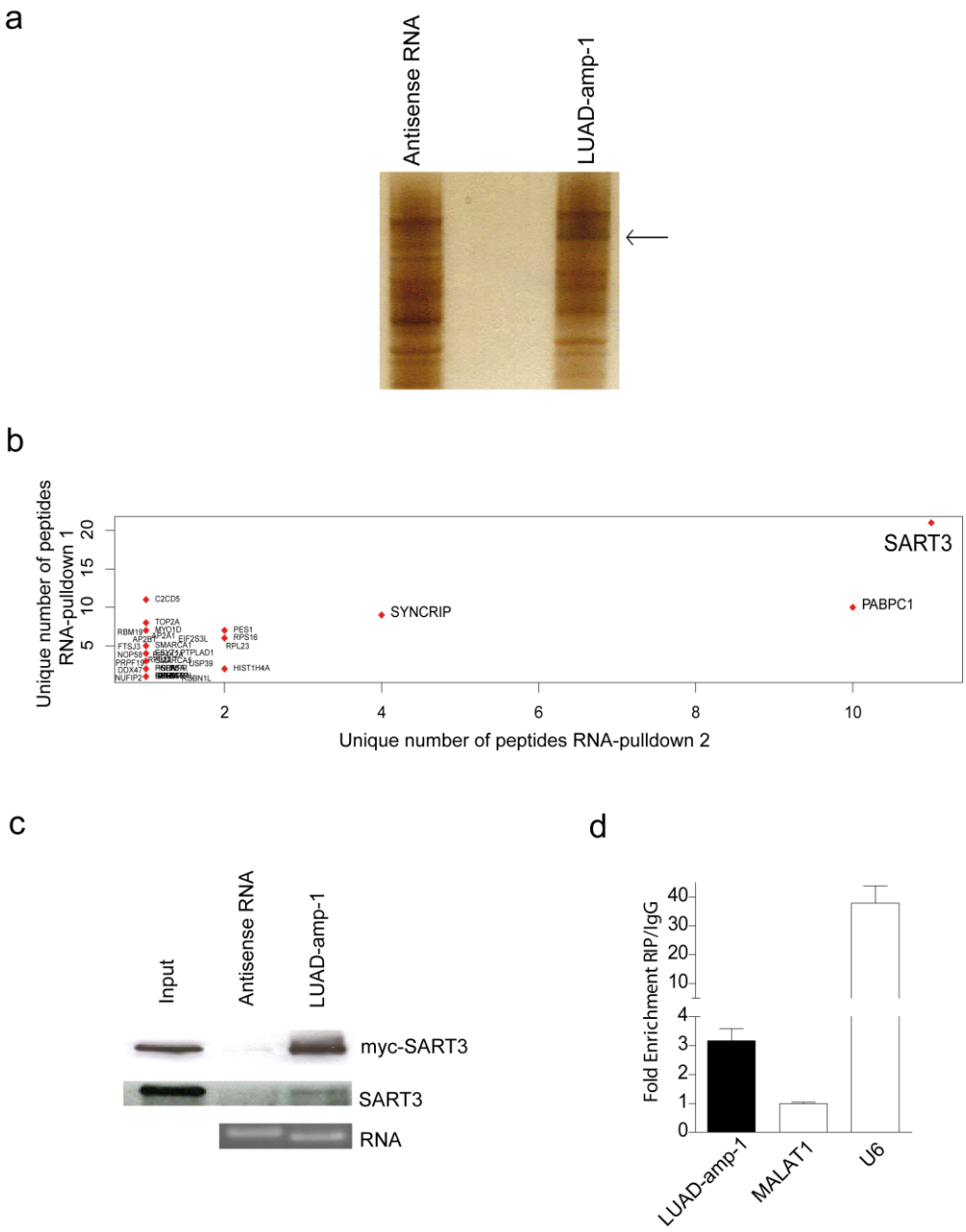

Figure 26. LUAD-amp-1 binds to SART3. (a) Silver stained gel of the proteins bound by the *in-vitro* transcribed antisense (left lane) and LUAD-amp-1 RNA (right lane), the differential band is indicated with an arrow. (b) Dot plot showing the number of peptides for each of the proteins identified in the two RNA-pulldown experiments. The plotted proteins were exclusively identified in the LUAD-amp-1 lane. SART3 came out as the top candidate with 21 and 11 unique peptides. (c) Validation of LUAD-amp-1 and SART3 interaction by western blot. The upper row shows the RNA-pulldown, using a protein extract of HCC95 cells transfected with myc-SART3, blotted with myc-tag antibody; the middle row shows the endogenous SART3 (RNA-pulldown by western blot and the lower row shows the biotinylated RNAs (antisense and LUAD-amp-1) used to pulldown SART3. (d) RNA immunoprecipitation (RIP) of SART3 antibody, followed by qRT-PCR of the bound RNAs. LUAD-amp-1 levels were quantified by qRT-PCR, the obtained values are represented relative to the values obtained with IgG, MALAT1 and U6 were used as positive and negative controls, respectively.

SART3 has been described as a ubiquitously expressed RNA-binding protein involved in several biological processes including pre-mRNA splicing, regulation of gene transcription, and protein localization. It has been widely characterized as a recycling factor during splicing. SART3 interacts with U6 and takes it to the Cajal Bodies for the next round of splicing [16]. SART3 localization has been described as mainly nuclear in tumors, however SART3 can also be found in the cytoplasm of proliferating cells (normal and malignant) [256], [257], [258]. We confirmed the presence of SART3 in both the nucleus and cytoplasm of A549, HCC95 and H1648 cells (Figure 27). These results are consistent with the interaction between LUAD-amp-1, mainly cytoplasmic, and SART3.

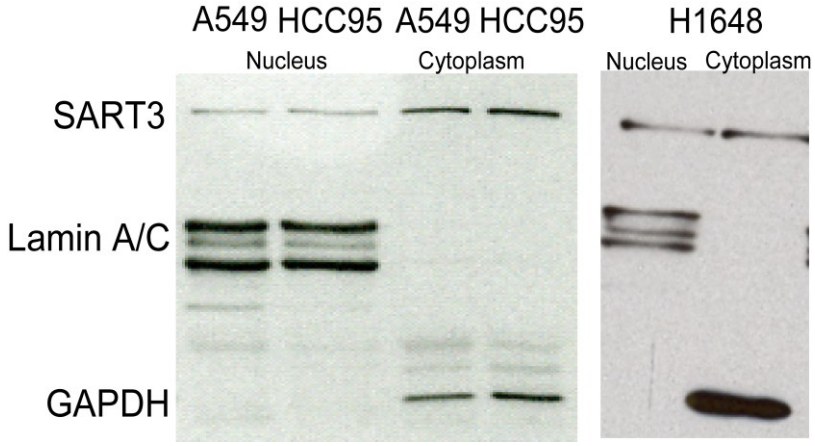

Figure 27. SART3 has a nuclear and cytoplasmic localization. (a) SART3 localization in A549 HCC95 and H1648 cells, identified by western blot using SART3 antibody. Lamina A/C and GAPDH were used as control of nuclear and cytoplasmic fractions respectively.

3.4.2 LUAD-amp-1 regulates SART3 localization

Since one of the mechanisms whereby lncRNAs exert their functions is by regulating protein localization [259], [260], we investigated if LUAD-amp-1 could affect SART3 protein localization. To test this hypothesis we performed immunofluorescence experiments using a myc-tagged version of SART3. Interestingly upon LUAD-amp-1 inhibition, myc-SART3 changed its localization from mostly nuclear to cytoplasmic. Instead of being localized to the nucleus we observed a punctuate pattern in the cytoplasm corresponding to SART3 (Figure 28a). The quantification of the cytoplasmic foci showed a

significant increase in their number in the LUAD-amp-1 knockdown condition compared to the controls (Figure 28b). Together these results suggest that LUAD-amp-1 molecular mechanism involves not only the binding to SART3, but also the regulation of its localization

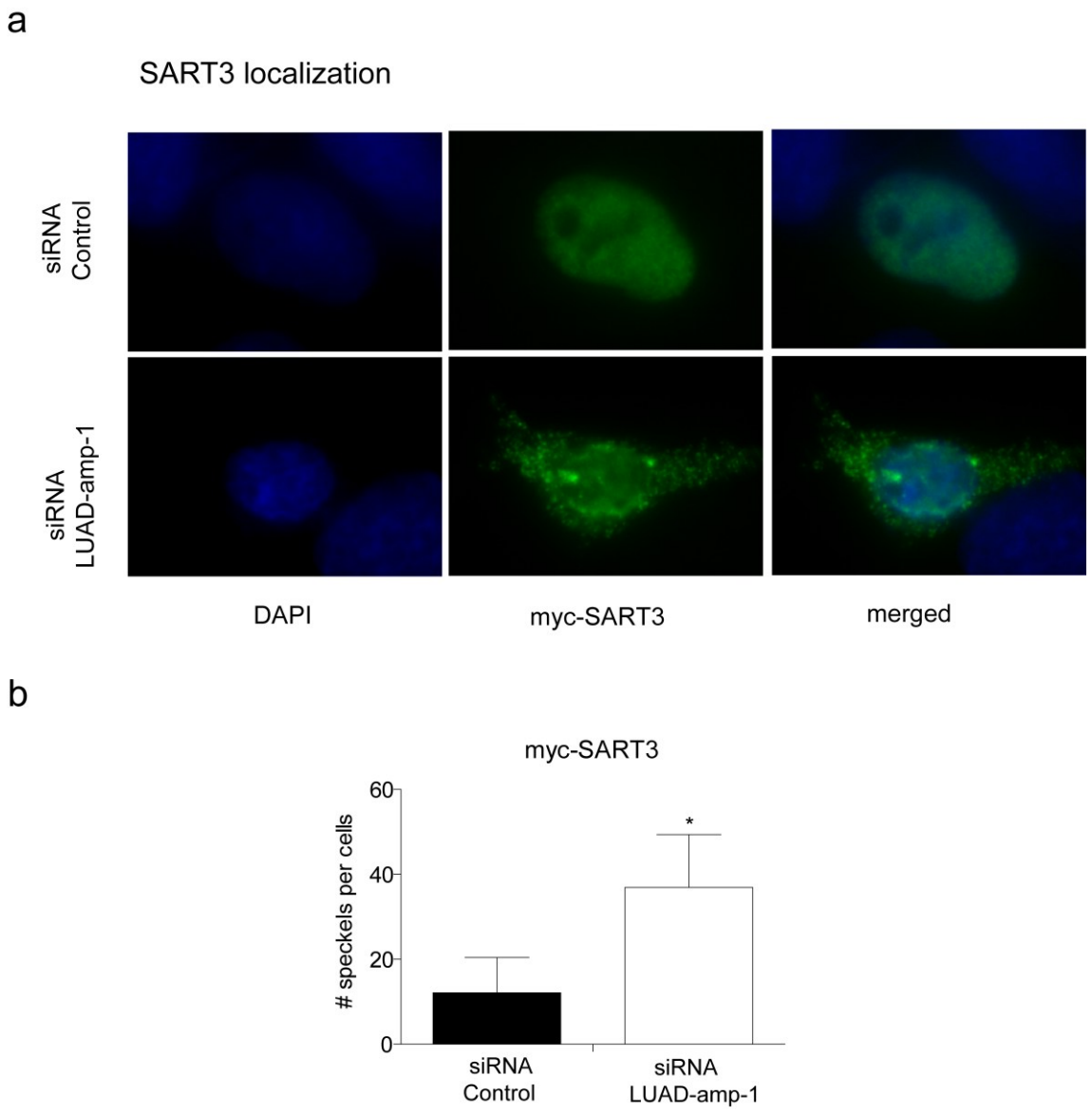

Figure 28. LUAD-amp-1 alters SART3 localization. (a) HCC95 cells transfected with myc-SART3 expressing plasmid, and either with the control siRNAs or siRNAs targeting LUAD-amp-1 were fixed with 4% PFA and stained for immunofluorescence using myc-tag antibody (green). DNA was stained using DAPI (blue). (b) The number of cytoplasmic dots in each of the two conditions was quantified with cell profiler software. The presented barplot shows the number of dots per cells (mean ± SD) obtained after quantifying a total of 20 images per condition, from three biological replicates. Statistical significance is represented as (*) for p-value ≤ 0.05.

3.4.3 LUAD-amp-1 regulates USP4 translocation to the nucleus

The interaction between LUAD-amp-1 and SART3 together with the effect observed in SART3 localization upon LUAD-amp-1 inhibition, led us to focus on the roles that have been described for SART3. Among them, the direct interaction between SART3 and the ubiquitin specific protease USP4 has been demonstrated with structural data. Moreover, we found SART3-USP4 interaction worth to explore because several publications demonstrate that USP4 is involved in the regulation of the NF-κB pathway [261], [262], [263]. USP4 downregulates NF-κB activity by deubiquitinating several proteins which are polyubiquitinated upon TNFα treatment (*RIP1, TAK1* and *TRAF2)*.

It has been reported that the SART3 and USP4 interaction results in the translocation of USP4 from the cytoplasm to the nucleus. Due to the fact that USP4 is not able to get into the nucleus because it lacks a nuclear localization signal (NLS) [264], the interaction between these two proteins is critical for USP4 function in the cell. We sought to investigate if the identified interaction of LUAD-amp-1 with SART3 could have a role on USP4 nuclear translocation. To do so, HCC95 cells were transfected with myc-USP4 and SART3 expressing plasmids, in the presence and absence of LUAD-amp-1. USP4 subcellular localization was then detected using the myc-tag antibody (Figure 29a). As previously described by others [264], the overexpression of SART3 resulted in nuclear translocation of USP4. However, this only occurred in the control cells where LUAD-amp-1 was present. In contrast, USP4 translocation was prevented after LUAD-amp-1 inhibition. This result is consistent with the previously observed relocalization of SART3 in the cytoplasm after LUAD-amp-1 inhibition (Figure 29b).

The presented data supports a mechanism where LUAD-amp-1 is necessary for the nuclear translocation of USP4 mediated by SART3. Since USP4 has been described as a negative regulator of the NF-κB pathway, it remains to be explored if the change in localization of USP4, where LUAD-amp-1 is implicated, has an effect on the pathway activity.

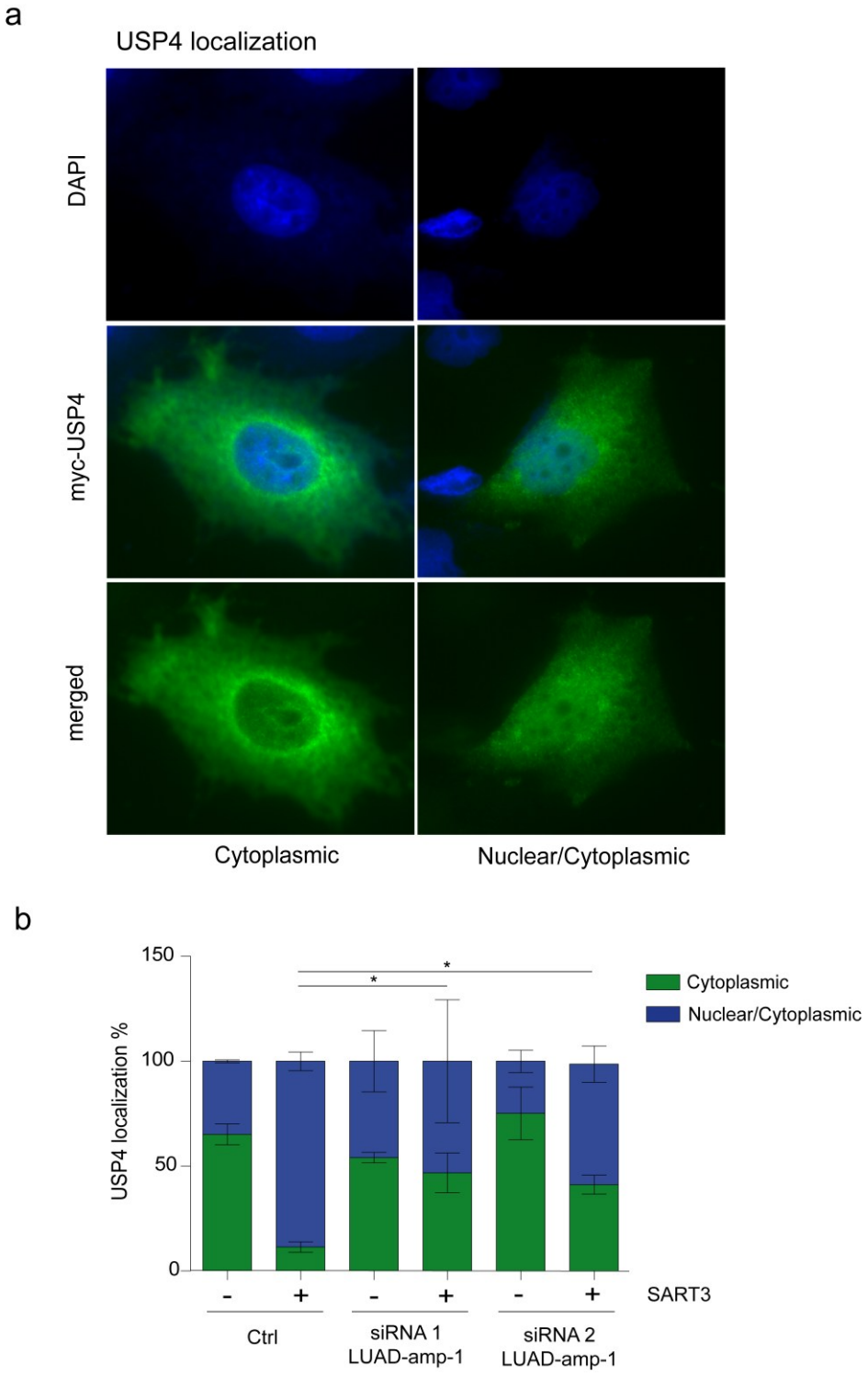

Figure 29. LUAD-amp-1 alters USP4 translocation mediated by SART3. (a) HCC95 cells transfected with the myc-USP4 expressing plasmid. The immunofluorescence images show the localization of myc-USP4 either distributed mainly in the cytoplasm (cytoplasmic) or localized also in the nucleus (nuclear/cytoplasmic). myc-USP4 is identified in green and DNA in blue. (b) Quantification of cells having a cytoplasmic (green) or nuclear/cytoplasmic (blue) localization of myc-USP4, in different conditions: with or without SART3 overexpression and with or without LUAD-amp-1 inhibition. The indicated numbers represent the percentage of cells (mean ± SD) obtained by quantifying 15 images per condition, from three biological replicates. Statistical significance is represented as (*) for p-value ≤ 0.05.

4. DISCUSSION

4.1 SCNAs harbor functional lncRNAs

The identification of cancer-associated genes has been mainly focused in analyzing the genetic alterations that target protein-coding genes. However, restricting only to protein-coding genes misses the opportunity to discover new cancer driver genes that reside in the noncoding part of the genome. Consistently, increasing evidences demonstrate that lncRNAs are implicated in cancer, acting as oncogenes or tumor suppressors [265] .Taking advantage of the copy-number profiles of tumor samples generated by TCGA, we analyzed the focal peaks of the SCNA focusing on those alterations mapping to the noncoding genome. With our analysis we identified a list of lncRNAs which copy number is altered in tumor samples, suggesting that they could have a role in cancer. Furthermore, the use of complementary gene annotations (Gencode and *MiTranscriptome*) allowed us to identified 13 SCNAs harboring recently annotated genes that await functional validation.

To gain further insight into the role that copy number altered lncRNAs could have we performed two computational analyses using the transcription factors and chromatin states associated with them. The enrichment of transcription factors such as MYC, MAX and JUND in the amplified lncRNAs set suggests an oncogenic regulation in line with their amplified status. For the case of the deleted lncRNAs not many transcription factors were enriched, this could reflect either that the deleted lncRNAs are not causally involved in cancer or that the frequent deletions are due to specific features of the region (e.g. chromosomal fragile sites, CFS) independent of the genes harbored inside them.

The analysis of the chromatin states associated with the copy number altered regions (protein coding and lncRNAs) identified an enrichment of enhancers. Integrating these results with additional datasets of lncRNAs bound to the chromatin could help identify a subset of copy number altered lncRNAs with enhancer like function (eRNAs). Nevertheless, the characterized lncRNA of our work LUAD-amp-1 is not associated to the chromatin and acts through a different mechanism. On the other hand, we found an enrichment of insulator regions in the deleted lncRNAs group. Interestingly this enrichment was observed exclusively in the lncRNA but no in the proteins, suggesting that the presence of lncRNAs could be related to the insulators functions of demarcating different chromatin domains. Since insulator regions are bound by the protein CTCF,

future experiments could help identify if the molecular mechanism of the some of these deleted lncRNAs involves CTCF binding.

Even though we used comprehensive annotations (Gencode and *MiTranscriptome*) there were still SCNAs lacking an annotated gene, suggesting the presence of regulatory elements inside them. In fact, other group published that some of these regions do harbor regulatory elements termed superenhancers, which are amplified in several cancer types [242]. Some of the identified superenhancers are associated with *MYC* promoter regulating its expression. These results support the validity of our approach to discover cancer driver events.

It is important to mention that our analysis takes into account exclusively the SCNA from the TCGA data. It could be interesting to use a similar approach to analyze additional tumor samples and defined the alterations shared among the different cohorts. Furthermore at the time when the TCGA data was downloaded, the analysis of several tumor types was still ongoing or restricted this is why only 25 tumor types were used. An update using the complete TCGA data release could lead to the identification of novel copy number alterations and lncRNAs involved in the disease.

4.2 Discovery of lncRNA as new cancer players

Integrative analysis of the lncRNAs inside the SCNAs in conjunction with differential expression data in tumor samples narrowed down the list to 20 putatively functional lncRNAs targeted by SCNAs whose expression is altered in cancer. Furthemore, the reported list harbors some lncRNAs with previously described roles in cancer (*PVT1* and *CASC8*), supporting the reliability of our analysis.

We focused on the characterization of LUAD-amp-1, however experimental validation of the additional candidates using a large phenotypic screening coupled to different inhibition strategies (shRNAs, ASOs, CRISPR/Cas9) could lead to the identification of more functional candidates. Similar approaches have been applied by others resulting in the identification of lncRNAs involved in ovarian cancer (*FAL1* [140], *OVAL1* [266]) and melanoma (*SAMMSON* [141]).

Interestingly, all these analysis [140], [266], [141] and also ours, focus exclusively on the role of amplified lncRNAs that can be acting as oncogenes, leaving the deleted candidates aside. This could be due to the fact that the functional role is easier to define for oncogenes that are found active in cancer cell lines allowing the experimental validation using loss of function experiments. Moreover, oncogenes whose inhibition can lead to tumor cell death are of great clinical interest as novel targets for cancer therapies. Nonetheless, we consider that the experimental validation of the deleted lncRNAs is worth exploring, since it could lead to the discovery of new mechanisms involved in the control of normal cell growth. Current CRISPR/Cas9 technologies could be used to explore the role of particular chromosomal deletions in cancer development by generating cells with chromosomal alterations that mimic human cancers.

4.3 LUAD-amp-1 a novel oncogenic lncRNA

Our SCNA analysis in combination with experimental validations identified a copy number altered lncRNAs that we named LUAD-amp-1; showing that indeed the lncRNA pinpointed by the focal peak of an amplified region is involved in lung cancer progression. Even though the focal peak maps exclusively to LUAD-amp-1; we cannot exclude that its neighbor genes (*PLAT* and *IKBKB*) are also co-amplified; however we consider that our experimental data is quite compelling in demonstrating the oncogenic role of LUAD-amp-1 independent of its neighbor genes. Furthermore, inhibition of LUAD-amp-1 did not have an effect on the expression levels of *PLAT* or *IKBKB* (Figure 1); supporting that LUAD-amp-1 exerts its function through a different mechanism independent of its neighbor genes.

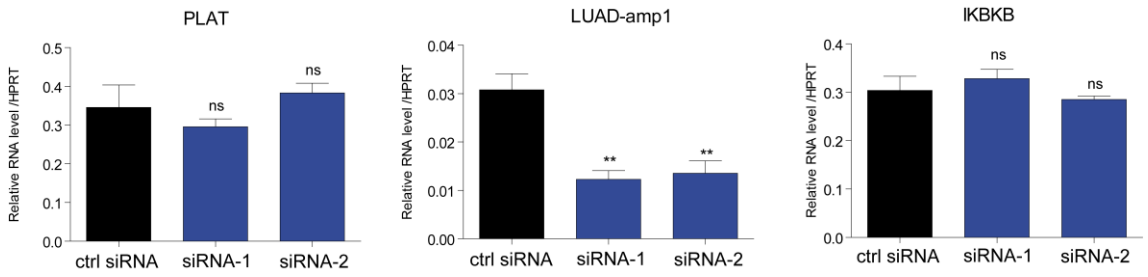

Figure 1. LUAD-amp-1 inhibition has no effect on its neighbor genes *PLAT* and *IKBKB*. HCC95 cells transfected with control siRNA or siRNAs 1, 2 targeting LUAD-amp-1. The expression levels of LUAD-amp-1 neighbor genes *PLAT* and *IKBKB* was assessed by qPCR. The represented graphs show the mean ± SD of two independent experiments.

With our analysis we identified LUAD-amp-1 as an oncogenic lncRNAs, amplified in lung cancer, and whose overexpression is associated with its amplification specifically in lung adenocarcinoma (LUAD). In addition, LUAD-amp-1 is overexpressed in lung squamous (LUSC), we propose that these is due to a different mechanism involving hypomethylation of its promoter region; suggesting that overexpression of LUAD-amp-1 could be achieved by different mechanisms depending on the tumor type.

Since LUAD-amp-1 is expressed in LUSC we explored if its expression could serve as a classifier of the four LUSC subtypes (primitive, classical, secretory and basal). These four subytpes were defined using mRNA expression profiles and are associated to different survival outcomes and biological processes [267]. We found that LUAD-amp-1 has a higher expression in the basal subtype (Figure 2). This subytpe is associated with cell adhesion processes and has a similar survival outcome compared to the classical and secretory subtypes. Further validation of LUAD-amp-1 specific expression in additional tumor cohorts could reinforce the possible use of LUAD-amp-1 as a subytpe stratifier and prognostic biomarker. As previously mentioned lncRNAs are expressed in a more tissue specific manner compared to protein coding genes. It could be interesting to explore which lncRNAs are differentially expressed in each of the LUSC subtypes, to further use them to stratify patients for a more precise prognosis.

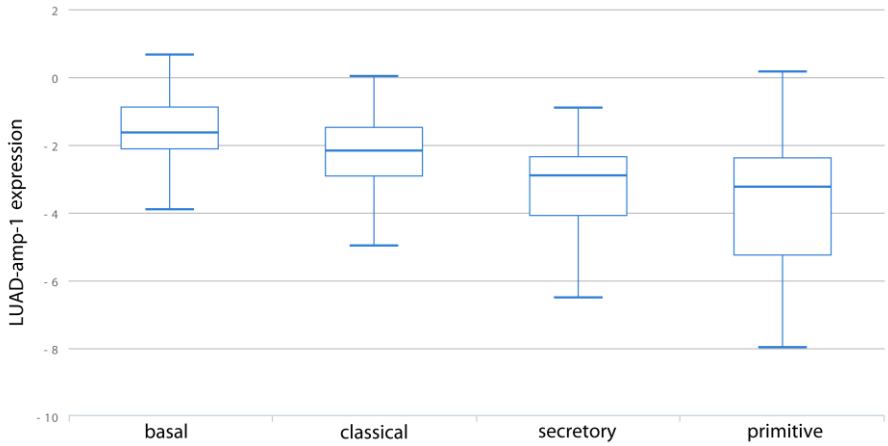

Figure 2. Expression of LUAD-amp-1 in the four LUSC subtypes. The data was obtained from the TANRIC platform.

4.4 LUAD-amp-1 and the hallmarks of cancer

Experimental validation in lung cancer cell lines showed the oncogenic potential of LUAD-amp-1. We carried out inhibition experiments using two different techniques CRISPR/Cas9 and siRNAs. Using these complementary techniques we demonstrate that the functional molecule, involved in tumor progression, is exclusively the RNA encoded by the amplified region. Downregulation of LUAD-amp-1 resulted in impaired cell proliferation and decrease colony formation. Additional experiments (wound healing, soft agar, etc.) could be explored to test if LUAD-amp-1 is involved in other cancer phenotypes.

In vivo tumor formation experiments helped us confirm the *in vitro* results. We performed these experiments with the two cell lines H1648 and HCC95 previously used for the *in vitro*, however no tumors were generated with H1648 cells. In contrast, HCC95 formed small tumors; 45 days post injection we started to measure the tumors and followed for a total of 78 days showing a significant difference between the control group and the two groups where LUAD-amp-1 was inhibited. Using the CRISPR/Cas9 engineered cells we also observed a significant reduction on tumor growth upon LUAD-amp-1 knock down.

Moreover, we also assessed the effects of stably overexpressing LUAD-amp-1 demonstrating its capacity to increase cell proliferation and colony formation *in vitro*. Interestingly, transient overexpression did not result on differential changes in cell proliferation between the control group and the cells overexpressing LUAD-amp-1. However, we did observe significant differences in the colony formation experiment. The differences between these two assays suggest that LUAD-amp-1 could be preferentially involved in a mechanism regulating the clonogenic capacity of the cells. One of our hypotheses is that LUAD-amp-1 via the regulation of TNFα levels participates in the autocrine signaling mediated by TNFα itself. Since cells growing in isolation are highly dependent on autocrine signals for survival, it is plausible to think that LUAD-amp-1 overexpression has a clearer effect on cell proliferation in this experimental setting.

Ongoing experiments include *in vitro* testing of the oncogenic transformation of normal epithelial cells after LUAD-amp-1 overexpression.

4.5 LUAD-amp-1 a novel NF-κB target

Digging into the molecular mechanism by which LUAD-amp-1 promotes cell proliferation and inhibits apoptosis, we find out that LUAD-amp-1 is a NF-κB regulated transcript. The NF-κB subunit p65 binds to two regions surrounding LUAD-amp-1 locus, which contain the consensus sequence for NF-κB binding. NF-κB signaling initiates by TNFα binding to the TNFR1 receptor, followed by signal transduction and proteosome mediated degradation of IκB what results into nuclear translocation of activated NF-κB inducing the expression of its target genes. p65 binding induces the expression of LUAD-amp-1 in a TNFα dependent manner; similarly to what occurs with other p65 targets. Moreover, NF-κB regulates the expression of other lncRNAs for example *NKILA* [132] and *Lethe* [104]. Furthermore, inhibition of p65 decreases LUAD-amp-1 expression level. These evidences led us to conclude that LUAD-amp-1 is also a direct target of NF-κB. Further investigation is required to examine if other stimulus (IL-1β) or the presence of activated oncogenes, such as *KRAS,* are associated with LUAD-amp-1 induction.

NF-κB has a wide variety of targets including inflammatory, proapoptotic and antiapoptotic genes [254]. The NF-κB mediated regulation of LUAD-amp-1 is consistent with its oncogenic function; since NF-κB overactivation has been described for many cancers. Interestingly we found out that LUAD-amp-1 inhibition results in the induction of several NF-κB targets; moreover this induction is increased upon NF-κB activation after TNF-α treatment. These findings suggest that LUAD-amp-1 could have a negative effect on the pathway. Similar to LUAD-amp-1, there are other direct NF-κB targets that turn off NF-κB activation. For example, TNFAIP3/A20 deubiquitinates upstream proteins of the pathway (RIP1) and turns it off. Moreover IκBα is also a target upregulated by NF-κB, which controls NF-κB by sequestering it in the cytoplasm.

TNFα itself is induced by NF-κB, what creates a positive feedback loop among them; demonstrating another mechanism for regulating NF-κB activity. Because LUAD-amp-1 is induced upon TNFα treatment, we explored if LUAD-amp-1 was somehow implicated in TNFα regulation. We did not observe any changes in TNFα levels in the microarray data where LUAD-amp-1 was inhibited. However, in time course experiments where NF-κB was

activated we observed an increase in TNFα expression and a higher induction when LUAD-amp-1 was inhibited. These data suggests that LUAD-amp-1 is implicated in the regulation of TNFα once the pathway becomes activated. We think that additional gene expression experiments (microarray or RNA-seq) activating NF-κB at different time points, could help us define a better picture of LUAD-amp-1 role in the regulation dynamics of TNFα and other NF-κB targets.

Among the hallmarks of cancer it has been recognized that inflammation is an enabling characteristic of tumors. Since we found that LUAD-amp-1 is involved in the regulation of TNFα and other cytokines (CXCL10 and CXCL8) and TNF-receptors (TNFRSF9/CD137), all critical players in inflammation and based on the results with xenograft mice where we observed a dramatic reduction in the size of the tumors after LUAD-amp-1 inhibition, we speculate that LUAD-amp-1 could be contributing to the inflammatory component of the tumor microenviroment. As follow up it would be interesting to explore this idea.

4.6 LUAD-amp-1 molecular mechanism: binding SART3

Similar to LUAD-amp-1 other lncRNAs have been described to affect the NF-κB pathway at different levels. For example, *Lethe* affects NF-κB target activation by directly binding to NF-κB and inhibiting its association to the DNA [104]. Another example is the lncRNA *NKILA* that binds IκB masking its phosphorylation site [132]. Since LUAD-amp-1 is a cytoplasmic lncRNA we reasoned that it could be exerting its function by a similar mechanism binding to a component of the pathway. Further experiments showed that LUAD-amp-1 interacts with the protein SART3. LUAD-amp-1 is the first lncRNA to be identified as a SART3 binder. No evidence so far has linked SART3 directly to NF-κB, however some unpublished data suggest that 'ectopic expression of SART3 induces the translocation of NF-κB/p65 from the cytoplasm to the nucleus' (American Association of Cancer Research, abstract #4557), suggesting a new mechanism for the regulation of NF-κB activity. It remains to be tested if the change in the localization of SART3 observed after LUAD-amp-1 inhibition has an effect on p65 translocation.

SART3 is a protein composed of a N-terminal HAT (half-a-tetratricopeptide repeat) domain a NLS and two C-terminal RNA recognition motifs (RRMs). SART3 is involved in RNA

splicing. It binds directly to the snRNA U6, transiently interacts with the U4/U6 snRNPs, and promotes the recycling of U6 from the spliceosome. THE RRM motifs of SART3 bind to U6 at the hexanucleotide sequence 'ACAGAG'. Interestingly, LUAD-amp-1 contains five 'ACAGA' nucleotides. Based on these data we hypothesized that the RRM motifs of SART3 could specifically bind to 'ACAGA' sequences in LUAD-amp-1. We reasoned that SART3 function in recycling U6 from the spliceosome is not altered by LUAD-amp-1 binding, since SART3 association with U6 occurs in the nucleus while SART3 and LUAD-amp-1 takes place in the cytoplasm.

SART3 expression is very low in normal and non-proliferating cells, but increases in tumors and proliferating cells, supporting a possible role in cancer. Moreover, the subcellular localization of SART3 has been described as mainly nuclear but it can also be in the cytoplasm of proliferating and malignant cells [257], it is reasonable to think that SART3 localization is central for its role in cancer. We found that SART3 is present in the cytoplasm of the cell lines used for exploring LUAD-amp-1 function (A549, HCC95, H1648) making plausible that the interaction between them occurs in the cytoplasm. Interestingly, immunofluorescence experiments showed that upon LUAD-amp-1 inhibition SART3 localizes to cytoplasmic foci. It still remains to be explored the nature of these foci, experiments using protein markers of specific subcellular compartments should be considered for future studies. In addition RNA FISH of LUAD-amp-1 could help us get a better picture of its colocalization with SART3. Moreover, it would be interesting to explore if the change in SART3 localization also alters some of its nuclear functions for example mRNA splicing or its association with the oncoprotein YB-1 [268].

LUAD-amp-1 cytoplasmic localization and its interaction with SART3 led us to explore the mechanisms described so far for this protein in the cytoplasm. Previous work and the presented results demonstrate that SART3 overexpression promotes the nuclear translocation of USP4; we decided to focus on SART3 interaction with USP4 because of USP4 negative regulation of the NF-κB pathway. It is possible that the change of localization of USP4 observed after LUAD-amp-1 inhibition could have an impact on the ubiquitination state of USP4 target proteins (RIP1, TRAF2, and TAK1) altering the NF-κB pathway.

On the other hand, SART3 also interacts with other proteins such as RNSP1, which is involved in splicing. Another hypothesis that remains to be explored is if the change of localization of SART3, after LUAD-amp-1 inhibition, has an effect on mRNA splicing regulation.

Immunofluorescence experiments show that LUAD-amp-1 inhibition results in the cytoplasmic retention of USP4 in the presence of SART3. Structural data shows that the binding of SART3 to USP4 occurs between the HAT repeat 7 of SART3 and the 'domain in USPs' (DUSP) of USP4, leaving the catalytic domain and the ubiquitin like domain (UBL) of USP4 free [264]. Considering that LUAD-amp-1 binds the RRMs of SART3, we can imagine a ribonucleoprotein complex formed by SART3 simultaneously binding USP4 and LUAD-amp-1. LUAD-amp-1 could act as a stabilizer of SART3 and USP4 interaction or even affect USP4 function by interacting with the catalytic site or the ubiquitin like (UBL) domain.

Interestingly, USP15 is a paralog of USP4, and similarly to USP4, SART3 also promotes the entry of USP15 into the nucleus [269]; because of this we can easily imagine that LUAD-amp-1 is involved not only in the regulation of USP4 but also USP15. Moreover, we reason that SART3 interaction with other proteins could also be altered by the presence or absence of LUAD-amp-1. These results implicate LUAD-amp-1 as a new player into the growing body of evidences of DUBs activity in the NF-κB pathway.

So far other mechanism have been proposed for cytoplasmic lncRNAs to explain its role as oncogenes by altering the NF-κB pathway [104], [149]. However, we propose that LUAD-amp-1 uses a completely novel mechanism, our results point to changes in the localization of USP4 and consequently the ubiquitination state of the pathway.

4.7 Summary

Overall, the obtained data identifies LUAD-amp-1 as a lncRNA targeted by amplification in lung adenocarcinoma tumors. Furthermore, LUAD-amp-1 is also overexpressed in lung squamous carcinoma by an epigenetic mechanism. *In vitro* and *in vivo* experiments support and oncogenic role for LUAD-amp-1, affecting cell proliferation. In addition we found that LUAD-amp-1 is and NF-κB target which is induced upon TNFα treatment. Consistent with LUAD-amp-1 dependency on the NF-κB pathway for its induction, we found out that LUAD-amp-1 is also involved in the regulation of other NF-κB targets; suggesting that LUAD-amp-1 could act as a fine tuner of NF-κB activity. Among the genes affected by LUAD-amp-1 we identified TNFα itself (Figure 3). So far our data supports a working model where LUAD-amp-1 expression is induced upon TNFα treatment via the activation of the NF-κB pathway. LUAD-amp-1 is then exported to the cytoplasm where it binds SART3. We hypothesized that this association has an impact on SART3 function. This is support by our results showing that LUAD-amp-1 inhibition affects the nuclear translocation of SART3 binding partner USP4, suggesting that LUAD-amp-1 might be involved in SART3 function and/or SART3/USP4 association. Since USP4 is as a negative regulator of the NF-κB pathway, we propose that LUAD-amp-1 modulates NF-κB activity via altering the localization or function of USP4, via SART3.

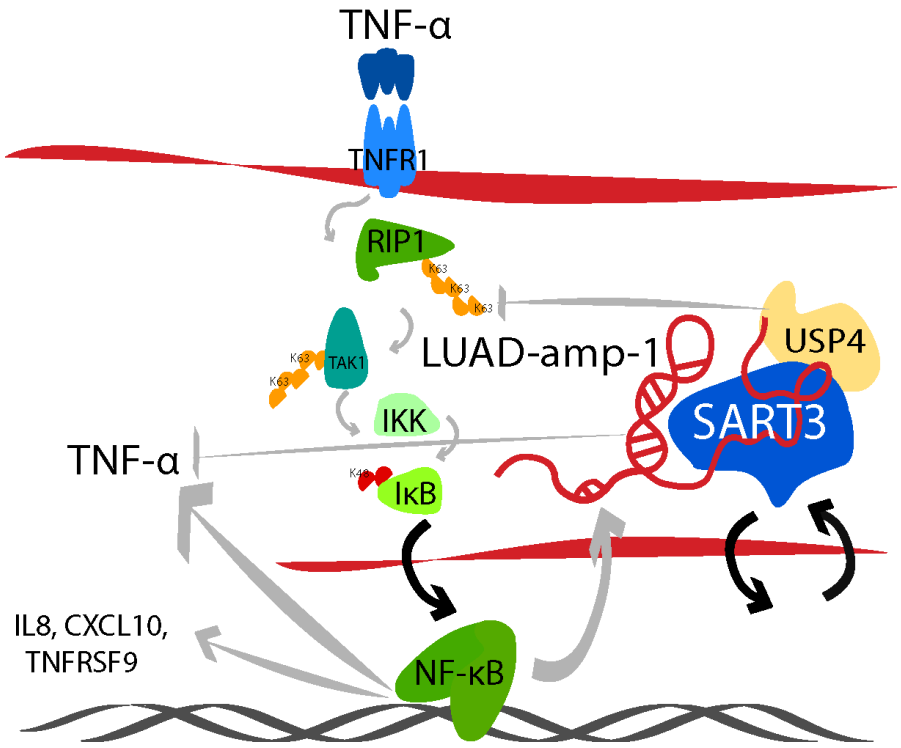

Figure 3. Proposed molecular mechanistic model for LUAD-amp-1. The NF-κB pathway is activated through the binding of TNFα to its cognate receptor (TNFR1). After binding TNFR1 forms an heterodimer initiating a signaling cascade dependent on several proteins (TAK1, RIP1 among many others); some of these proteins assemble complexes via polyubiquitin chains to transduce the activating signal, resulting in the nuclear translocation of NF-κB. Once in the nucleus NF-κB induces the expression of LUAD-amp-1 and other target genes (IL8, CXCL10, TNFRSF9, TNFα). LUAD-amp-1 is then exported to the cytoplasm where it binds SART3. LUAD-amp-1 promotes the nuclear translocation of SART3 and its associated proteins USP4, a negative regulator of the NF-κB pathway.

5. CONCLUSIONS

-The focal peak of 103 somatic copy number alterations (SCNAs) map to noncoding RNA genes. 36 of these alterations contain only long noncoding RNAs (lncRNAs).

-LncRNAs targeted by amplifications are enriched for oncogenic transcription factors. In addition frequently deleted lncRNAs are enriched for insulator marks.

-LUAD-amp-1 is a lncRNA frequently amplified, overexpressed and hypomethylated in lung cancer.

-LUAD-amp-1 is involved in lung cancer cell proliferation, colony formation and apoptosis *in vitro*.

-Inhibition of LUAD-amp-1 reduces tumor formation in a xenograft mouse model.

-LUAD-amp-1 is a direct transcriptional target of NF-κB, induced upon activation of the NF-κB pathway by TNFα treatment.

-LUAD-amp-1 alters the expression of NF-κB target genes including TNFα itself.

-LUAD-amp-1 localizes to the cytoplasm where it interacts with SART3 and regulates its localization.

-LUAD-amp-1 interaction with SART3 alters the localization of USP4, a protein involved in the negative regulation of the NF-κB pathway.

6. MATERIALS AND METHODS

6.1 Data Analysis

The SCNA data was retrieved from the September 2013 run of the TCGA Firehose pipeline (ref). These data contains the focal deletions, and amplifications obtained from GISTIC analysis of Affymetrix 6.0 SNP arrays (q-value < 0.25). Only the data of those tumor types released without restrictions before 2015 was used for the rest of the analysis. Using in house R and perl scripts, all the deletions and amplifications were merged into one file. SCNAs regions were annotated with the Gencode v19; the biotypes from this annotation were used to classify each gene as coding or non-coding. Expression data for the candidate lncRNAs was obtained from *MiTranscriptome* [32].

LUAD-amp-1 expression was analyzed in other lung cancer cohorts GSE18842, and GSE19804 with the probe 231378_at from the Affymetrix HG-U133_Plus_2 microarray. Additionally, LUAD-amp-1 expression data was obtained from the Cancer Cell Line Encyclopedia (CCLE) [270]. Methylation data obtained from the TCGA LUAD HumanMethylation450 BeadChip was retrieved from the TCGA Wanderer interface [271].

6.2 DNA extraction

Genomic DNA was obtained from 2 X 10^5 cells using the DNA extraction kit (QIAGEN). Copy number was assessed by quantitative genomic PCR using primers recognizing LUAD-amp-1 locus (Table qRT-PCR primers, in_LUADAMP1_gDNA). Data was normalized to *PEX19* gene located in chromosome 1p36.23, a region with no significant aneuploidy in the cell lines studied. LUAD-amp-1 loci from the HCC95 cells engineered with the CRISPR/Cas9 were quantified using the same methodology.

6.3 Cell Lines

All the human cell lines used were cultured at 37ºC with 5% of CO_2, using either DMEM, RPM1 or ACL4 medium supplemented with 10% of fetal bovine serum (FBS) and 1% PenStrep (100 U/mL Penicilium and 100 µg/mL Streptomycin). LUAD-amp-1 experiments were carried out in the lung cancer cell lines: HCC95, H1648, A549 and BJ. (Table Cell lines)

6.4 RNA extraction, qPCR, and primer design

Total RNA was extracted using Trizol reagent (Sigma). The concentration and 260/280 of the purified RNA was obtained using Nanodrop. RNA samples were then treated with DNAseI (Invitrogen), for 15 minutes at room temperature followed by 10 minutes incubation at 65ºC with a concentration of 5 mM EDTA to protect RNA from being degraded during DNAseI inactivation. After DNAseI treatment 1 µg of RNA was reverse transcribed with the High Capacity Kit (Applied Biosystems) in a 20 µl volume reaction.
All quantitative PCRs qRT-PCR primers used were designed with the Universal ProbeLibrary Assay Design Center (Roche) or the Primer3 software (http://primer3.ut.ee/) (Table qRT-PCR primers).

All qRT-PCR were carried out in a 384-well format doing quadruplicate reactions in a 10 µl volume. SYBR Green Master Mix (Applied Biosystems) was used for obtaining the amplification signal with the following settings: 15 min at 95^{o}C followed by 40 cycle of 95^{o}C for 30s, and 60^{o}C for 30s in the ViiA7 Real Time PCR system (Applied Biosystems).

6.5 RNAi, transient transfection

At least two siRNAs were designed to target each of the identified copy number altered lncRNAs. SiRNA were designed using the i-Score designer tool [272]. The sequences of the siRNAs and negative control used are indicated in (Table siRNAs). 2 x 10^5 cells per well were plated in a 6 well plate. The next day, siRNAs at a final concentration of 30 nM per well were transfected using Lipofectamine 2000 (Invitrogen) following manufacturer's protocol. Cells were then harvested 24 or 48 hours depending on the experiment.

6.6 Cell proliferation (MTS)

For LUAD-amp-1 experiments one day after transfection 1000 HCC95 or 1500 H1648 cells were plated in 96 well plate in a final volume of 100 µl. Proliferation was then measured using 20 µl of CellTiter96 Aqueous Non-Radioactive Cell Proliferation Assay (MTS) kit (Promega) and proliferation every 2 days up to day 6. For A549 cells overexpressing LUAD-amp-1, the MTS experiment was carried out using 500 cells per well and cell proliferation was measured, since these cells have a higher proliferation rate.

6.7 Colony formation

HCC95 cells were transfected with the corresponding siRNAs, as previously described; twenty-four hours after cells were trypsinized and counted. 500 cells were plated in 6-well plates and left to form colonies over a period of 10-14 days. The colonies formed were washed twice with PBS and then fixed using 0.5% glutaraldehyde, for 20 minutes. After another PBS wash, colonies were stained with 1% crystal violet in 35% of methanol solution. The number of colonies formed was counted manually. Crystal violet staining was solubilized with 10% acetic acid, and absorbance was measured at 570 nm wavelengths with a plate reader.

6.8 Xenograft Models *in vivo* experiments

Eight weeks old female BALB/c Rag2-/- mice were used for the xenograft experiments. To generate the tumors 1.5 million HCC95 cells were resuspended in 100 μl of PBS, mixed with one volume of Matrigel® Matrix (Corning) and injected subcutaneously into the flank of mice. Tumor growth was followed every three days by perpendicular caliper measurements. Tumor volume was calculated using the formula $(\pi/6)$*length*width^2. 78 days after tumor implantation or when tumor size was bigger than 1000 mm^3, mice were sacrificed and tumors were harvested.

6.9 Western Blot

Cells were lysed in RIPA buffer (50 mM Tris-HCl ph7.4, 150 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 0.1% SDS), by incubating the samples on ice for 15 minutes, followed by centrifugation at 13,000 rpm for 15 minutes. Protein concentration was determined with the Pierce BCA Protein Assay Kit (Thermo Fisher). Equal amounts of protein were resolved by SDS-PAGE, and blotted onto nitrocellulose membranes (Bio-Rad). Membranes were blocked for 2 hours with 5% (bovine serum albumin) BSA or non-fat milk dissolved in PBS with 0.05% Tween 20 (PBST), followed by overnight incubation with the primary antibody at 4^{o}C. After 1-hour incubation at room temperature with the secondary antibody, membranes were washed in PBST. The ECL reagent was used for antibody detection. The antibodies used in this study are listed in (Table antibodies). The relative expression levels of proteins were obtained based on the intensity of the western blot bands using the

software Image Studio. The quantified intensities were normalized to those of the loading reference protein and the fold change relative to the control condition was then calculated.

6.10 Microarray

Total RNA from HCC95 cells transfected with control siRNA or siRNAs targeting LUAD-amp-1 was extracted using Trizol,, RNA quality was assessed using Bioanalyzer (Agilent) and hybridized to Affymetrix GeneChip Human U133 Plus 2.0 Array. Data was normalized using the Robust Multichip Average (RMA) algorithm. The subsequent steps for the microarray analysis were carried out with R, using the BioConductor packages: affy [273] and limma [274].

6.11 CRISPR/Cas9

To delete exon D of LUAD-amp-1 two single-guide RNAs (sgRNAs) flanking the region were designed using the tool available at: http://crispr.mit.edu/. The sgRNA with the highest score was selected; sequences of the sgRNAs used are listed in (Table sgRNAs). sgRNAs were cloned in pX330 vector by using BbSI restriction enzyme as described in the Zhang lab CRISPR protocol: https://www.addgene.org/crispr/zhang/

To generate the genomic deletion, 3 x 10^6 HCC95 cells were co-transfected with 4.5 ug of each pX330-sgRNA and 1 μg of GFP expressing plasmid. Control cells were transfected with empty pX330 vector. 24 hours after transfection, GFP positive cells were sorted in six 96 well-plates using the BD FACSAria IIu cytometer. After 2 days we selected those wells having only one cell/colony and we allowed them to grow until they reached confluency, adding new media every 3 days. Cells were then trypsinized and split into two plates. Genomic DNA was then extracted with the QuickExtract reagent (Epicentre), and directly used for PCR reaction.

Genotyping was done using PCR primers upstream and downstream of the sgRNA cleavage sites (Table primers). PCR products were then run in an agarose gel to check for the amplicon size and compare it to the wild-type (800-bp) or expected deletion (300-bp). PCR products of the obtained deletions were then gel purified and sent for sequencing.

Sequencing results are presented in the appendix (sequence CRISPR/Cas9 clones), as well as the sequence of the primers used for genotyping.

6.12 LUAD-amp-1 overexpression

LUAD-amp-1 cDNA sequence was obtained from IMAGE Consortium, clone id: HU4_p940H1163D. LUAD-amp-1 was then cloned between *EcoRI-BamHI* sites in pcDNA3.0 vector (Invitrogen), and between *EcoRI-XhoI* sites in pMSCVneo retroviral vector (Clontech), for transient and stable overexpression respectively. To generate A549 cells with stable overexpression of LUAD-amp-1 the pMSCVneo-LUAD-amp-1 construct was co-transfected in HEK293T with the helper plasmids encoding Gag-Pol and VSV-G genes. The generated retroviruses were then used for A549 transduction followed by neomycin selection.

6.13 TNFα treatment

HCC95 cells were treated with 10 ng/ml of TNFα (R&D systems, 210-TA) for different time points.

6.14 Apoptosis quantification by Annexin V detection

$2x10^6$ HCC95 cells were detached with acutase, to minimize membrane damage, washed with PBS, and resuspended in annexin V binding buffer (PE Annexin V Apoptosis Detection Kit I, BD Pharmingen). Cells were then pelleted by centrifugation (1200 rcf, 5 min), and incubated at room temperature with PE Annexin V in a buffer containing 7-AAD for 15 min. The number of Annexin V and 7-AAD positive cells was measured by flow cytometry using FACSCalibur (BD Biosciences) and FlowJO analysis software. A total of 15000 cells were acquired for each condition. Cells undergoing early stages apoptosis were gated as PE Annexin V positive and 7-AAD negative; cells in late stages of apoptosis were selected as PE Annexin V and 7-AAD positive.

6.15 Nuclear/cytoplasmic fractionation

3×10^6 cells were lysed in 500 µl of lysis buffer (20 mM Tris-HCl pH 7.5%, 0.1% NP-40, 280 mM NaCl, 3 mM $MgCl_2$, and RNasin (Promega)), and incubated on ice for 10 min. The cell lysate was then layered over 500 µl of sucrose cushion (50% sucrose in cell lysis buffer),

and centrifuged at 13,000 rpm at 4ºC for 10 min; the resulting supernatant corresponded to the cytoplasmic fraction. Nuclei pellet was then resuspended in 500 ul of lysis buffer containing triton (10 mM Tris, 100 mM NaCl, 1 mM EGTA, 300 mM sucrose, 0.5 mM NaVO3, 50 mM NaF, 1 mM phenylmethylsulphonyl fluoride, 0.5% triton X-100, protease inhibitor cocktail and RNAsin (Promega)); RNA from all fractions was then extracted using 1 ml of Trizol Reagent.

6.16 RNA pull-down assay

LUAD-amp-1 was in vitro transcribed using the T7/T3 RNA polymerase kit from Promega. For this the pT3T7D-*PacI* plasmid was digested with *EcoRI* or *NotI* to generate the sense and antisense sequence of LUAD-amp-1 respectively. After digestion the plasmids were used as template for in vitro transcription and biotin-16-UTP was incorporated. The resulting RNA was treated with RNAse-free DNAse I (Promega) and purified using the MicroSpin G-50 columns (GE Healthcare). HCC95 protein lysate was prepared with the lysis buffer (20 mM Tris-HCl at pH 7.5, 100 mM KCl, 5 mM MgCl2, 0.5% NP-40, protease inhibitors (Roche), RNase inhibitor (100 U / ml), and 10 mM DTT). The protein lysate was then pre-cleared using Dynabeads® MyOne Streptavidin C1 (Invitrogen) for one hour with rotation at 4ºC. In vitro transcribed RNA was incubated with the pre-cleared protein lysate diluted one to one with Buffer A (150mM KCl, 25mM Tris pH 7.4, 5mM EDTA, 0.5% NP-40, protease inhibitors (Roche), RNase inhibitor (100 U /ml), and 0.5 mM DTT), for one hour with rotation followed by one hour with streptavidin beads. Beads were then washed five times using Buffer A, and resuspended in SDS-PAGE sample Buffer. Interacting proteins were resolved in a NuPAGE Novex 4-12% bis-Tris gel (Invitrogen), and visualized by silver staining with the SilverQuest Silver Staining Kit (ThermoFisher) following manufacturers protocol. For mass spectrometry analysis the observed differential bands were cut and sent to Taplin Mass Spectrometry Facility. A replicate of the pull-down experiment was performed to narrow down the associated protein candidates.

6.17 RNA immunoprecipiation (RIP)

1 x 10^7 A549 cells overexpressing LUAD-amp-1 were lysed with lysis buffer (20 mM Tris-HCl at pH 7.5, 100 mM KCl, 5 mM MgCl2, 0.5% NP-40, protease inhibitors (Roche), RNase inhibitor (100 U / ml), and 10 mM DTT). Protein lysate was then incubated with pre-

washed Protein-A magnetic beads for one hour with rotation at 4ºC, for pre-clearing. Extract was diluted up to 1 ml with RIP Buffer, and incubated either with Normal Rabbit IgG (Cell Signaling, 2729S) or with anti-SART3 (abcam, ab155765), overnight with rotation at 4ºC. Protein-A magnetic beads were added for one hour, and then washed five times with Buffer A (150 mM KCl, 25mM Tris pH 7.4, 5mM EDTA, 0.5% NP-40, protease inhibitors (Roche), RNase inhibitor (100 U /ml), and 0.5 mM DTT), for the last wash PBS was used. RNA was recovered from the beads by resuspending them with Trizol Reagent.

6.18 Immunofluorescence

HCC95 cells were grown on glass coverslips, and transfected with the corresponding plasmids and siRNAs. 72 hours after transfection, the cells were fixed with 4% methanol-free paraformaldehyde, and were incubated with the anti-Myc (Cell Signaling #2276) antibody for 30 minutes followed by incubation with the donkey anti-mouse antibody coupled to Alexa Fluor® 488 (ThermoFisher) for 30 more minutes. Coverslips were then washed three times with washing buffer (0.5% NP-40, 0.01% Na azide, diluted in PBS), briefly air-dried, and mounted in the DAPI-containing VectaShield medium (Vector). Cells were visualized using the 20x or 40x objectives with the automated microscope Zeiss Axio Imager M1, and images were captured with the ZEN microscopy software. Foci quantification was performed with ImageJ and CellProfiler software.

6.19 Statistical analysis

All experimental data was plotted and analyzed using the GraphPad statistical software. Data is represented as the average of at least three biological replicates. Significance was obtained using the statistical test corresponding to each type of data analyzed. P-values are given using the following thresholds: ns for p-value > 0.05; * for p-value ≤ 0.05; ** for p-value ≤ 0.01; *** for p-value ≤ 0.001

APPENDIX

Table qRT-PCR primers

Target	Forward primer	Reverse primer	Use
LUAD-AMP-1	CAGCCTGAGGTTCATCGTTC	TCAGCAGTGAGGACATTCTGAT	qRT-PCR
LUAD-AMP-1_IS	GACAGAATCTGGCAACAGATGA	TTGATGTCCCCAATATACCTGA	qRT-PCR
LUAD-AMP-1_exonC	AGCCGGAAACACAGAAAGG	GGGATCCCAATTGTAGAGCA	qRT-PCR
LUAD-AMP-1_exonD	AAAGGTCACTTCCTGGCAGA	GGAATCTCAGGTTAGACTTTCCA	qRT-PCR
IKBKB	AATGGAGCAGGCTGTGGA	CATGGGGCTCCTCTGTAAGT	qRT-PCR
PLAT	GCTGACGTGGGAGTACTGTG	CTGAGGCTGGCTGTACTGTCT	qRT-PCR
p65	TCATGAAGAAGAGTCCTTTCAGC	GGATGACGTAAAGGGATAGGG	qRT-PCR
CXCL10	TTCTGATTTGCTGCCTTATC	CTTGGATTAACAGGTTGATTACT	qRT-PCR
CCL20	GCTGCTTTGATGTCAGTGCT	GAAGAATACGGTCTGTGTATCCAA	qRT-PCR
CXCL8	GAGTGGACCACACTGCGCCA	TCCACAACCCTCTGCACCCAGT	qRT-PCR
TNFRSF9	TGTGACATATGCAGGCAGTG	CTGGAGTGCAGTCACACTCTG	qRT-PCR
IL1A	GGTTGAGTTTAAGCCAATCCA	TGCTGACCTAGGCTTGATGA	qRT-PCR
SERPINE2	GGTCCGGAATGTGAACTTTG	ATCTGGGGACAGCAGATTGT	qRT-PCR
PTGES	CATGTGAGTCCCTGTGATGG	CTGCAGCAAAGACATCCAAA	qRT-PCR
TNFα	CAGCCTCTTCTCCTTCCTGAT	GCCAGAGGGCTGATTAGAGA	qRT-PCR
SART3	AGAAGGGTTGATTTCAAACAAGA	CTCCAAGGCACGAGTAAAGG	qRT-PCR
USP4	ATGAGGACCACACTCCAACG	GCTGTCAAAGCCCACATACTT	qRT-PCR
U6	GCTTGCTTCAGCACATA	AAAAACATGGAACTCTTCACG	qRT-PCR
in_LUADAMP1_gDNA	TTGATGTCCCCAATATACCTGA	TGTGGTTTCTTTATGTGGTTGC	gDNA
out_LUADAMP1_gDNA	TGCCCACTGTCTGGCAAA	CCACCATGCTTGGCACTAT	gDNA
PEX19_gDNA	AGCCTAGCCAGGGCTTACTC	CTGAGGTCAACCTGCTCACA	gDNA

Table cell lines

Cell line	Description	Tissue	Type
A549	epithelial cell line derived from a lung carcinoma tissue	lung	cancer
BJ	skin fibroblast	skin	normal
H1648	stage 3A lung adenocarcinoma	lung	cancer
HCC95	squamous cell lung carcinoma	lung	cancer

Table antibodies

Antibody	Reference	Provider	Host
anti-Phospho-NF-kB p65 (Ser536) (93H1)	#3033	Cell Signaling	Rabbit monoclonal
anti-GAPDH	ab9484	Abcam	Mouse monoclonal
anti-Lamin AC	#2032	Cell Signaling	Rabbit polyclonal
anti-SART3	ab155765	Abcam	Rabbit polyclonal
anti-USP4	A300-830A	Bethyl	Rabbit polyclonal
anti-Myc-Tag (9B11)	#2276	Cell Signaling	Mouse monoclonal

Table plasmids

Plasmid Name	Use/Annotations
pX330_empty	CRISPR/CAS9 system
pX330_lex3	CRISPR/CAS9 system; sgRNA cloning
pX330_rex3	CRISPR/CAS9 system; sgRNA cloning
pT3T7D_LUAD_comercial_clone_365492	I.M.A.G.E MGC clone number 365492
pcDNA3-LUAD-amp-1	Cloned between EcoR1 and Not1
pMSCV_neo_LUAD-amp-1	Cloned between EcoR1 and XhoI

pcs2-myc-SART3	provided by Michael Rape
pcs2-myc-USP4	provided by Michael Rape

Table siRNAs

Target gene	Sequence
LUADAMP1_siRNA_1	GGUAUAUUGGGGACAUCAA
LUADAMP1_siRNA_2	GGAUAUGGAGAAAAUUAUU
LUADAMP1_siRNA_3	GGAUUACAUUAGUGGACAA
SART3_siRNA_1	GGAUAUAGCUGUUCAGAAA
SART3_siRNA_2	CGUGGAGUAUGAAAAUGAA
p65_siRNA	GCCCUAUCCCUUUACGUCA
control_siRNA	CAGUCGCGUUUGCGACUGGC

Table TFBS.counts

Transcription_factors (ENCODE)	LncRNA_amplified
MYC	6.9324E-07
POLR2A	1.07194E-06
SMARCB1	2.26343E-06
FOXP2	3.80056E-06
MAX	6.97296E-06
REST	9.36887E-06
TFAP2C	1.00129E-05
TFAP2A	1.31117E-05
EP300	1.57993E-05
TCF7L2	2.62234E-05
MAZ	4.18025E-05
PAX5	4.88405E-05
TAF1	6.11723E-05
GABPA	8.01344E-05
FOXA1	8.2386E-05
NR3C1	9.18768E-05
CHD2	0.000193602
RCOR1	0.00025656
GATA3	0.000258896
YY1	0.000351893

Transcription_factors (ENCODE)	LncRNA_deleted
POU5F1	0.008020652
SP2	0.010328699
BACH1	0.012393121
PBX3	0.020892665
ZKSCAN1	0.037966923
NANOG	0.038473671
GRp20	0.062104164
STAT2	0.098943122
NR3C1	0.10667074
EZH2	0.118496672
PPARGC1A	0.118948116
SETDB1	0.125210416
SRF	0.130927078
NFYA	0.135183878
ZNF263	0.209070614
NRF1	0.225796215
ESRRA	0.228084765
CTBP2	0.236199576
RDBP	0.241172099
SMC3	0.263687046

FOXA2	0.000360633	E2F4	0.270027759
HMGN3	0.000455431	ZNF274	0.275364427
BCLAF1	0.000478672	HSF1	0.287125839
JUND	0.000520109	MAFK	0.289423073
E2F4	0.000783235	GABPA	0.299066519
SMC3	0.000959296	IRF3	0.314288087
ETS1	0.001173525	HNF4G	0.34754442
RFX5	0.001187981	KDM5B	0.361189282
BHLHE40	0.001289257	GATA3	0.363784744
CCNT2	0.001317007	WRNIP1	0.372425088
SMARCC1	0.001477002	RAD21	0.388836162
FOXM1	0.001546473	REST	0.394229033
EGR1	0.001574755	SP4	0.398648877
ATF1	0.001616713	HDAC6	0.408744787
GTF2F1	0.002197745	CTCF	0.415491521
USF1	0.00227126	MAX	0.416492612
RBBP5	0.002433801	SIRT6	0.424708206
HNF4A	0.002465337	CHD2	0.425097793
ZBTB33	0.002588752	FOSL2	0.456896782
PML	0.003036444	ELK1	0.45946467
ZBTB7A	0.003090402	ATF3	0.468195506
TCF12	0.003109834	BCL3	0.47232061
SIX5	0.003615311	PHF8	0.494029223
CTBP2	0.003658934	BRF2	0.502514908
CEBPB	0.003705032	STAT5A	0.509090474
NR2F2	0.003745347	RBBP5	0.514545104
TBP	0.003747814	ESR1	0.522112094
BRCA1	0.003753588	SAP30	0.528556779
MEF2A	0.004067062	ELF1	0.532257136
STAT5A	0.004097129	RPC155	0.53751591
MXI1	0.004137928	BRCA1	0.544522977
STAT1	0.004371717	ELK4	0.554961649
E2F1	0.004404069	EBF1	0.55758452
JUN	0.004953601	TRIM28	0.558533817
TCF3	0.004981122	TCF12	0.563230959
CTCF	0.005119954	NFYB	0.563605356
ELF1	0.005142752	IKZF1	0.5681906
GATA1	0.005356035	E2F1	0.571879281
NR2C2	0.005368331	YY1	0.579334205
ZKSCAN1	0.00542652	HDAC1	0.580396242
CREB1	0.005497914	THAP1	0.584276458
TBL1XR1	0.005605784	SIN3AK20	0.587382356
HDAC2	0.006254841	TCF7L2	0.601896391
SP1	0.007100916	CREB1	0.603302464
CBX3	0.007195023	MYC	0.606146345
UBTF	0.007284788	HDAC2	0.608655078

USF2	0.007627332	CHD1	0.61008486
SMARCA4	0.008517914	TAF1	0.620921125
BCL3	0.008635632	CEBPD	0.625149051
HDAC1	0.009840579	FOXP2	0.634016765
TRIM28	0.009873093	TCF3	0.634395009
SIN3AK20	0.010450428	GTF2F1	0.63637949
CHD1	0.01067875	PRDM1	0.644594393
POU2F2	0.0108482	IRF4	0.646476772
IRF1	0.010935747	SMARCA4	0.657171667
SIRT6	0.011316895	MEF2C	0.659708409
ZNF263	0.01237869	ZNF143	0.659778077
KAP1	0.012527462	MXI1	0.663152958
E2F6	0.012662194	GTF2B	0.669516011
BRF2	0.012739938	SIX5	0.671147697
ESRRA	0.013868897	MAFF	0.677378807
TAL1	0.014417984	EGR1	0.693293562
PRDM1	0.014710224	GATA1	0.695344681
STAT3	0.014888562	CTCFL	0.695658849
SAP30	0.015852206	POLR2A	0.705169719
TEAD4	0.016526998	TAF7	0.709326358
STAT2	0.016655905	ZZZ3	0.713498341
FOSL2	0.017618778	ATF1	0.727617727
PHF8	0.019322829	SP1	0.747368582
ELK4	0.02081719	IRF1	0.758494081
ATF3	0.023574199	GATA2	0.767906784
RXRA	0.023826207	BATF	0.77100607
NRF1	0.024457271	E2F6	0.778572798
ATF2	0.024746266	RUNX3	0.781237937
GTF2B	0.027420951	STAT1	0.781426885
HNF4G	0.028878124	FOXA2	0.786002412
ELK1	0.02901738	HNF4A	0.788269133
HSF1	0.031811193	NR2C2	0.791774388
KDM5B	0.032074562	MAZ	0.799175115
CTCFL	0.033812464	UBTF	0.799249414
TAF7	0.033820687	BCLAF1	0.802404326
RELA	0.034247838	POU2F2	0.808576938
MEF2C	0.037677135	KAP1	0.809106212
FOSL1	0.040569337	FOXA1	0.817137111
MAFF	0.046094835	RFX5	0.821875928
RUNX3	0.050447529	SUZ12	0.823361855
MYBL2	0.051495506	ZBTB7A	0.82617114
JUNB	0.054920885	BCL11A	0.830258185
BACH1	0.057394551	RXRA	0.830518062
POLR3G	0.058576947	NFATC1	0.840425359
GATA2	0.062551522	ETS1	0.848315984
EBF1	0.066809517	TFAP2C	0.863981309

PBX3	0.068187499	PAX5	0.864993391
ZEB1	0.068946794	FOSL1	0.87174985
NFIC	0.069798895	FOXM1	0.87202827
CEBPD	0.071832321	CCNT2	0.874712373
NANOG	0.073515598	NFE2	0.878352446
ZNF274	0.076007785	BRF1	0.880957946
ZNF143	0.081594661	MTA3	0.883129238
ARID3A	0.083155653	GTF3C2	0.888849544
RPC155	0.084829869	ZNF217	0.890708039
GRp20	0.087710186	POLR3G	0.89358289
RAD21	0.089412605	TEAD4	0.896941281
POU5F1	0.090637404	HDAC8	0.898763877
MAFK	0.092350329	TBP	0.899462274
ESR1	0.095665746	TFAP2A	0.902421374
MBD4	0.104949692	HMGN3	0.910281316
FOS	0.105018223	TBL1XR1	0.913153226
BDP1	0.138118296	JUN	0.917760935
NFYB	0.146691324	KDM5A	0.920895754
IKZF1	0.148842372	MYBL2	0.923118015
SPI1	0.152010957	BHLHE40	0.923426254
GTF3C2	0.157846369	JUNB	0.926095847
SIN3A	0.158553666	ATF2	0.929122275
RDBP	0.175304087	FOS	0.933005327
SRF	0.176369664	ZBTB33	0.933100268
SMARCC2	0.180736131	PML	0.937555337
SETDB1	0.188085235	FAM48A	0.940289622
BCL11A	0.191772508	SIN3A	0.94043254
IRF3	0.214205815	RELA	0.941583395
KDM5A	0.214205815	SREBP1	0.941961593
SUZ12	0.224261271	CBX3	0.942064271
NFYA	0.236063053	STAT3	0.95787768
NFE2	0.250369269	USF2	0.961731839
WRNIP1	0.268767147	MBD4	0.964031645
FAM48A	0.296783424	TAL1	0.965809952
BRF1	0.315967075	SPI1	0.96697569
IRF4	0.317080418	NR2F2	0.968788062
NFATC1	0.324559057	USF1	0.969664887
THAP1	0.356086708	ZEB1	0.969846921
EZH2	0.365108346	EP300	0.973984135
SP4	0.385394372	BDP1	0.975244264
ZNF217	0.385731131	ARID3A	0.977167235
HDAC8	0.481410598	JUND	0.977567097
ZZZ3	0.493366589	RCOR1	0.982087088
MTA3	0.494277019	SMARCB1	0.985534666
HDAC6	0.607732248	CEBPB	0.986333109
SREBP1	0.808876552	MEF2A	0.987050036

BATF	0.825661301	SMARCC2	0.993782607
SP2	0.840782506	SMARCC1	0.99415232
PPARGC1A	0.848100108	NFIC	0.998779198

Table SNPs.lncRNAs

SNPs	Associated Disease	chr	Start	End	LncRNA-ID
rs2048672	Breast cancer	chr7	130626519	130794935	*MKLN1-AS1/PINT*
rs10094872	Urinary bladder cancer	chr8	128698588	128746213	*CASC11*
rs9642880	Bladder cancer	chr8	128698588	128746213	*CASC11*
rs9642880	Urinary bladder cancer	chr8	128698588	128746213	*CASC11*
rs1447295	Prostate cancer	chr8	128302062	128494384	*CASC8*
rs7014346	Colorectal cancer	chr8	128302062	128494384	*CASC8*
rs6983267	Colorectal cancer	chr8	128302062	128494384	*CASC8*
rs6983267	Prostate cancer	chr8	128302062	128494384	*CASC8*
rs6983267	Prostate cancer (early onset)	chr8	128302062	128494384	*CASC8*
rs10505477	Prostate cancer (early onset)	chr8	128302062	128494384	*CASC8*
rs10505477	Colorectal cancer	chr8	128302062	128494384	*CASC8*
rs2392780	Breast cancer (early onset)	chr8	128302062	128494384	*CASC8*
rs2392780	Breast cancer (early onset)	chr8	128351519	128404876	*RP11-382A18.2*
rs1562430	Breast cancer	chr8	128302062	128494384	*CASC8*
rs1562430	Breast cancer	chr8	128351519	128404876	*RP11-382A18.2*
rs13281615	Breast cancer	chr8	128302062	128494384	*CASC8*
rs13281615	Breast cancer	chr8	128351519	128404876	*RP11-382A18.2*
rs445114	Prostate cancer	chr8	128302062	128494384	*CASC8*
rs16902094	Prostate cancer	chr8	128302062	128494384	*CASC8*
rs704017	Colorectal cancer	chr10	80703085	80827652	*ZMIZ1-AS1*
rs1353747	Breast cancer	chr5	58335588	58359330	*RP11-266N13.2*
rs1017226	Breast cancer (early onset)	chr5	56137843	56157991	*AC008937.2*
rs1011970	Breast cancer	chr9	21994777	22121096	*CDKN2B-AS1*
rs2823779	Response to radiotherapy in cancer (late toxicity)	chr21	17442842	17999716	*LINC00478*
rs17142289	Response to radiotherapy in cancer (late toxicity)	chr6	6346698	6623004	*LY86-AS1*
rs10774214	Colorectal cancer	chr12	4357931	4385350	*RP11-264F23.3*
rs9874556	Pancreatic cancer	chr3	3292371	3668980	*AC026188.1*

Sequence CRISPR/Cas9 clones

Primers used for genomic DNA amplification

>out_LUADAMP1_gDNA_fwd
TGCCCACTGTCTGGCAAA

>out_LUADAMP1_gDNA_rev
CCACCATGCTTGGCACTAT

>clone.23
CTAGTCTCCTTATAACAGGGACACAATTGAAAACCTTCGGGCATGGCGCTCCTCACATCCCAGATG
GGGTGGTGGCCGGGCAGAGGCACTCCTCATTTGCCANACNGTGGGNAANNNNNNGNNNGCCN
CTGNNNGNNNNNNNCTGGGNTGTGAGGAGCGCCATGNCNANNGTTTTCAATTGTGTCCTGTTA
TAAGGAGACTAGATTCTTACTGAATTTATGAAAATAACTATATAGTGCCAANCAGGNNGNAN

>clone.24
CTAGTCTCCTTATAACAGNGGNCACAATTGAAAACCTTCGGGNANNNNGCTCCTCACATCCCAGA
TGGGGTGGTGGCCGGGCAGAGGCACTCCTCATTTGCCANANNGTGNGNAANNNANNGNCNNN
CNCNGNNNGNNNNNNNCTGGGNTGTGAGGAGCGCCANGNNNANGTTTTCNATTGTNNNNGTT
ATAANGAGACTAGATTCTTACTGAATTTATGAAAATAACTATATAGTGCCNNNAGGNGGNA

REFERENCES

1. Ghosh S, Hayden MS: **New regulators of NF-kappaB in inflammation.** *Nat Rev Immunol* 2008, **8:**837-848.
2. Oeckinghaus A, Hayden MS, Ghosh S: **Crosstalk in NF-κB signaling pathways.** *Nat Immunol* 2011, **12:**695-708.
3. Gilbert W: **Origin of Life - the Rna World.** *Nature* 1986, **319:**618-618.
4. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S: **The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme.** *Cell* 1983, **35:**849-857.
5. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR: **Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena.** *Cell* 1982, **31:**147-157.
6. Crick FH: **On protein synthesis.** *Symp Soc Exp Biol* 1958, **12:**138-163.
7. Palade GE: **A small particulate component of the cytoplasm.** *J Biophys Biochem Cytol* 1955, **1:**59-68.
8. Hoagland MB, Stephenson ML, Scott JF, Hecht LI, Zamecnik PC: **A soluble ribonucleic acid intermediate in protein synthesis.** *J Biol Chem* 1958, **231:**241-257.
9. Weinberg RA, Penman S: **Small molecular weight monodisperse nuclear RNA.** *J Mol Biol* 1968, **38:**289-304.
10. Jacob F, Monod J: **Genetic regulatory mechanisms in the synthesis of proteins.** *J Mol Biol* 1961, **3:**318-356.
11. Gilbert W, Muller-Hill B: **Isolation of the lac repressor.** *Proc Natl Acad Sci U S A* 1966, **56:**1891-1898.
12. Lee RC, Feinbaum RL, Ambros V: **The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.** *Cell* 1993, **75:**843-854.
13. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G: **The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.** *Nature* 2000, **403:**901-906.
14. Aravin AA, Hannon GJ, Brennecke J: **The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race.** *Science* 2007, **318:**761-764.
15. Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest AR, Grimmond SM, Schroder K, et al: **Tiny RNAs associated with transcription start sites in animals.** *Nat Genet* 2009, **41:**572-578.
16. Taft RJ, Simons C, Nahkuri S, Oey H, Korbie DJ, Mercer TR, Holst J, Ritchie W, Wong JJ, Rasko JE, et al: **Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans.** *Nat Struct Mol Biol* 2010, **17:**1030-1034.
17. Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF: **A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome.** *Nature* 1991, **349:**38-44.
18. Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S: **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:**515-526.
19. Meller VH, Wu KH, Roman G, Kuroda MI, Davis RL: **roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system.** *Cell* 1997, **88:**445-457.
20. Lee JT, Davidow LS, Warshawsky D: **Tsix, a gene antisense to Xist at the X-inactivation centre.** *Nature Genetics* 1999, **21:**400-404.
21. Lyle R, Watanabe D, te Vruchte D, Lerchner W, Smrzka OW, Wutz A, Schageman J, Hahner L, Davies C, Barlow DP: **The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1.** *Nat Genet* 2000, **25:**19-21.
22. Hangauer MJ, Vaughn IW, McManus MT: **Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs.** *PLoS Genet* 2013, **9:**e1003569.
23. Ohno S: **So much "junk" DNA in our genome.** *Brookhaven Symp Biol* 1972, **23:**366-370.

24. de Koning APJ, Gu WJ, Castoe TA, Batzer MA, Pollock DD: **Repetitive Elements May Comprise Over Two-Thirds of the Human Genome.** *Plos Genetics* 2011, **7**.
25. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, et al: **The transcriptional landscape of the mammalian genome.** *Science* 2005, **309:**1559-1563.
26. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, et al: **Global identification of human transcribed sequences with genome tiling arrays.** *Science* 2004, **306:**2242-2246.
27. Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, Gingeras TR: **Large-scale transcriptional activity in chromosomes 21 and 22.** *Science* 2002, **296:**916-919.
28. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY: **Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.** *Cell* 2007, **129:**1311-1323.
29. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, et al: **Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.** *Nature* 2009, **458:**223-227.
30. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.** *Nature* 2007, **448:**553-560.
31. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al: **Landscape of transcription in human cells.** *Nature* 2012, **489:**101-108.
32. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, et al: **The landscape of long noncoding RNAs in the human transcriptome.** *Nat Genet* 2015, **47:**199-208.
33. Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJ, Gough J, Denisenko E, Schmeier S, Poulsen TM, Severin J, et al: **An atlas of human long non-coding RNAs with accurate 5' ends.** *Nature* 2017, **543:**199-204.
34. Wang KC, Chang HY: **Molecular mechanisms of long noncoding RNAs.** *Mol Cell* 2011, **43:**904-914.
35. Ulitsky I, Bartel DP: **lincRNAs: genomics, evolution, and mechanisms.** *Cell* 2013, **154:**26-46.
36. Mele M, Mattioli K, Mallard W, Shechner DM, Gerhardinger C, Rinn JL: **Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs.** *Genome Res* 2017, **27:**27-37.
37. Rinn JL, Chang HY: **Genome regulation by long noncoding RNAs.** *Annu Rev Biochem* 2012, **81:**145-166.
38. Dinger ME, Pang KC, Mercer TR, Mattick JS: **Differentiating Protein-Coding and Noncoding RNA: Challenges and Ambiguities.** *Plos Computational Biology* 2008, **4**.
39. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, et al: **Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.** *Nature* 2002, **420:**563-573.
40. Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH, Wright F: **Codon Usage Patterns in Escherichia-Coli, Bacillus-Subtilis, Saccharomyces-Cerevisiae, Schizosaccharomyces-Pombe, Drosophila-Melanogaster and Homo-Sapiens - a Review of the Considerable within-Species Diversity.** *Nucleic Acids Research* 1988, **16:**8207-8211.
41. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G: **CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine.** *Nucleic Acids Research* 2007, **35:**W345-W349.
42. Wang L, Park HJ, Dasari S, Wang SQ, Kocher JP, Li W: **CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model.** *Nucleic Acids Research* 2013, **41**.
43. Lin MF, Jungreis I, Kellis M: **PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions.** *Bioinformatics* 2011, **27:**I275-I282.

44. Housman G, Ulitsky I: **Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs.** *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* 2016, **1859:**31-40.
45. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al: **Pfam: the protein families database.** *Nucleic Acids Research* 2014, **42:**D222-D230.
46. Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching.** *Nucleic Acids Research* 2011, **39:**W29-W37.
47. Ingolia NT, Lareau LF, Weissman JS: **Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes.** *Cell* 2011, **147:**789-802.
48. Kisselev LL, Buckingham RH: **Translational termination comes of age.** *Trends Biochem Sci* 2000, **25:**561-566.
49. Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES: **Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins.** *Cell* 2013, **154:**240-251.
50. Sun H, Chen C, Shi M, Wang D, Liu M, Li D, Yang P, Li Y, Xie L: **Integration of mass spectrometry and RNA-Seq data to confirm human ab initio predicted genes and lncRNAs.** *Proteomics* 2014, **14:**2760-2768.
51. Louro R, Smirnova AS, Verjovski-Almeida S: **Long intronic noncoding RNA transcription: Expression noise or expression choice?** *Genomics* 2009, **93:**291-298.
52. Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I: **Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species.** *Cell Rep* 2015, **11:**1110-1122.
53. Marques AC, Ponting CP: **Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness.** *Genome Biol* 2009, **10:**R124.
54. Pang KC, Frith MC, Mattick JS: **Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function.** *Trends Genet* 2006, **22:**1-5.
55. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP: **Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution.** *Cell* 2011, **147:**1537-1550.
56. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, Burch CL, Weeks KM: **Architecture and secondary structure of an entire HIV-1 RNA genome.** *Nature* 2009, **460:**711-716.
57. Hawkes EJ, Hennelly SP, Novikova IV, Irwin JA, Dean C, Sanbonmatsu KY: **COOLAIR Antisense RNAs Form Evolutionarily Conserved Elaborate Secondary Structures.** *Cell Rep* 2016, **16:**3087-3096.
58. Xue Z, Hennelly S, Doyle B, Gulati AA, Novikova IV, Sanbonmatsu KY, Boyer LA: **A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage.** *Mol Cell* 2016, **64:**37-50.
59. Jeon Y, Lee JT: **YY1 tethers Xist RNA to the inactive X nucleation center.** *Cell* 2011, **146:**119-133.
60. Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maier VK, Kingston RE, Lee JT: **High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation.** *Nature* 2013, **504:**465-469.
61. Wang J, Mager J, Chen Y, Schneider E, Cross JC, Nagy A, Magnuson T: **Imprinted X inactivation maintained by a mouse Polycomb group gene.** *Nat Genet* 2001, **28:**371-375.
62. Chaumeil J, Le Baccon P, Wutz A, Heard E: **A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced.** *Genes Dev* 2006, **20:**2223-2237.
63. Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY: **Systematic discovery of Xist RNA binding proteins.** *Cell* 2015, **161:**404-416.
64. McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, et al: **The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3.** *Nature* 2015, **521:**232-236.

65. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, et al: **Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression.** *Proc Natl Acad Sci U S A* 2009, **106:**11667-11672.
66. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, et al: **Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis.** *Nature* 2010, **464:**1071-1076.
67. Portoso M, Ragazzini R, Brenčič Ž, Moiani A, Michaud A, Vassilev I, Wassef M, Servant N, Sargueil B, Margueron R: **PRC2 is dispensable for HOTAIR-mediated transcriptional repression.** *EMBO J* 2017, **36:**981-994.
68. Chalei V, Sansom SN, Kong L, Lee S, Montiel JF, Vance KW, Ponting CP: **The long non-coding RNA Dali is an epigenetic regulator of neural differentiation.** *Elife* 2014, **3:**e04530.
69. Wang L, Zhao Y, Bao X, Zhu X, Kwok YK, Sun K, Chen X, Huang Y, Jauch R, Esteban MA, et al: **LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration.** *Cell Res* 2015, **25:**335-350.
70. Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J, Figueroa ME, De Figueiredo Pontes LL, Alberich-Jorda M, Zhang P, et al: **DNMT1-interacting RNAs block gene-specific DNA methylation.** *Nature* 2013, **503:**371-376.
71. Merry CR, Forrest ME, Sabers JN, Beard L, Gao XH, Hatzoglou M, Jackson MW, Wang Z, Markowitz SD, Khalil AM: **DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer.** *Hum Mol Genet* 2015, **24:**6240-6253.
72. Bao X, Wu H, Zhu X, Guo X, Hutchins AP, Luo Z, Song H, Chen Y, Lai K, Yin M, et al: **The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters.** *Cell Res* 2015, **25:**80-92.
73. Schmitz KM, Mayer C, Postepska A, Grummt I: **Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes.** *Genes Dev* 2010, **24:**2264-2269.
74. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C: **Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation.** *Mol Cell* 2008, **32:**232-246.
75. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P: **The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin.** *Science* 2008, **322:**1717-1720.
76. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY: **Long noncoding RNA as modular scaffold of histone modification complexes.** *Science* 2010, **329:**689-693.
77. Ding J, Xie M, Lian Y, Zhu Y, Peng P, Wang J, Wang L, Wang K: **Long noncoding RNA HOXA-AS2 represses P21 and KLF2 expression transcription by binding with EZH2, LSD1 in colorectal cancer.** *Oncogenesis* 2017, **6:**e288.
78. Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R: **Activating RNAs associate with Mediator to enhance chromatin architecture and transcription.** *Nature* 2013, **494:**497-501.
79. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, et al: **A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.** *Nature* 2011, **472:**120-124.
80. Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A, Macura K, Bläss G, Kellis M, Werber M, Herrmann BG: **The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse.** *Dev Cell* 2013, **24:**206-214.
81. Yang YW, Flynn RA, Chen Y, Qu K, Wan B, Wang KC, Lei M, Chang HY: **Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency.** *Elife* 2014, **3:**e02046.

82. Mayer C, Neubert M, Grummt I: **The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus.** *EMBO Rep* 2008, **9:**774-780.
83. Wehner S, Dörrich AK, Ciba P, Wilde A, Marz M: **pRNA: NoRC-associated RNA of rRNA operons.** *RNA Biol* 2014, **11:**3-9.
84. Schoeftner S, Sengupta AK, Kubicek S, Mechtler K, Spahn L, Koseki H, Jenuwein T, Wutz A: **Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing.** *EMBO J* 2006, **25:**3110-3122.
85. Yang L, Lin C, Liu W, Zhang J, Ohgi KA, Grinstein JD, Dorrestein PC, Rosenfeld MG: **ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs.** *Cell* 2011, **147:**773-788.
86. Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM: **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:**662-674.
87. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT: **Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome.** *Science* 2008, **322:**750-756.
88. Heo JB, Sung S: **Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA.** *Science* 2011, **331:**76-79.
89. Kaneko S, Bonasio R, Saldaña-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A, Reinberg D: **Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin.** *Mol Cell* 2014, **53:**290-300.
90. Marín-Béjar O, Marchese FP, Athie A, Sánchez Y, González J, Segura V, Huang L, Moreno I, Navarro A, Monzó M, et al: **Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2.** *Genome Biol* 2013, **14:**R104.
91. Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y: **Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene.** *Oncogene* 2011, **30:**1956-1962.
92. Ng SY, Johnson R, Stanton LW: **Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors.** *EMBO J* 2012, **31:**522-533.
93. Wang D, Ding L, Wang L, Zhao Y, Sun Z, Karnes RJ, Zhang J, Huang H: **LncRNA MALAT1 enhances oncogenic activities of EZH2 in castration-resistant prostate cancer.** *Oncotarget* 2015, **6:**41045-41055.
94. Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, Vergara IA, Davicioni E, Erho N, Ghadessi M, et al: **The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex.** *Nat Genet* 2013, **45:**1392-1398.
95. Wang Y, He L, Du Y, Zhu P, Huang G, Luo J, Yan X, Ye B, Li C, Xia P, et al: **The long noncoding RNA lncTCF7 promotes self-renewal of human liver cancer stem cells through activation of Wnt signaling.** *Cell Stem Cell* 2015, **16:**413-425.
96. Kawaguchi T, Tanigawa A, Naganuma T, Ohkawa Y, Souquere S, Pierron G, Hirose T: **SWI/SNF chromatin-remodeling complexes function in noncoding RNA-dependent assembly of nuclear bodies.** *Proc Natl Acad Sci U S A* 2015, **112:**4304-4309.
97. Wang X, Gong Y, Jin B, Wu C, Yang J, Wang L, Zhang Z, Mao Z: **Long non-coding RNA urothelial carcinoma associated 1 induces cell replication by inhibiting BRG1 in 5637 cells.** *Oncol Rep* 2014, **32:**1281-1290.
98. Wang S, Zhang X, Yuan Y, Tan M, Zhang L, Xue X, Yan Y, Han L, Xu Z: **BRG1 expression is increased in thoracic aortic aneurysms and regulates proliferation and apoptosis of vascular smooth muscle cells through the long non-coding RNA HIF1A-AS1 in vitro.** *Eur J Cardiothorac Surg* 2015, **47:**439-446.
99. Cajigas I, Leib DE, Cochrane J, Luo H, Swyter KR, Chen S, Clark BS, Thompson J, Yates JR, Kingston RE, Kohtz JD: **Evf2 lncRNA/BRG1/DLX1 interactions reveal RNA-dependent inhibition of chromatin remodeling.** *Development* 2015, **142:**2641-2652.
100. Hu G, Gong AY, Wang Y, Ma S, Chen X, Chen J, Su CJ, Shibata A, Strauss-Soukup JK, Drescher KM, Chen XM: **LincRNA-Cox2 Promotes Late Inflammatory Gene Transcription in Macrophages through Modulating SWI/SNF-Mediated Chromatin Remodeling.** *J Immunol* 2016, **196:**2799-2808.

101. Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP: **Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor.** *Sci Signal* 2010, **3:**ra8.
102. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, et al: **Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters.** *Nat Genet* 2011, **43:**621-629.
103. Sun S, Del Rosario BC, Szanto A, Ogawa Y, Jeon Y, Lee JT: **Jpx RNA activates Xist by evicting CTCF.** *Cell* 2013, **153:**1537-1551.
104. Rapicavoli NA, Qu K, Zhang J, Mikhail M, Laberge RM, Chang HY: **A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics.** *Elife* 2013, **2:**e00762.
105. Schmitt AM, Garcia JT, Hung T, Flynn RA, Shen Y, Qu K, Payumo AY, Peres-da-Silva A, Broz DK, Baum R, et al: **An inducible long noncoding RNA amplifies DNA damage signaling.** *Nat Genet* 2016, **48:**1370-1376.
106. Shamovsky I, Ivannikov M, Kandel ES, Gershon D, Nudler E: **RNA-mediated response to heat shock in mammalian cells.** *Nature* 2006, **440:**556-560.
107. Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD: **The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator.** *Genes Dev* 2006, **20:**1470-1484.
108. Berghoff EG, Clark MF, Chen S, Cajigas I, Leib DE, Kohtz JD: **Evf2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes.** *Development* 2013, **140:**4407-4416.
109. Ng SY, Bogu GK, Soh BS, Stanton LW: **The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis.** *Mol Cell* 2013, **51:**349-359.
110. Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R: **Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription.** *Nature* 2008, **454:**126-130.
111. Sharma S, Findlay GM, Bandukwala HS, Oberdoerffer S, Baust B, Li Z, Schmidt V, Hogan PG, Sacks DB, Rao A: **Dephosphorylation of the nuclear factor of activated T cells (NFAT) transcription factor is regulated by an RNA-protein scaffold complex.** *Proc Natl Acad Sci U S A* 2011, **108:**11381-11386.
112. Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McClanahan P, Hendrickson DG, Sauvageau M, Kelley DR, et al: **Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre.** *Nat Struct Mol Biol* 2014, **21:**198-206.
113. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell* 2014, **159:**1665-1680.
114. Splinter E, de Wit E, Nora EP, Klous P, van de Werken HJ, Zhu Y, Kaaij LJ, van Ijcken W, Gribnau J, Heard E, de Laat W: **The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA.** *Genes Dev* 2011, **25:**1371-1383.
115. Chen CK, Blanco M, Jackson C, Aznauryan E, Ollikainen N, Surka C, Chow A, Cerase A, McDonel P, Guttman M: **Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing.** *Science* 2016, **354:**468-472.
116. Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB: **An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles.** *Mol Cell* 2009, **33:**717-726.
117. Sasaki YT, Ideue T, Sano M, Mituyama T, Hirose T: **MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles.** *Proc Natl Acad Sci U S A* 2009, **106:**2525-2530.
118. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, et al: **The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation.** *Mol Cell* 2010, **39:**925-938.

119. West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev RI, Wang PI, Tolstorukov MY, Kingston RE: **The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites.** *Mol Cell* 2014, **55:**791-802.
120. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al: **Widespread transcription at neuronal activity-regulated enhancers.** *Nature* 2010, **465:**182-187.
121. Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, et al: **Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation.** *Nature* 2013, **498:**516-520.
122. Xiang JF, Yin QF, Chen T, Zhang Y, Zhang XO, Wu Z, Zhang S, Wang HB, Ge J, Lu X, et al: **Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus.** *Cell Res* 2014, **24:**513-531.
123. Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, et al: **Long noncoding RNAs with enhancer-like function in human cells.** *Cell* 2010, **143:**46-58.
124. Mousavi K, Zare H, Dell'orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager GL, Sartorelli V: **eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci.** *Mol Cell* 2013, **51:**606-617.
125. Maruyama A, Mimura J, Itoh K: **Non-coding RNA derived from the region adjacent to the human HO-1 E2 enhancer selectively regulates HO-1 gene induction by modulating Pol II binding.** *Nucleic Acids Res* 2014, **42:**13599-13614.
126. Schaukowitch K, Joo JY, Liu X, Watts JK, Martinez C, Kim TK: **Enhancer RNA facilitates NELF release from immediate early genes.** *Mol Cell* 2014, **56:**29-42.
127. Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, Jangi M, Giallourakis CC, Sharp PA, Young RA: **Transcription factor trapping by RNA in gene regulatory elements.** *Science* 2015, **350:**978-981.
128. Gong C, Maquat LE: **lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements.** *Nature* 2011, **470:**284-288.
129. Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K, Lee CS, Flockhart RJ, Groff AF, Chow J, et al: **Control of somatic tissue differentiation by the long non-coding RNA TINCR.** *Nature* 2013, **493:**231-235.
130. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C, et al: **Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat.** *Nature* 2012, **491:**454-457.
131. Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, Huarte M, Zhan M, Becker KG, Gorospe M: **LincRNA-p21 suppresses target mRNA translation.** *Mol Cell* 2012, **47:**648-655.
132. Liu B, Sun L, Liu Q, Gong C, Yao Y, Lv X, Lin L, Yao H, Su F, Li D, et al: **A cytoplasmic NF-kappaB interacting long noncoding RNA blocks IkappaB phosphorylation and suppresses breast cancer metastasis.** *Cancer Cell* 2015, **27:**370-381.
133. Hanahan D, Weinberg RA: **The hallmarks of cancer.** *Cell* 2000, **100:**57-70.
134. Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, Kominsky HD, et al: **Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression.** *Nat Biotechnol* 2011, **29:**742-749.
135. Prensner JR, Chen W, Iyer MK, Cao Q, Ma T, Han S, Sahu A, Malik R, Wilder-Romans K, Navone N, et al: **PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer.** *Cancer Res* 2014, **74:**1651-1660.
136. Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, Merkurjev D, Ohgi KA, Meng D, Zhang J, et al: **lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs.** *Nature* 2013, **500:**598-602.
137. Hung CL, Wang LY, Yu YL, Chen HW, Srivastava S, Petrovics G, Kung HJ: **A long noncoding RNA connects c-Myc to tumor metabolism.** *Proc Natl Acad Sci U S A* 2014, **111:**18697-18702.

138. Tseng YY, Moriarity BS, Gong W, Akiyama R, Tiwari A, Kawakami H, Ronning P, Reuland B, Guenther K, Beadnell TC, et al: **PVT1 dependence in cancer with MYC copy-number increase.** *Nature* 2014, **512:**82-86.
139. Kim T, Jeon YJ, Cui R, Lee JH, Peng Y, Kim SH, Tili E, Alder H, Croce CM: **Role of MYC-regulated long noncoding RNAs in cell cycle regulation and tumorigenesis.** *J Natl Cancer Inst* 2015, **107**.
140. Hu X, Feng Y, Zhang D, Zhao SD, Hu Z, Greshock J, Zhang Y, Yang L, Zhong X, Wang LP, et al: **A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer.** *Cancer Cell* 2014, **26:**344-357.
141. Leucci E, Vendramin R, Spinazzi M, Laurette P, Fiers M, Wouters J, Radaelli E, Eyckerman S, Leonelli C, Vanderheyden K, et al: **Melanoma addiction to the long non-coding RNA SAMMSON.** *Nature* 2016, **531:**518-522.
142. Panzitt K, Tschernatsch MM, Guelly C, Moustafa T, Stradner M, Strohmaier HM, Buck CR, Denk H, Schroeder R, Trauner M, Zatloukal K: **Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA.** *Gastroenterology* 2007, **132:**330-342.
143. Zhao Y, Guo Q, Chen J, Hu J, Wang S, Sun Y: **Role of long non-coding RNA HULC in cell proliferation, apoptosis and tumor metastasis of gastric cancer: a clinical and in vitro investigation.** *Oncol Rep* 2014, **31:**358-364.
144. Matouk IJ, Abbasi I, Hochberg A, Galun E, Dweik H, Akkawi M: **Highly upregulated in liver cancer noncoding RNA is overexpressed in hepatic colorectal metastasis.** *Eur J Gastroenterol Hepatol* 2009, **21:**688-692.
145. Peng W, Gao W, Feng J: **Long noncoding RNA HULC is a novel biomarker of poor prognosis in patients with pancreatic cancer.** *Med Oncol* 2014, **31:**346.
146. Lu Z, Xiao Z, Liu F, Cui M, Li W, Yang Z, Li J, Ye L, Zhang X: **Long non-coding RNA HULC promotes tumor angiogenesis in liver cancer by up-regulating sphingosine kinase 1 (SPHK1).** *Oncotarget* 2016, **7:**241-254.
147. Ji P, Diederichs S, Wang W, Böing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, et al: **MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer.** *Oncogene* 2003, **22:**8031-8041.
148. Gutschner T, Hämmerle M, Eissmann M, Hsu J, Kim Y, Hung G, Revenko A, Arun G, Stentrup M, Gross M, et al: **The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells.** *Cancer Res* 2013, **73:**1180-1189.
149. Liu B, Sun L, Liu Q, Gong C, Yao Y, Lv X, Lin L, Yao H, Su F, Li D, et al: **A cytoplasmic NF-κB interacting long noncoding RNA blocks IκB phosphorylation and suppresses breast cancer metastasis.** *Cancer Cell* 2015, **27:**370-381.
150. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, et al: **A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response.** *Cell* 2010, **142:**409-419.
151. Dimitrova N, Zamudio JR, Jong RM, Soukup D, Resnick R, Sarma K, Ward AJ, Raj A, Lee JT, Sharp PA, Jacks T: **LincRNA-p21 activates p21 in cis to promote Polycomb target gene expression and to enforce the G1/S checkpoint.** *Mol Cell* 2014, **54:**777-790.
152. Sánchez Y, Segura V, Marín-Béjar O, Athie A, Marchese FP, González J, Bujanda L, Guo S, Matheu A, Huarte M: **Genome-wide analysis of the human p53 transcriptional network unveils a lncRNA tumour suppressor signature.** *Nat Commun* 2014, **5:**5812.
153. Léveillé N, Melo CA, Rooijers K, Díaz-Lagares A, Melo SA, Korkmaz G, Lopes R, Akbari Moqadam F, Maia AR, Wijchers PJ, et al: **Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA.** *Nat Commun* 2015, **6:**6520.
154. Zhou Y, Zhong Y, Wang Y, Zhang X, Batista DL, Gejman R, Ansell PJ, Zhao J, Weng C, Klibanski A: **Activation of p53 by MEG3 non-coding RNA.** *J Biol Chem* 2007, **282:**24731-24742.
155. Pandey GK, Mitra S, Subhash S, Hertwig F, Kanduri M, Mishra K, Fransson S, Ganeshram A, Mondal T, Bandaru S, et al: **The Risk-Associated Long Noncoding RNA**

NBAT-1 Controls Neuroblastoma Progression by Regulating Cell Proliferation and Neuronal Differentiation. *Cancer Cell* 2014, **26:**722-737.
156. Mourtada-Maarabouni M, Pickard MR, Hedge VL, Farzaneh F, Williams GT: **GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer.** *Oncogene* 2009, **28:**195-208.
157. Poliseno L, Haimovic A, Christos PJ, Vega Y Saenz de Miera EC, Shapiro R, Pavlick A, Berman RS, Darvishian F, Osman I: **Deletion of PTENP1 pseudogene in human melanoma.** *J Invest Dermatol* 2011, **131:**2497-2500.
158. Yang F, Huo XS, Yuan SX, Zhang L, Zhou WP, Wang F, Sun SH: **Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis.** *Mol Cell* 2013, **49:**1083-1096.
159. Ouyang B, Bracken B, Burke B, Chung E, Liang J, Ho SM: **A duplex quantitative polymerase chain reaction assay based on quantification of alpha-methylacyl-CoA racemase transcripts and prostate cancer antigen 3 in urine sediments improved diagnostic accuracy for prostate cancer.** *J Urol* 2009, **181:**2508-2513; discussion 2513-2504.
160. Prensner JR, Zhao S, Erho N, Schipper M, Iyer MK, Dhanasekaran SM, Magi-Galluzzi C, Mehra R, Sahu A, Siddiqui J, et al: **RNA biomarkers associated with metastatic progression in prostate cancer: a multi-institutional high-throughput analysis of SChLAP1.** *Lancet Oncol* 2014, **15:**1469-1480.
161. Teschendorff AE, Lee SH, Jones A, Fiegl H, Kalwa M, Wagner W, Chindera K, Evans I, Dubeau L, Orjalo A, et al: **HOTAIR and its surrogate DNA methylation signature indicate carboplatin resistance in ovarian cancer.** *Genome Med* 2015, **7:**108.
162. Serghiou S, Kyriakopoulou A, Ioannidis JP: **Long noncoding RNAs as novel predictors of survival in human cancer: a systematic review and meta-analysis.** *Mol Cancer* 2016, **15:**50.
163. Garraway LA, Lander ES: **Lessons from the cancer genome.** *Cell* 2013, **153:**17-37.
164. Nowell PC: **The clonal evolution of tumor cell populations.** *Science* 1976, **194:**23-28.
165. Ashworth A, Lord CJ, Reis-Filho JS: **Genetic interactions in cancer progression and treatment.** *Cell* 2011, **145:**30-38.
166. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW: **Cancer genome landscapes.** *Science* 2013, **339:**1546-1558.
167. MITELMAN F: **CATALOG OF CHROMOSOME-ABERRATIONS IN CANCER.** *Cytogenetics and Cell Genetics* 1983, **36:**1-&.
168. Negrini S, Gorgoulis VG, Halazonetis TD: **Genomic instability--an evolving hallmark of cancer.** *Nat Rev Mol Cell Biol* 2010, **11:**220-228.
169. Burrell RA, McClelland SE, Endesfelder D, Groth P, Weller MC, Shaikh N, Domingo E, Kanu N, Dewhurst SM, Gronroos E, et al: **Replication stress links structural and numerical cancer chromosomal instability.** *Nature* 2013, **494:**492-496.
170. Schimke RT, Kaufman RJ, Alt FW, Kellems RF: **Gene amplification and drug resistance in cultured murine cells.** *Science* 1978, **202:**1051-1055.
171. Kaufman RJ, Brown PC, Schimke RT: **Amplified dihydrofolate reductase genes in unstably methotrexate-resistant cells are associated with double minute chromosomes.** *Proc Natl Acad Sci U S A* 1979, **76:**5669-5673.
172. Biedler JL, Spengler BA: **Metaphase chromosome anomaly: association with drug resistance and cell-specific products.** *Science* 1976, **191:**185-187.
173. Dulbecco R: **A turning point in cancer research: sequencing the human genome.** *Science* 1986, **231:**1055-1056.
174. Dickson D: **Wellcome funds cancer database.** *Nature* 1999, **401:**729.
175. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM, Network CGAR: **The Cancer Genome Atlas Pan-Cancer analysis project.** *Nat Genet* 2013, **45:**1113-1120.
176. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabé RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, et al: **International network of cancer genome projects.** *Nature* 2010, **464:**993-998.

177. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, et al: **The landscape of somatic copy-number alteration across human cancers.** *Nature* 2010, **463:**899-905.
178. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhang CZ, Wala J, Mermel CH, et al: **Pan-cancer patterns of somatic copy number alteration.** *Nat Genet* 2013, **45:**1134-1140.
179. Pathmanathan N, Bilous AM: **HER2 testing in breast cancer: an overview of current techniques and recent developments.** *Pathology* 2012, **44:**587-595.
180. Cappuzzo FM, PhD, Hirsch FRM, PhD, Bunn PAJM, Franklin WAM, Varella-Garcia MP: **EGFR Amplification in**
Lung Cancer: Time for Thawing the Frozen FISH? , vol. Volume 38. pp. 36-37: Oncology Times; 2016:36-37.
181. Hu L, Ru K, Zhang L, Huang Y, Zhu X, Liu H, Zetterberg A, Cheng T, Miao W: **Fluorescence in situ hybridization (FISH): an increasingly demanded tool for biomarker research and personalized medicine.** *Biomark Res* 2014, **2:**3.
182. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D: **Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors.** *Science* 1992, **258:**818-821.
183. Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, et al: **High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.** *Nat Genet* 1998, **20:**207-211.
184. Mantripragada KK, Buckley PG, de Ståhl TD, Dumanski JP: **Genomic microarrays in the spotlight.** *Drug Discov Today* 2004, **9:**S45-52.
185. Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, Haden K, Li J, Shaw CA, Belmont J, et al: **High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping.** *Genome Res* 2006, **16:**1136-1148.
186. Nowak D, Hofmann WK, Koeffler HP: **Genome-wide Mapping of Copy Number Variations Using SNP Arrays.** *Transfus Med Hemother* 2009, **36:**246-251.
187. Alkan C, Coe BP, Eichler EE: **Genome structural variation discovery and genotyping.** *Nat Rev Genet* 2011, **12:**363-376.
188. Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS: **A census of amplified and overexpressed human cancer genes.** *Nat Rev Cancer* 2010, **10:**59-64.
189. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G: **GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers.** *Genome Biol* 2011, **12:**R41.
190. Bignell GR, Greenman CD, Davies H, Butler AP, Edkins S, Andrews JM, Buck G, Chen L, Beare D, Latimer C, et al: **Signatures of mutation and selection in the cancer genome.** *Nature* 2010, **463:**893-898.
191. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, Vivanco I, Lee JC, Huang JH, Alexander S, et al: **Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma.** *Proc Natl Acad Sci U S A* 2007, **104:**20007-20012.
192. Catherine, Goodman C: *Pathology*
E-Book: Implications for the Physical Therapist. **Elsevier**
Health Sciences; 2014.
193. Didkowska J, Wojciechowska U, Mańczuk M, Łobaszewski J: **Lung cancer epidemiology: contemporary and future challenges worldwide.** *Ann Transl Med* 2016, **4:**150.
194. Travis WD: **Classification of lung cancer.** *Semin Roentgenol* 2011, **46:**178-186.
195. Herbst RS, Heymach JV, Lippman SM: **Lung cancer.** *N Engl J Med* 2008, **359:**1367-1380.
196. Harris K, Khachaturova I, Azab B, Maniatis T, Murukutla S, Chalhoub M, Hatoum H, Kilkenny T, Elsayegh D, Maroun R, Alkaied H: **Small cell lung cancer doubling time and its effect on clinical presentation: a concise review.** *Clin Med Insights Oncol* 2012, **6:**199-203.

197. Perez-Moreno P, Brambilla E, Thomas R, Soria JC: **Squamous cell carcinoma of the lung: molecular subtypes and therapeutic opportunities.** *Clin Cancer Res* 2012, **18:**2443-2451.
198. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, et al: **EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy.** *Science* 2004, **304:**1497-1500.
199. Network CGAR: **Comprehensive molecular profiling of lung adenocarcinoma.** *Nature* 2014, **511:**543-550.
200. Network CGAR: **Comprehensive genomic characterization of squamous cell lung cancers.** *Nature* 2012, **489:**519-525.
201. Wistuba II, Behrens C, Virmani AK, Mele G, Milchgrub S, Girard L, Fondon JW, Garner HR, McKay B, Latif F, et al: **High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints.** *Cancer Res* 2000, **60:**1949-1960.
202. Sen R, Baltimore D: **Multiple nuclear factors interact with the immunoglobulin enhancer sequences.** *Cell* 1986, **46:**705-716.
203. Oeckinghaus A, Ghosh S: **The NF-kappaB family of transcription factors and its regulation.** *Cold Spring Harb Perspect Biol* 2009, **1:**a000034.
204. Chen FE, Ghosh G: **Regulation of DNA binding by Rel/NF-kappaB transcription factors: structural views.** *Oncogene* 1999, **18:**6845-6852.
205. Huxford T, Ghosh G: **A structural guide to proteins of the NF-kappaB signaling module.** *Cold Spring Harb Perspect Biol* 2009, **1:**a000075.
206. Ozes ON, Mayo LD, Gustin JA, Pfeffer SR, Pfeffer LM, Donner DB: **NF-kappaB activation by tumour necrosis factor requires the Akt serine-threonine kinase.** *Nature* 1999, **401:**82-85.
207. Chen ZJ, Parent L, Maniatis T: **Site-specific phosphorylation of IkappaBalpha by a novel ubiquitination-dependent protein kinase activity.** *Cell* 1996, **84:**853-862.
208. Wertz IE, O'Rourke KM, Zhou H, Eby M, Aravind L, Seshagiri S, Wu P, Wiesmann C, Baker R, Boone DL, et al: **De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-kappaB signalling.** *Nature* 2004, **430:**694-699.
209. Trompouki E, Hatzivassiliou E, Tsichritzis T, Farmer H, Ashworth A, Mosialos G: **CYLD is a deubiquitinating enzyme that negatively regulates NF-kappaB activation by TNFR family members.** *Nature* 2003, **424:**793-796.
210. Ruland J: **Return to homeostasis: downregulation of NF-κB responses.** *Nat Immunol* 2011, **12:**709-714.
211. Hoffmann A, Levchenko A, Scott ML, Baltimore D: **The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation.** *Science* 2002, **298:**1241-1245.
212. Gilmore T, Gapuzan ME, Kalaitzidis D, Starczynowski D: **Rel/NF-kappa B/I kappa B signal transduction in the generation and treatment of human cancer.** *Cancer Lett* 2002, **181:**1-9.
213. Migliazza A, Lombardi L, Rocchi M, Trecca D, Chang CC, Antonacci R, Fracchiolla NS, Ciana P, Maiolo AT, Neri A: **Heterogeneous chromosomal aberrations generate 3' truncations of the NFKB2/lyt-10 gene in lymphoid malignancies.** *Blood* 1994, **84:**3850-3860.
214. Ohno H, Takimoto G, McKeithan TW: **The candidate proto-oncogene bcl-3 is related to genes implicated in cell lineage determination and cell cycle control.** *Cell* 1990, **60:**991-997.
215. Caamaño JH, Perez P, Lira SA, Bravo R: **Constitutive expression of Bc1-3 in thymocytes increases the DNA binding of NF-kappaB1 (p50) homodimers in vivo.** *Mol Cell Biol* 1996, **16:**1342-1348.
216. Karin M: **NF-kappaB as a critical link between inflammation and cancer.** *Cold Spring Harb Perspect Biol* 2009, **1:**a000141.

217. Boehm JS, Zhao JJ, Yao J, Kim SY, Firestein R, Dunn IF, Sjostrom SK, Garraway LA, Weremowicz S, Richardson AL, et al: **Integrative genomic approaches identify IKBKE as a breast cancer oncogene.** *Cell* 2007, **129:**1065-1079.
218. Pflueger D, Terry S, Sboner A, Habegger L, Esgueva R, Lin PC, Svensson MA, Kitabayashi N, Moss BJ, MacDonald TY, et al: **Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing.** *Genome Res* 2011, **21:**56-67.
219. Shen RR, Zhou AY, Kim E, O'Connell JT, Hagerstrand D, Beroukhim R, Hahn WC: **TRAF2 is an NF-κB-activating oncogene in epithelial cancers.** *Oncogene* 2015, **34:**209-216.
220. Starczynowski DT, Lockwood WW, Deléhouzée S, Chari R, Wegrzyn J, Fuller M, Tsao MS, Lam S, Gazdar AF, Lam WL, Karsan A: **TRAF6 is an amplified oncogene bridging the RAS and NF-κB pathways in human lung cancer.** *J Clin Invest* 2011, **121:**4095-4105.
221. Guttridge DC, Albanese C, Reuther JY, Pestell RG, Baldwin AS: **NF-kappaB controls cell growth and differentiation through transcriptional regulation of cyclin D1.** *Mol Cell Biol* 1999, **19:**5785-5799.
222. Perkins ND, Felzien LK, Betts JC, Leung K, Beach DH, Nabel GJ: **Regulation of NF-kappaB by cyclin-dependent kinases associated with the p300 coactivator.** *Science* 1997, **275:**523-527.
223. La Rosa FA, Pierce JW, Sonenshein GE: **Differential regulation of the c-myc oncogene promoter by the NF-kappa B rel family of transcription factors.** *Mol Cell Biol* 1994, **14:**1039-1044.
224. Kiriakidis S, Andreakos E, Monaco C, Foxwell B, Feldmann M, Paleolog E: **VEGF expression in human macrophages is NF-kappaB-dependent: studies using adenoviruses expressing the endogenous NF-kappaB inhibitor IkappaBalpha and a kinase-defective form of the IkappaB kinase 2.** *J Cell Sci* 2003, **116:**665-674.
225. Deveraux QL, Roy N, Stennicke HR, Van Arsdale T, Zhou Q, Srinivasula SM, Alnemri ES, Salvesen GS, Reed JC: **IAPs block apoptotic events induced by caspase-8 and cytochrome c by direct inhibition of distinct caspases.** *EMBO J* 1998, **17:**2215-2223.
226. Wang CY, Mayo MW, Korneluk RG, Goeddel DV, Baldwin AS: **NF-kappaB antiapoptosis: induction of TRAF1 and TRAF2 and c-IAP1 and c-IAP2 to suppress caspase-8 activation.** *Science* 1998, **281:**1680-1683.
227. Micheau O, Lens S, Gaide O, Alevizopoulos K, Tschopp J: **NF-kappaB signals induce the expression of c-FLIP.** *Mol Cell Biol* 2001, **21:**5299-5305.
228. Jin F, Liu X, Zhou Z, Yue P, Lotan R, Khuri FR, Chung LW, Sun SY: **Activation of nuclear factor-kappaB contributes to induction of death receptors and apoptosis by the synthetic retinoid CD437 in DU145 human prostate cancer cells.** *Cancer Res* 2005, **65:**6354-6363.
229. Campbell KJ, Rocha S, Perkins ND: **Active repression of antiapoptotic gene expression by RelA(p65) NF-kappa B.** *Mol Cell* 2004, **13:**853-865.
230. Old LJ: **Tumor necrosis factor.** *Sci Am* 1988, **258:**59-60, 69-75.
231. Green S, Dobrjansky A, Chiasson MA: **Murine tumor necrosis-inducing factor: purification and effects on myelomonocytic leukemia cells.** *J Natl Cancer Inst* 1982, **68:**997-1003.
232. Black RA, Rauch CT, Kozlosky CJ, Peschon JJ, Slack JL, Wolfson MF, Castner BJ, Stocking KL, Reddy P, Srinivasan S, et al: **A metalloproteinase disintegrin that releases tumour-necrosis factor-alpha from cells.** *Nature* 1997, **385:**729-733.
233. Vandenabeele P, Declercq W, Beyaert R, Fiers W: **Two tumour necrosis factor receptors: structure and function.** *Trends Cell Biol* 1995, **5:**392-399.
234. Natoli G, Costanzo A, Ianni A, Templeton DJ, Woodgett JR, Balsano C, Levrero M: **Activation of SAPK/JNK by TNF receptor 1 through a noncytotoxic TRAF2-dependent pathway.** *Science* 1997, **275:**200-203.
235. Devin A, Cook A, Lin Y, Rodriguez Y, Kelliher M, Liu Z: **The distinct roles of TRAF2 and RIP in IKK activation by TNF-R1: TRAF2 recruits IKK to TNF-R1 while RIP mediates IKK activation.** *Immunity* 2000, **12:**419-429.
236. Hershko A, Ciechanover A: **The ubiquitin system.** *Annu Rev Biochem* 1998, **67:**425-479.

237. Kirisako T, Kamei K, Murata S, Kato M, Fukumoto H, Kanie M, Sano S, Tokunaga F, Tanaka K, Iwai K: **A ubiquitin ligase complex assembles linear polyubiquitin chains.** *EMBO J* 2006, **25:**4877-4887.
238. Kanayama A, Seth RB, Sun L, Ea CK, Hong M, Shaito A, Chiu YH, Deng L, Chen ZJ: **TAB2 and TAB3 activate the NF-kappaB pathway through binding to polyubiquitin chains.** *Mol Cell* 2004, **15:**535-548.
239. Regamey A, Hohl D, Liu JW, Roger T, Kogerman P, Toftgard R, Huber M: **The tumor suppressor CYLD interacts with TRIP and regulates negatively nuclear factor kappaB activation by tumor necrosis factor.** *J Exp Med* 2003, **198:**1959-1964.
240. Kovalenko A, Chable-Bessia C, Cantarella G, Israël A, Wallach D, Courtois G: **The tumour suppressor CYLD negatively regulates NF-kappaB signalling by deubiquitination.** *Nature* 2003, **424:**801-805.
241. Harhaj EW, Dixit VM: **Deubiquitinases in the regulation of NF-κB signaling.** *Cell Res* 2011, **21:**22-39.
242. Zhang X, Choi PS, Francis JM, Imielinski M, Watanabe H, Cherniack AD, Meyerson M: **Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers.** *Nat Genet* 2015.
243. Tamborero D, Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Kandoth C, Reimand J, Lawrence MS, Getz G, Bader GD, Ding L, Lopez-Bigas N: **Comprehensive identification of mutational cancer driver genes across 12 tumor types.** *Sci Rep* 2013, **3:**2650.
244. Gaszner M, Felsenfeld G: **Insulators: exploiting transcriptional and epigenetic mechanisms.** *Nat Rev Genet* 2006, **7:**703-713.
245. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H: **The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.** *Nucleic Acids Res* 2014, **42:**D1001-1006.
246. Kim T, Cui R, Jeon YJ, Lee JH, Lee JH, Sim H, Park JK, Fadda P, Tili E, Nakanishi H, et al: **Long-range interaction and correlation between MYC enhancer and oncogenic long noncoding RNA CARLo-5.** *Proceedings of the National Academy of Sciences of the United States of America* 2014, **111:**4173-4178.
247. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, et al: **Genome-wide association study of prostate cancer identifies a second risk locus at 8q24.** *Nat Genet* 2007, **39:**645-649.
248. Hu R, Zhong P, Xiong L, Duan L: **Long Noncoding RNA Cancer Susceptibility Candidate 8 Suppresses the Proliferation of Bladder Cancer Cells via Regulating Glycolysis.** *DNA Cell Biol* 2017.
249. Buiting K, Nazlican H, Galetzka D, Wawrzik M, Gross S, Horsthemke B: **C15orf2 and a novel noncoding transcript from the Prader-Willi/Angelman syndrome region show monoallelic expression in fetal brain.** *Genomics* 2007, **89:**588-595.
250. Huang da W, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic Acids Res* 2009, **37:**1-13.
251. Brown JD, Lin CY, Duan Q, Griffin G, Federation AJ, Paranal RM, Bair S, Newton G, Lichtman AH, Kung AL, et al: **NF-kappaB directs dynamic super enhancer formation in inflammation and atherogenesis.** *Mol Cell* 2014, **56:**219-231.
252. Raskatov JA, Meier JL, Puckett JW, Yang F, Ramakrishnan P, Dervan PB: **Modulation of NF-kappaB-dependent gene transcription using programmable DNA minor groove binders.** *Proc Natl Acad Sci U S A* 2012, **109:**1023-1028.
253. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B: **A high-resolution map of the three-dimensional chromatin interactome in human cells.** *Nature* 2013, **503:**290-294.
254. Zhou A, Scoggin S, Gaynor RB, Williams NS: **Identification of NF-kappa B-regulated genes induced by TNFalpha utilizing expression profiling and RNA interference.** *Oncogene* 2003, **22:**2054-2064.
255. Tian B, Nowak DE, Brasier AR: **A TNF-induced gene expression program under oscillatory NF-kappaB control.** *BMC Genomics* 2005, **6:**137.

256. Sasatomi T, Suefuji Y, Matsunaga K, Yamana H, Miyagi Y, Araki Y, Ogata Y, Itoh K, Shirouzu K: **Expression of tumor rejection antigens in colorectal carcinomas.** *Cancer* 2002, **94:**1636-1641.
257. Suefuji Y, Sasatomi T, Shichijo S, Nakagawa S, Deguchi H, Koga T, Kameyama T, Itoh K: **Expression of SART3 antigen and induction of CTLs by SART3-derived peptides in breast cancer patients.** *Br J Cancer* 2001, **84:**915-919.
258. Yang D, Nakao M, Shichijo S, Sasatomi T, Takasu H, Matsumoto H, Mori K, Hayashi A, Yamana H, Shirouzu K, Itoh K: **Identification of a gene coding for a protein possessing shared tumor epitopes capable of inducing HLA-A24-restricted cytotoxic T lymphocytes in cancer patients.** *Cancer Res* 1999, **59:**4056-4063.
259. Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG: **A strategy for probing the function of noncoding RNAs finds a repressor of NFAT.** *Science* 2005, **309:**1570-1573.
260. Zhao L, Guo H, Zhou B, Feng J, Li Y, Han T, Liu L, Li L, Zhang S, Liu Y, et al: **Long non-coding RNA SNHG5 suppresses gastric cancer progression by trapping MTA2 in the cytosol.** *Oncogene* 2016, **35:**5770-5780.
261. Hou X, Wang L, Zhang L, Pan X, Zhao W: **Ubiquitin-specific protease 4 promotes TNF-α-induced apoptosis by deubiquitination of RIP1 in head and neck squamous cell carcinoma.** *FEBS Lett* 2013, **587:**311-316.
262. Fan YH, Yu Y, Mao RF, Tan XJ, Xu GF, Zhang H, Lu XB, Fu SB, Yang J: **USP4 targets TAK1 to downregulate TNFα-induced NF-κB activation.** *Cell Death Differ* 2011, **18:**1547-1560.
263. Xiao N, Li H, Luo J, Wang R, Chen H, Chen J, Wang P: **Ubiquitin-specific protease 4 (USP4) targets TRAF2 and TRAF6 for deubiquitination and inhibits TNFα-induced cancer cell migration.** *Biochem J* 2012, **441:**979-986.
264. Song EJ, Werner SL, Neubauer J, Stegmeier F, Aspden J, Rio D, Harper JW, Elledge SJ, Kirschner MW, Rape M: **The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome.** *Genes Dev* 2010, **24:**1434-1447.
265. Huarte M: **The emerging role of lncRNAs in cancer.** *Nat Med* 2015, **21:**1253-1261.
266. Akrami R, Jacobsen A, Hoell J, Schultz N, Sander C, Larsson E: **Comprehensive analysis of long non-coding RNAs in ovarian cancer reveals global patterns and targeted DNA amplification.** *PLoS One* 2013, **8:**e80306.
267. Wilkerson MD, Yin X, Hoadley KA, Liu Y, Hayward MC, Cabanski CR, Muldrew K, Miller CR, Randell SH, Socinski MA, et al: **Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types.** *Clin Cancer Res* 2010, **16:**4864-4875.
268. Timani KA, Liu Y, He JJ: **Tip110 interacts with YB-1 and regulates each other's function.** *BMC Mol Biol* 2013, **14:**14.
269. Vlasschaert C, Xia X, Coulombe J, Gray DA: **Evolution of the highly networked deubiquitinating enzymes USP4, USP15, and USP11.** *BMC Evol Biol* 2015, **15:**230.
270. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehar J, Kryukov GV, Sonkin D, et al: **The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.** *Nature* 2012, **483:**603-607.
271. Diez-Villanueva A, Mallona I, Peinado MA: **Wanderer, an interactive viewer to explore DNA methylation and gene expression data in human cancer.** *Epigenetics Chromatin* 2015, **8:**22.
272. Ichihara M, Murakumo Y, Masuda A, Matsuura T, Asai N, Jijiwa M, Ishida M, Shinmi J, Yatsuya H, Qiao S, et al: **Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities.** *Nucleic Acids Res* 2007, **35:**e123.
273. Gautier L CL, Bolstad BM and Irizarry RA: **affy—analysis of Affymetrix GeneChip data at the probe level.** *Bioinformatics* 2004, **20:** pp. 307–315.
274. **limma powers differential expression analyses for RNA-sequencing and microarray studies.**

ACKNOWLEDGEMENTS

I would like to start by thanking my thesis supervisor *Maite*, for the great opportunity to join her lab, for her academic guidance through the Ph.D. process, and for always encouraging me to explore new ideas.

I thank the University of Navarra and the Center for Applied Medical Research (CIMA) and acknowledge the Spanish Ministry of Economy and Competitivity (MINECO) for supporting me through the FPI predoctoral fellowship.

I would also like to thank *Erik* and his lab for giving me the opportunity to experience a different research perspective during my stay in Sweden.

It has been five years with a lot of experiences and lessons on the way, without the help of *Oskar* it would have been totally different. Thanks *Oskar* for showing me a different lifestyle. I would also like to express my gratitude to *Oskar's* family for always being there to help me in whatever I needed.

I thank Jovanna for all her technical support, but most importantly for always being there to listen to me. So many laughs will be difficult to forget.

I thank my lab mates, *Francesco* and *Ivan* for feeding me during months with our 'tupper club' deal and for all the scientific and non-scientific advices. Thanks *Ivan* for giving me shelter at your house/hotel during the last months. Thanks also to *Aina* and *Elena* for sharing with me the experience of being a Ph.D. student and the frustrating/fun moments at lab. I would also like to thank *Jacques* for helping me at the beginning of the project. Thank you all for making the lab a fun/great place to work.

Thanks to all of you who always had an advice, recommendation or smile to share with me, thank you all: *Celia, Cris C., Leire, Lola, Marisol, Marina, Naiara, Nere, Nerea M., Isa C*.

These Ph.D. years have been quite academically and personally challenging; without the help of many people it would have been even harder. Thanks to the running club, *Aitor, Jone, Iñaki* and *Oskar* for those evenings of training and all the races we ran together, I will always remember those moments. *Patri* thanks for those long honest chats and for always being there in the good and bad times.

I would like to express my gratitude to *Kasia* my best roommate and friend, you made of our flat in Calle Mayor a really nice place to live, where I felt like home. Thanks to my international group of friends, *Aislinn, Dany, Momo, Richard, Unai* with whom I shared many fun nights out in the Casco Viejo.

I would like to mention my gratitude to my good old-friends from bachelor degree: *Ana, Cynthia, Rocío* and *Rodrigo*; even though we took different paths, I will always remember those long nights studying together and day dreaming about our future plans. To *Gustavo P*. for being so patient while teaching and guiding me during the first year in a lab. To *Julian* for encouraging me to continue the scientific career, and to never give up. To my friends back in Mexico (*Alicia, Fabiola, Silvia, Nora*) for your true friendship that has survived no matter the time and distance.

Thanks *Emre* for your patience during all these years, for your unlimited support and for being an inspiration to enjoy the ride together.

I thank to my mom, for all her love and for teaching me there is always a bright side no matter how difficult life can get. To my dad for his endless support and for teaching me with his example, the meaning of hard work. To my brother *David* for showing me that never is enough and that everything in life is possible. I thank all my family for their support no matter the distance between us.

PUBLICATIONS

FAL1ing inside an Amplicon.
Athie A, Huarte M. Preview Cancer Cell. 2014 Sep 8;26(3):303-4. doi: 10.1016/j.ccr.2014.08.009.
PMID: 25203317
http://www.ncbi.nlm.nih.gov/pubmed/25203317

Genome-wide analysis of the human p53 transcriptional network unveils a lncRNA tumor suppressor signature.
Sánchez Y, Segura V, **Athie A**, Marín-Béjar O, Marchese FP, González J, Bujanda L, Guo S, Matheu A, Huarte M. Nat Commun. 2014 Dec 19;5:5812. doi: 10.1038/ncomms6812.
PMID: 25524025
https://www.ncbi.nlm.nih.gov/pubmed/25524025