Transcriptional characterization and functional analysis of long non-coding RNA/protein-coding gene pairs encoded in the human genome

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)
submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by
Anne-Susann Musahl
from Berlin

Berlin, 2015

The present work was carried out from March 2012 until April 2015 at the Max Planck Institute for Molecular Genetics under the supervision of Dr. Ulf Andersson Ørom.

1st Reviewer: Dr. Ulf Andersson Ørom
2nd Reviewer: Prof. Dr. Markus Wahl

Date of Defence: 10.07.2015

Für Küsschen und das Entchen.

Acknowledgements

I would like to thank my PhD advisor Dr. Ulf Ørom for giving me the opportunity to work in his lab and on this exciting project. Ulf provided me with guidance and support but also gave me the freedom to develop and follow own research ideas. I also want to thank Prof. Dr. Markus Wahl for reviewing this PhD thesis.

Many thanks goes to my friends and colleagues in the lab: Dubravka Vučićević, Thomas Conrad, Evgenia Ntini, Annita Louloupi, Alexander Kiefer, Julia Liz, Maja Gehre, Antonia Hilbig and Masha Kenda, you guys provided such a friendly and helpful working environment that enabled me to finish this PhD work. Special thanks goes to Dusa, Thomas and Evgenia for many insightful scientific discussions and their endless technical and moral support to pass all big and small problems that emerged in the last three years.

Also, this work would never have been possible without the help of my co-authors at the Max Planck Institute and in Paris and Villejuif (France), therefore I would like to thank Annalisa Marsico, Ruping Sun, Marcus W. Albrecht, Guido Kroemer, Oliver Kepp, Xing Huang and Sylvie Rusakiewicz.

Besonders dankbar bin ich meinen Eltern, für ihre dauerhafte Unterstützung und Ermutigung ohne die meine schulische und universitäre Ausbildung nicht möglich gewesen wäre.

Last, but certainly not least I would like to endlessly thank Marcus for always standing at my side and his continuous encouragement. Your love, friendship and support means everything to me.

Contents

Abbreviations

7SL	signal recognition particle RNA
AF9	ALL1-fused gene from chromosome 9
AFF	AF4/FMR2
APC	adenomatosis polyposis coli
APC/C	anaphase-promoting complex/cyclosome
ASO	antisense oligonucleotide
ATM	ataxia-telangiectasia mutated gene product
bp	base pair
BRCA1	breast cancer gene 1
BRE	TFIIB recognition element
BRMS1	breast cancer metastasis suppressor 1
CAGE	cap analysis of gene expression
CALR	calreticulin
cas	CRISPR-associated
CBP	CREB-binding protein
CBX6	chromobox protein homolog 6
CDK	cyclin-dependent kinase
CDKN1A	cyclin-dependent kinase inhibitor 1A
cDNA	complementary DNA
CDR1as	cerebellar degeneration-related protein 1 transcript
CGI	CpG islands
CHK2	checkpoint homologue 2
CIITA	class II, major histocompatibility complex, transactivator
conc	concentration
COPRS	coordinator of PRMT5
CPE	core promoter element
CPSF	cleavage and polyadenylation specificity factor
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSTF	cleavage stimulating factor
CTCF	CCCTC-binding factor

Continued on next page

Continued from previous page

C_t	threshold cycle
CXCL5	chemokine (C-X-C motif) ligand 5
DCE	downstream core elements
DEPC	diethylpyrocarbonate
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DSE	downstream sequence element
DSIF	DRB-sensitivity-inducing factor
dsiRNA	dicer-substrate RNA
EDTA	ethylenediaminetetraacetic acid
ENL	eleven-nineteen leukemia
EPCAM	epithelial cell adhesion molecule
eRNA	enhancer RNA
ERV	endogenous retrovirus
EZR	ezrin
FBS	fetal bovine serum
GO	gene ontology
GTF	general transcription factor
H3K27ac	histone 3 lysine 27 acetylation
H3K4me	histone 3 lysine 4 methylation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hESC	human embryonic stem cell
ICD	immunogenic cell death
IL6	interleukin 6
Inr	initiator
kb	kilobase pair
LINE	long interpersed nuclear element
LTR	long terminal repeat
$MgCl_2$	magnesium chloride
MLH1	mutL homolog 1
MM	mismatch
MSH2	mutS homolog 2
MTE	motif ten element
MTX	mitoxantrone
NaCl	sodium chloride
NaOH	sodium hydroxide
ncRNA	non-coding RNA
NELF	Negative elongation factor

Continued on next page

Continued from previous page

nt	nucleotide(s)
NTP	nucleoside triphosphate
OD	optical density
ORF	open reading frame
p300	E1A binding protein p300
Paf	polymerase associated factor
PAP	poly(A) polymerase
PAS	poly(A) signal
PBS	phosphate-buffered saline
PCG	protein coding gene
PCR	polymerase chain reaction
PIC	pre-initiation complex
Pol II	RNA polymerase II
poly(A)	polyadenylation
pRB	retinoblastoma protein
PRC2	polycomb repressive complex 2
PROMT	promoter upstream antisense transcript
P-TEFb	positive transcription elongation factor b
PTEN	phosphatase and tensin homolog
PVDF	polyvinylidene difluoride
qRT-PCR	quantitative real-time PCR
RACE	rapid amplification of cDNA ends
RB1	retinoblastoma 1
RNA	ribonucleic acid
RPKM	reads per kilobase per million reads
scaRNA	Cajal body-associated RNA
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SDS	sodium dodecylsulfate
SEC	super elongation complex
Ser	serine
SINE	short interpersed nuclear element
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
TBE buffer	TRIS-borate-EDTA buffer
TBP	TATA box-binding protein
TE	transposable element
TFII	transcription factor, RNA polymerase II
TIMP	tissue inhibitor of metalloproteinase

Continued on next page

Continued from previous page

TLS	translocated in liposarcoma
T_m	melting temperature
Tris-HCl	tris-(hydroxymethyl)-aminomethane hydrochloride
TSS	transcription start site
Uchl1	ubiquitin carboxy-terminal hydrolase L1
USE	upstream sequence element
UTR	unstranslated region

Abstract

Many protein-coding gene (PCG) promoters in the human genome initiate transcription in two directions, thereby expressing an mRNA and an upstream non-coding RNA (ncRNA). Diverse species of these promoter-associated ncRNAs are abundantly detected in genome-wide transcriptome studies but the functions of these non-coding transcripts remain mostly elusive. In this thesis, a set of 1,107 long ncRNA/PCG pairs that are expressed from bidirectional promoters is defined. These bidirectional promoters exhibit a high degree of sequence conservation and mediate linked expression of paired genes. This is determined by expression quantification and reporter assays of selected candidates.
Expression of these long ncRNA/PCG pairs is detected to frequently occur from promoters of cancer-related proteins. One of the bidirectional promoters mediates simultaneous expression of the tumor suppressor gene RB1 and ncRNA-RB1 as detected by assaying the effects of retinoblastoma-associated point mutations in a bidirectional reporter assay. The linked expression of both genes is further shown by mutation of core promoter elements residing in both promoter directions. Changes of single or few base-pairs, is found to affect transcription initiation in both promoter directions equally.
To determine the functionality of paired genes and their involvement in common biological pathways, ncRNA-RB1 and RB1 mRNA were individually depleted in a cell culture system. This revealed that both genes are not regulating each other's expression and that ncRNA-RB1 conveys regulatory effects that are different but also to a certain degree overlapping to the RB1 controlled transcriptional program. NcRNA-RB1 positively regulates the expression of calreticulin (CALR), an endoplasmic reticulum-sessile chaperone that can translocate to the surface of tumor cells after chemotherapy, thereby serving as an 'eat-me-signal' to phagocytes. Knock-down of the nuclear-retained ncRNA-RB1 in tumor cells reduces the expression of the CALR gene on chromatin, impairs translocation of the CALR protein to the cell surface upon treatment with anthracylines, and consequently inhibits uptake of the cells by macrophages. In conclusion, co-transcription of ncRNA-RB1 from the bidirectional RB1 promoter provides a positive link between the regulation of two the tumor suppressors RB1 and CALR. Loss of expression of either gene product of the ncRNA-RB1/RB1 pair entails the abolition of additional tumor-inhibitory mechanisms.

Zusammenfassung

Viele Promotoren Protein-kodierender Gene im menschlichen Genom initiieren die Transkription in zwei Richtungen und exprimieren dabei eine Boten-RNA (mRNA) sowie eine nicht-kodierende RNA (ncRNA), welche upstream des Promoters liegt. Verschiedene Varianten solcher Promoter-assoziierten, ncRNAs wurden kürzlich in Genom-weiten Transkriptionsstudien detektiert. Dennoch sind ihre Funktionen bisher weitgehend ungeklärt.

In dieser Doktorarbeit, wird ein Set bestehend aus 1107 Paaren langer ncRNAs (>200 bp) und Protein-kodierender Gene definiert, die von bidirektionellen Promotoren exprimiert werden. Entsprechende Promotoren weisen ein hohes Maß an Sequenzkonservierung auf und initiieren gleichzeitig die Expression von Genpaaren. Durch Quantifizierung der Expression und Verwendung von Reporter Assays für ausgewählte Kandidatengenpaare wurde dieses Verhalten nachgewiesen.

Viele dieser bidirektionellen Promotoren exprimieren Gene, die im Zusammenhang zur Entstehung von Tumoren stehen. Einer dieser Promotoren vermittelt die gleichzeitige Expression des Tumorsuppressors RB1 und der ncRNA-RB1. Dieses wird mittels eines Reporter-Assays gezeigt, welcher die Auswirkungen von Retinoblastoma-assoziierten Punktmutationen auf die Bidirektionionalität des Promotors nachvollzieht. Weiterhin wird die gekoppelte Expression der Genpaare durch Mutation von Core-Promoterelementen gezeigt, welche sich in beide Richtungen des bidirektionellen ncRNA-RB1/RB1 Promoters befinden. Dabei beeinflusste die artifizielle Veränderung einzelner oder einiger weniger Basenpaare die Transkriptionsinitiation in beide Promoterrichtungen.

Um die Funktionalität beider Gene eines Genpaares sowie ihre Beteiligung in gemeinsamen biologischen Stoffwechselwegen aufzuklären, wurden die ncRNA-RB1 und die RB1 mRNA einzeln inaktiviert. Dieser Versuch zeigte, dass beide Gene nicht gegenseitig ihre Expression beeinflussen und die ncRNA-RB1 regulatorische Effekte besitzt, die unterschiedlich von, andererseits aber auch überlappend mit der transkriptionellen Regulation durch RB1 sind. Unabhängig von RB1 beeinflusst die ncRNA-RB1 die Expression von Calreticulin (CALR), eines Chaperons des endoplasmatischen Retikulums, positiv.

Nach Behandlung mit spezifischen Chemotherapeutika kann CALR zur Zelloberfläche von Tumorzellen translozieren und dort als Fress-Signal für phagozytierende Zellen dienen. Der Knock-down der nukleären ncRNA-RB1 in Tumorzellen reduziert die Transkrip-

tion des CALR-Gens und verhindert nachfolgend die Translokation des CALR-Proteins zur Zelloberfläche als Auswirkung der Behandlung mit Anthracyclinen. Die Konsequenz daraus ist eine verhinderte Aufnahme der ncRNA-RB1 knock-down Zellen durch Makrophagen.

Als Ergebnis stellt die gleichzeitige Transkription von ncRNA-RB1 und RB1 von einem gemeinsamen bidirektionalen Promoter eine Verknüpfung zwischen der Regulation der zwei Tumorsuppressoren RB1 und CALR her. Der Verlust der Expression jedes Gens des Paares ncRNA-RB1/RB1 führt zur Beeinträchtigung Tumor unterdrückender Mechanismen in der Zelle.

Chapter 1

Introduction

1.1 Preface

Exact spatial and temporal expression of genes is fundamental for all biological processes of the cell, such as proliferation, differentiation, aging or apoptosis.
The first gene regulatory model was introduced by F. Jacob and J. Monod in 1961 and proposed that transcription initiation is controlled by the interaction of regulators with specific sequence elements in the DNA. These regulators were suggested to be represented by proteins or RNA, assuming that a regulator of gene expression might also be transcribed in order to fulfill its regulatory function [JACOB and MONOD, 1961]. This model proved to be very true, as transcriptional regulation by proteins, as well as by RNAs has been confirmed. More recently, the function of RNA as a regulator of gene expression has been elucidated and this discovery is gaining importance with the description and functional characterization of an increasing number of non-coding RNA (ncRNA) species [Derrien et al., 2012],[Ørom et al., 2010]. For decades transcriptional activation or repression of genes has been considered to be exclusively mediated by proteins, so-called transcription factors, and the additional regulatory functions of ncRNAs in the transcription process exemplifies the complexity of transcriptional regulation and its importance.
This introduction will give an overview on the principles and outcomes of this sophisticated process.

1.2 Encoding of genetic information in the human genome

The human genome consists of 3.2 billion base pairs (bp) organized into 23 chromosomes that are estimated to encode for a number of 60,483 genes including protein-coding genes, non-coding RNA genes and pseudogenes [The GENCODE Consortium, 2014].
In order to utilize the genetic information thereby building and organizing the various cell types present in the human body, the genes need to be expressed. The central dogma of molecular biology was established more than 50 years ago and describes the flow of

genetic information to be mostly unidirectional from DNA into RNA and into protein as the final functional product [CRICK, 1958],[Crick, 1970]. The biological processes responsible for the transfer of information between these macromolecules are termed 'transcription', describing the copy of DNA into RNA, and 'translation', descriptive for the conversion of RNA information into protein sequence. Although the dogma was accurate, it considered RNA to solely be the template for protein synthesis, a perception that was challenged by the discovery of many ncRNA species. These transcripts are encoded by their own genes and represent a final product of genetic information as they are not translated [Liu and Maxwell, 1990].

DNA as the coding form of genetic information in the genome is organized into chromatin allowing its compaction and regulation. The fundamental unit of chromatin is the nucleosome with 147 base pairs (bp) of DNA wrapped around an octamer of histone proteins in 1 3/4 superhelical turns [Finch et al., 1977]. The center of the histone octamer is built of two dimers of histones H3 and H4, surrounded by two dimers of histones H2A and H2B [Klug et al., 1980]. Histone H1 binds to the linker DNA in between the nucleosomes and is required to organize higher order chromatin structures thereby achieving higher compaction of the DNA [Bednar et al., 1998]. Besides organizing chromatin structure, the assembly of DNA with histones also regulates gene expression, as the presence of nucleosomes affects the accessibility and recognition of regulatory DNA sequences and the process of transcription itself. Additionally, posttranslational modifications at the unstructured core histone tails directly affect chromatin structure and the interaction of modifying factors with chromatin. These modifications include acetylation, methylation, phosphorylation, ubiquitylation and others and can be found at specific amino acids of the core histone tails [Bhaumik et al., 2007]. For example, trimethylation of histone 3 lysine 9 (H3K9me3) is a marker for the transcriptionally inactive and tightly compacted heterochromatin due to its recognition by the heterochromatin protein 1 [Bannister et al., 2001]. Methylations of histone 3 lysine 4, such as mono- (H3K4me1) and trimethylation (H3K4me3), are present in transcriptional active or accessible euchromatin and are indicative for regulatory sequences [Bhaumik et al., 2007]. Thus, transcription is controlled by DNA sequence but also depends highly on the mobilization of nucleosomes and modification of histones.

Altogether, the complex process of gene expression is not only regulated at all steps of the transcription process, including transcription initiation, elongation and termination but also during RNA maturation steps, at the level of translation and via post-translational modification of proteins. However, most regulation occurs at the stage of transcription initiation [Maston et al., 2006].

1.3 The transcription process

The machinery for the transcription of mRNA and many ncRNAs, consists of RNA polymerase II (Pol II) and of general transcription factors (GTFs) that mediate Pol II anchoring to promoters, DNA melting and transcription start site (TSS) recognition.
Transcription is initiated by the formation of a preinitiation complex (PIC) that contains Pol II, TFIIA (transcription factor, RNA polymerase II, A), TFIIB, TFIID, TFIIE, TFIIF, TFIIH and Mediator. At the core promoter, TFIID first interacts with its already bound subunit the TATA box-binding protein (TBP). This is followed by binding of TFIIA, thereby stabilizing TFIID–core promoter interactions. Then TFIIB associates with TBP and recruits the 12 subunit Pol II enzyme that is already in complex with TFIIF. Following TSS selection by Pol II, docking of TFIIE and TFIIH as recruited by the Mediator coactivator complex, completes PIC assembly [Esnault et al., 2008]. DNA melting is subsequently catalyzed by TFIIH (as review in [Juven-Gershon and Kadonaga, 2010],[Cheung and Cramer, 2012],[Grünberg and Hahn, 2013],[Kandiah et al., 2014]).
The initially abortively transcribing Pol II escapes the promoter when the nascent RNA has reached a length of 8-9 bp. Promoter escape is facilitated by phosphorylation of Pol II at Ser5 (serine5) within its C-terminal domain which is mediated by TFIIH [Ramanathan et al., 2001]. It facilitates dissociation of Pol II from the complex formed with Mediator [Søgaard and Svejstrup, 2007]. The transition of Pol II from initiation to elongation further results in dissociation of TFIIB from the pre-initiation complex. By functioning as an elongation factor, TFIIF can stay in association with Pol II, other GTFs such as TFIID, TFIIA, TFIIH and TFIIE as well as the Mediator complex remain associated with the core promoter, helping in subsequent re-initiation rounds by acting as a scaffold for re-initiation complex formation [Yudkovsky et al., 2000](as review in [Kwak and Lis, 2013],[Cheung and Cramer, 2012]). During early elongation, Pol II can also move backwards and this backtracking might be followed by transcriptional arrest. TFIIS reactivates arrested Pol II by stimulating the cleavage of nascent RNA [Bengal et al., 1991],[Cheung and Cramer, 2011].
In metazoans, early elongation of Pol II is paused at ~30% of actively transcribed genes, a process different from backtracking. Pol II pausing especially affects genes regulated by signaling pathways, e.g. developmental genes. It occurs between the promoter and the first (+1) nucleosome following processive transcription of 20-60 nucleotides (nt) of RNA [Guenther et al., 2007],[Muse et al., 2007],[Zeitlinger et al., 2007](as reviewed in [Adelman and Lis, 2012]). It involves interaction of the pausing factors NELF (Negative elongation factor) and DSIF (DRB-sensitivity-inducing factor) with Pol II [Yamaguchi et al., 1999]. Additional factors such as GDOWN1 and TFIIF have been associated with the stability of Pol II pausing [Cheng et al., 2012]. Promoter-proximal paused Pol II can continue productive elongation, however, transcription may also be terminated.
For pause release and maturation of Pol II into the elongating form, the kinase complex

P-TEFb (positive transcription elongation factor b) is recruited and mediates the phosphorylation of NELF and DSIF and also phosphorylates the C-terminal domain of Pol II at Ser2. As a consequence, NELF dissociates from the complex and DSIF continues to play the role of an elongation factor. Pol II phosphorylation at Ser2, allows for the recruitment of accessory elongation factors and stimulates mRNA processing [Czudnochowski et al., 2012](as reviewed in [Peterlin and Price, 2006]).

Resuming of elongation by Pol II is achieved by assembly of the Super Elongation Complex (SEC) that facilitates productive RNA synthesis. The SEC is assembled in interchangeable combinations of the AFF (AF4/FMR2) family members AFF1 and AFF4, the ELL (eleven-nineteen Lys-rich leukemia) family members ELL1, ELL2 and ELL3, of ENL (eleven-nineteen leukemia) and of AF9 (ALL1-fused gene from chromosome 9), but invariably contains P-TEFb. Furthermore, factors that mediate RNA-processing and modify chromatin also associate with Pol II during its progression through the gene body, such as TFIIS or the Paf1 (polymerase associated factor) complex (as reviewed in [Guo and Price, 2013]).

Processing of the nascent RNA occurs co-transcriptionally, with 5' capping being one of the first steps and realized during early elongation. Also, introns are spliced and polyadenylation (poly(A)) factors are recruited during productive elongation of Pol II (as reviewed in [Kwak and Lis, 2013],[Guo and Price, 2013]).

3' end formation and polyadenylation of the nascent RNA occurs when Pol II has passed a poly(A) signal (PAS) that has been selected to end the RNA message. The PAS is a 6 nt sequence motif with the consensus sequence AAUAAA which is preceded by upstream sequence elements (USEs) and followed by U- or GU-rich downstream sequence elements (DSEs). Recognition of the transcribed poly(A) signal and cleavage of the RNA is accomplished by the action of CPSF (cleavage and polyadenylation specificity factor) and CSTF (cleavage stimulating factor) complexes. The cleaved RNA is elongated by 250-300 untemplated adenosines through the action of poly(A) polymerase (PAP). The length of the poly(A) tail determines RNA stability and its presence allows nuclear export and translation of mRNAs (as reviewed in [Elkon et al., 2013],[Zheng and Tian, 2014],[Shatkin and Manley, 2000]).

Pol II is continuously transcribing after passing the poly(A)signal and termination of transcription by Pol II is achieved by degradation of the uncapped residual RNA while still being elongated. When exonuclease Xrn2 reaches the transcribing Pol II during its degradation process, termination of transcription is realized. This process is aided by slowing down Pol II at DNA encoded pause sites or by transcription of co-transcriptional cleavage sequences (as reviewed in [Proudfoot, 2011]).

1.4 Transcriptional regulatory motifs

Regulation of transcription is accomplished through cis-regulatory DNA sequences and binding of trans-acting factors to them. These regulatory sequences are referred to as promoters, enhancers, silencers or insulators and binding of activating or repressive transcription factors to them can positively or negatively affect the rate of transcription initiation (as reviewed in [Riethoven, 2010],[Maston et al., 2006]).
For example, transcriptional activation can be achieved by binding of sequence-specific activating transcription factors, e.g. ligand-dependent nuclear receptors, to such regulatory sequences. This binding can entail interaction of such sequence-specific factors with coactivator complexes, which do not possess DNA binding properties, such as the Mediator complex, the histone acetyltransferase p300 (E1A binding protein p300) or CBP (CREB-binding protein). The function of co-activator complexes often involves nucleosome remodeling or placement of activating histone modifications, thereby promoting PIC assembly, transcription elongation or re-initiation (as reviewed in [Rosenfeld et al., 2006]).
Transcriptional repression includes sequence-specific binding of repressors, such as unliganded or antagonist-bound nuclear receptors. This can be followed by their interaction with corepressor complexes that can antagonize the activity of coactivator complexes by e.g. positioning of histone marks that are repressive to transcription initiation (as reviewed in [Rosenfeld et al., 2006]).

1.4.1 Enhancers

Enhancers have classically been defined as DNA regions that positively influence the expression of target genes in an orientation-independent manner [Banerji et al., 1981]. Enhancers can be located up to several 100 kilobase pairs (kb) upstream and downstream of a gene's TSS, in unstranslated regions (UTRs), exons, introns or in intergenic regions [Lettice et al., 2003](as reviewed in [Bulger and Groudine, 2011]). Enhancer elements are characterized by the presence of high levels of the activating histone modification H3K4me1 and additionally by H3K27ac when the enhancer itself is active [E.N.C.O.D.E. Project Consortium et al., 2007],[Heintzman et al., 2009],[Creyghton et al., 2010]. At the same time, the activating histone mark H3K4me3 is mostly absent from enhancer sequences [E.N.C.O.D.E. Project Consortium et al., 2007]. Through the action of enhancers on their target promoters, highly tissue- or developmental stage-specific gene expression is assured [Amano et al., 2009]. On average, an enhancer has been observed to interact with 2.4 promoters and vice versa a promoter to associate with 4.9 enhancers [Andersson et al., 2014a].
The gene activating function of enhancers is accomplished by clustering of transcription factor binding sites thereby serving as platforms for and cooperative binding of tran-

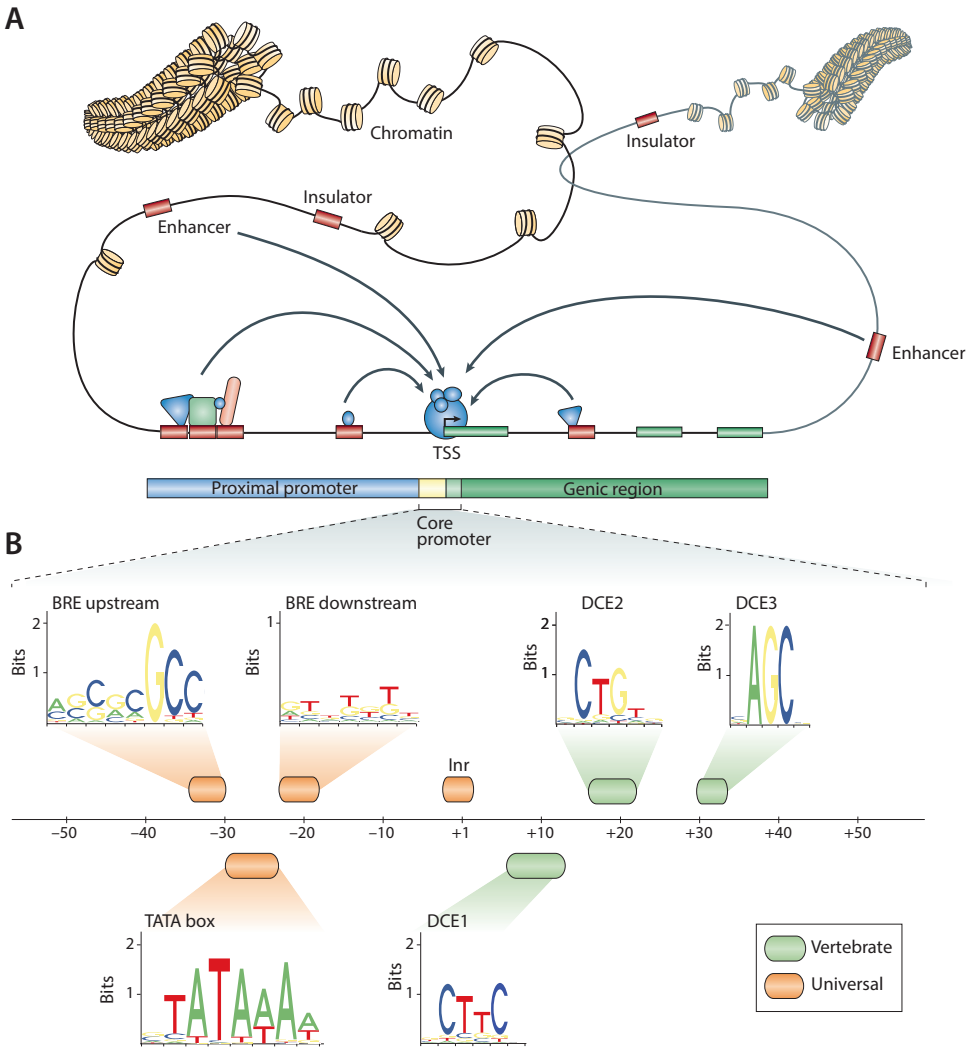

Figure 1.1: Transcriptional regulatory motifs. A. Regulatory sequences. DNA is compactly organized into chromatin by its wrapping around nucleosomes. It can be tightly organized or accessible to proteins as in active cis-regulatory DNA sequences. These regulatory sequences are referred to as promoters (including proximal and core promoters), enhancers, silencers or insulators and binding of activating or repressive transcription factors can positively or negatively affect the rate of transcription initiation of the TSS. **B.** Core promoter elements. The Pol II promoter is 50-100 nt in size and harbors several conserved sequence elements to initiate transcription at the TSS. Among them are the TATA box, Inr, BREu and BREd and DCE. Figure modified from [Lenhard et al., 2012].

scription factors. As a consequence, nucleosomes are excluded of from active enhancer regions [Stamatoyannopoulos et al., 1995],[Elgin, 1988]. Chromatin looping as mediated by the interaction of the transcriptional coactivator complexes such as Mediator and cohesion with enhancer-bound transcription factors and bring enhancer sequences and gene promoters in physical proximity [Kagey et al., 2010]. This positively influences PIC assembly at promoters [de Laat et al., 2008],[Heintzman et al., 2009]. More recently, these chromatin loopings have been suggested to be aided by long ncRNAs and their interaction with the Mediator complex [Lai et al., 2013].

1.4.2 Promoters

At the promoter, all regulation directed to a gene is converted into the rate of transcription initiation.
Promoters can be subdivided into proximal and core promoters, with proximal promoters being similar to enhancers in serving as transcription factor binding sites and with core promoters mediating the assembly of the PIC. Promoters vary greatly in the presence of specific regulatory elements and sequence motifs, generating a regulative complexity at the site of transcription initiation (as reviewed in [Riethoven, 2010],[Smale and Kadonaga, 2003]).

1.4.2.1 Proximal promoter

The upstream regulatory element of the promoter, typically extending from 50 bp up to 1 kb from the TSS, is referred to as the proximal promoter. It serves as binding site for specific transcription factors and integrates activating as well as repressive signals into the rate of transcription initiation at the core promoter. The proximal promoter region has been proposed to help tethering distal enhancers thereby conveying their activating effect to the core promoter (as reviewed in [Lenhard et al., 2012],[Maston et al., 2006]).

1.4.2.2 Core promoter

The RNA polymerase II core promoter is defined as the minimal DNA sequence sufficient to initiate transcription. This stretch of 50-100 nt harbors the TSS and functions as a platform to assemble the PIC. Several conserved sequence elements, so-called core promoter elements (CPE) have been identified in vertebrate core promoters due to their evolutionary conservation. Among these elements are the TATA box, Initiator (Inr), upstream and downstream TFIIB recognition element (BREu and BREd) and downstream core element (DCE). The motif ten element (MTE) has been described as an additional element, but its occurrence is less frequent. Additionally, the CCAAT box is conserved in core promoters but is not considered as a canonical CPE. The exact role of these CPEs in directing transcription initiations is mostly undefined due to their irregular distribution across promoters (as reviewed in [Smale and Kadonaga, 2003],[Kadonaga, 2012]).
TATA box and Inr elements are most frequently present in protein-coding gene (PCG) promoters, occurring often together but also separately. The initiator spans the TSS and its consensus sequence in the human genome has been determined as YYANWYY, with A being the first nucleotide to be transcribed (+1 nucleotide) [Javahery et al., 1994]. During PIC formation, the Inr is contacted by TFIID [Kaufmann and Smale, 1994]. For the many promoters lacking a Inr consensus motif, TSS selection is not random with the -1 and +1 nucleotides often being represented by a pyrimidine and a purine, respectively [Corden et al., 1980].
The TATA box is located between 28 and 34 nt upstream of the +1 nucleotide, with

a spacing of 30-31 nt being strongly preferred [Ponjavic et al., 2006]. It is bound by the TBP, subunit of TFIID during PIC formation [Patikoglou et al., 1999]. The TATA box consensus sequence has been found to be TATAT/AAAA/G [Ponjavic et al., 2006]. Although it is the best studied CPE, it only occurs in about 10% of mammalian core promoters [Carninci et al., 2006]. Many promoters without TATA box contain an Inr to direct accurate transcription initiation [Suzuki et al., 2001].

The two BRE motifs are located upstream and downstream of the TATA box, if present, and are bound by TFIIB during transcription initiation [Nikolov et al., 1995]. Their locations with respect to the TSS are at position ∼-35 and -20 (with respect to the motive midpoints), but promoters usually contain only one BRE element. Their consensus sequence is G/CG/CG/ACGCC and G/ATT/AT/GT/GT/GT/G for BREu and BREd, respectively [Lagrange et al., 1998]. The BRE motifs not only activate transcription, but are also be able to repress it [Evans et al., 2001].

Three different DCEs (DCE1-3) can be present in core promoters and are contacted by the TFIID complex during PIC assembly [Lewis et al., 2000]. These elements are located at positions ∼+9, +18 and +32 relative to the TSS and their consensus sequence has been determined to be CTTC, CTGT and AGC for DCE1-3, respectively [Lewis et al., 2000].

Diverse combinations as well as the presence or absence of certain CPEs in vertebrate core promoters allow for the integration of different transcriptional regulatory signals by recruiting different components of the transcription machinery. This possibility has been reported for the transcription of major histocompatibility complex (MHC) class I genes: different CPEs within the same promoter are used during constitutive and activated transcription, leading to the assembly of a canonical PIC (including TFIID) or non-canonical PIC (including the CIITA (class II, major histocompatibility complex, transactivator) and excluding TFIID), respectively [Howcroft et al., 2003].

As the presented elements are not universally found in core promoters and as about one third of promoters feature none of these CPEs, it has been suggested that more CPEs could be discovered [Kadonaga, 2012],[Roy and Singer, 2015].

In addition to CPEs, non-canonical promoter elements for Pol II promoters are increasingly recognized, such as CpG islands, ATG deserts and transcription initiation platforms (as reviewed in [Roy and Singer, 2015]).

CpG islands are regions in the DNA of 0.5-2 kb in length that exhibit a high density of cytosine nucleotides that reside directly next to guanosine nucleotides, thereby forming CpG dinucleotides. As CpG dinucleotides are substrates for DNA methylation, they are underrepresented in the genome due to the mutagenic properties of methylcytosine. Therefore, when occurring in a CpG island, these dinucleotides are unmethylated in all tissues and across all developmental stages. Most CpG islands are sites of transcription initiation owing to their destabilizing impact on nucleosomes and low nucleosome occupancy. About half of PCG promoters are associated with CpG islands, spanning the

proximal and core promoter. Thus, the presence of a CpG island was shown to be the best predictor for the existence of a gene (as reviewed in [Deaton and Bird, 2011]).
ATG deserts are non-canonical promoter elements mostly found in the context of TATA-less core promoters. ATG deserts span a DNA region of ± 1 kb around the TSS , are characterized by their depletion of ATG trinucleotides and occur independently of CpG islands [Lee et al., 2005].
Finally, transcription initiation platforms have been described as regulatory genomic sequences that are associated with promoters. These platforms span 0.4-10 kb in length, correlate with a high CpG content and recruit Pol II and GTFs, thereby overlapping with TSSs. Their presence is not specific to promoters and is also observed at enhancers [Koch et al., 2011].

1.4.2.3 Core promoter classes

Based on the presence of CPEs, histone modifications and transcriptional initiation patterns as were determined by genome-wide studies, a tripartition of metazoan core promoters has been suggested (as reviewed in [Lenhard et al., 2012]).
Type I promoters are characterized by their low CpG content and frequent occurrence of TATA-boxes and Inr-like sequences. Usage of a discrete TSS or of several TSSs but within a cluster of a few nucleotides, is a feature of this promoter class, also referred to as focused promoters. Selection of a defined TSS by the PIC is mediated by TBP binding to an available TATA-box. Tissue-specific genes are expressed from type I promoters, which are further characterized by an H3K4me3 pattern solely downstream of the TSS. This promoter class is majorly regulated by sequence modules residing close to the TSS [Ernst and Kellis, 2010],[Carninci et al., 2006].
Type II core promoters harbor short CpG islands around their TSS and initiate transcription of ubiquitously expressed genes and housekeeping genes. Several Inr elements can be found within core promoters of this class, so that TSS selection is dispersed and transcription initiation occurs within a region of up to 150 nt. The H3K4me3 histone modification is usually only present at the 5' end of the gene and overlaps the CpG island at this site. Nucleosome positioning around the TSSs is precise and regulation of this promoter type is achieved by the action of only few enhancers [Akalin et al., 2009],[Carninci et al., 2006].
Type III promoters mediate expression of developmental genes and harbor several large CpG islands that extend into the gene body. Besides the presence of H3K4me3, these promoters are simultaneously marked by H3K27me3, due to the binding of polycomb group proteins (PcG proteins). Both histone modifications are widely distributed across the promoter and gene body, indicating repression and activation so that this promoter type is also referred to as bivalent [Bernstein et al., 2006]. TSS selection from type III promoters is not specific but more focused than for type II promoters with the nucleo-

some positioning being also discrete. Regulation of this promoter class is mediated by numerous enhancers [Carninci et al., 2006].
In summary, two thirds of vertebrate promoters lack a distinct TSSs and instead initiate transcription within a window of 50-150 nt as has been determined by CAGE (cap analysis of gene expression)-based methods. In general, transcription initiation tends to be more focused in highly regulated genes. Intermediates between the focused and dispersed promoters exist, e.g. in promoters with several TSSs but one preferred initiation site (as reviewed in [Sandelin et al., 2007]). Additionally, transition between different TSS selection modes has been observed during early vertebrate development: In oocytes, TATA-like sequences are used for TSS selection in the absence of nucleosome positioning. Subsequently, TSS selection switches to a fixed position from the well-positioned first downstream nucleosome, suggesting that Inr-like sequences determine TSS selection at a later developmental stage [Haberle et al., 2014].

1.4.3 Silencers

Silencers confer negative regulation to the transcription initiation process and function in an orientation independent manner analogous to enhancers. These elements can be located within a proximal promoter region, be part of an enhancer region or represent an independent regulatory sequence. Mechanistically, silencers provide binding sites for repressive protein complexes, so called repressors, which can then interact with corepressors (as reviewed in [Maston et al., 2006]). Binding of repressors to certain DNA sequences interferes with binding of activators or GTFs, thereby mediating the repressive effect of silencers when they are located in the proximal promoter region [Harris et al., 2005],[Perissi et al., 2004]. Also, recruited co-repressors can possess histone modifying activity and thereby generate a repressive chromatin structure, e.g. when silencers are located more distal from the TSS [Srinivasan and Atchison, 2004].

1.4.4 Insulators

Insulators divide the genome into regulatory units by restricting the activity of enhancers and silencers to a specific set of proximal genes. They are position-dependent but orientation independent regulatory elements and have an average size of few kb (as reviewed in [Maston et al., 2006]). Unwanted interactions of cis-regulatory elements with promoters are prevented by enhancer-blocking activity when the insulator is located in between an enhancer and a promoter [Recillas-Targa et al., 2002]. In vertebrates, this functions is mediated by binding of CTCF (CCCTC-binding factor) [Bell et al., 1999]. Also, insulators function as barriers to the spread of heterochromatin and reside at the border of eu- and heterochromatin [Recillas-Targa et al., 2002].

1.5 Transcription of the human genome

Of the 3.2 billion base pairs representing the human genome, only about 1.5% are protein-coding [Lander et al., 2001]. At the same time, 75% of the genomic sequence has been found to be transcribed into RNA across different cell lines, with 25% being attributable to genomic output from protein-coding genes when including intronic sequences. In one cell line on average about 40% of the genome is covered by primary transcripts. Mapping these transcripts to the genome visualizes how the intergenic space is reduced and how neighboring transcription units increasingly overlap [Djebali et al., 2012]. The term 'pervasive transcription' has been used to refer to the variety of RNA species, apart from protein encoding RNAs and ncRNAs with well-established functions, that are often low abundant and now being detected using high-throughput deep-sequencing techniques. Many of these ncRNA species are detected at regulatory sequences such as promoters and enhancers. Thereby, a common theme is the bidirectional initiation of transcription at both DNA strands from these regulatory sequences, yielding a variety of ncRNA species (as reviewed in [Jensen et al., 2013]).

1.5.1 Bidirectional expression of PCGs

Initially, a bidirectional promoter was described as a stretch of DNA driving the expression of two PCGs that are encoded on opposite DNA strands and arranged in a head-to-head divergent orientation. The intervening promoter sequence was defined to be less than 1,000 bp in size and suggested to initiate transcription in both directions [Trinklein et al., 2004]. Early work established that more than 10% of human genes are encoded in a bidirectional conformation, an arrangement often conserved among mouse orthologs [Adachi and Lieber, 2002],[Trinklein et al., 2004]. This bias for divergently encoded PCGs is unique to mammalian genomes, when compared to organisms with similar genome size [Wakano et al., 2012]. For many of these gene pairs in the human genome, the intervening promoter sequence is short, more precisely less than 300 base pairs in size, and characterized by a high frequency of CpG islands [Adachi and Lieber, 2002],[Trinklein et al., 2004]. Sequence analysis of bidirectional promoters revealed that TATA box elements are underrepresented in comparison to unidirectional promoters, that drive the expression of one annotated gene, most likely due to an enrichment in CpG islands [Yang and Elnitski, 2008]. On the other hand, the CCAAT box sequence, a promoter element occurring 75-80 bp upstream of the Inr, is more prevalent in bidirectional promoter sequences [Yang and Elnitski, 2008]. With regard to transcription factor binding sites, GABPA, MYC, E2F1, E2F4, NRF-1, YY1 consensus sequences are enriched in bidirectional promoters, while the majority of motifs that are found in vertebrates are underrepresented [Lin et al., 2007]. Indeed, CCATT boxes, GAPBA, NRF-1 and YY1 binding sites are among the most conserved promoter motifs, bridging back to

the conservation of the bidirectional PCG arrangement [Xie et al., 2005].
Bidirectionally encoded PCGs are enriched in specific biological functions such as DNA repair, cell cycle regulation and regulation of metabolism [Adachi and Lieber, 2002]. In this context, expression of many of these paired genes has been found to be correlated, with genes being either simultaneously expressed or showing anti-correlated expression [Trinklein et al., 2004]. These non-directional expression patterns indicate that various regulatory mechanisms act on bidirectional promoters. Furthermore, genes expressed from bidirectional promoters are strongly associated with breast and ovarian cancer, due to their enrichment in DNA repair genes and consequent association with genome stability [Yang et al., 2007]. The high frequency of CpG islands in bidirectional promoters favors the inactivation of these genes trough DNA methylation, a major mechanism through which tumor suppressor genes are inactivated in cancer [Yang et al., 2007].

1.5.2 Inherent promoter bidirectionality

The recognition that many PCG promoters lack classical promoter elements and defined TSSs, but initiate transcription across a region of hundred base pairs, was recently expanded by the observation that bidirectional transcription initiation occurs at 50-80% of human promoters and generate transcripts that commence in the antisense direction relative to the mRNA TSS, even though no gene is annotated in this orientation [Duttke et al., 2015],[Core et al., 2008],[Core et al., 2014]. Many genome-wide studies in a wide range of organisms as distant as yeast and human established that eukaryotic PCG promoters also mediate expression of diverse-oriented upstream ncRNAs besides their expression of a PCG, a phenomenon termed divergent transcription [Core et al., 2008],[Preker et al., 2008],[Seila et al., 2008],[Neil et al., 2009]. These RNAs are generated through assembly of two independent PICs, one at each TSS of divergently transcribing promoters [Venters and Pugh, 2013].
In human, upstream antisense RNAs (uaRNAs) are grouped by their length into transcription start site-associated RNAs (<100 nt), which are attributable to abortive pausing and backtracking of Pol II in both promoter directions, as well as promoter upstream transcripts (PROMTs) (>100 nt) [Taft et al., 2009],[Seila et al., 2008],[Preker et al., 2008]. PROMTs are a class of mostly unstable RNAs as suggested by their low levels of expression and due to their stabilization upon depletion of the cellular exosome complex, the major eukaryotic 3'-5' exoribonuclease [Core et al., 2008],[Preker et al., 2008],[Seila et al., 2008]. Their emergence has long been suggested to be due to nucleosome depletion at active PCG promoters leading to aberrant transcription initiation of RNA polymerase II. This theory is supported by the observations that bidirectionality of promoters can be suppressed by nucleosome remodeling, histone deacetylation or gene loop formation [Whitehouse et al., 2007],[Marquardt et al., 2014],[Tan-Wong et al., 2012].
PROMTs share characteristics with mRNAs such as 3'polyadenylation and 5'capping

[Flynn et al., 2011],[Preker et al., 2011],[Ntini et al., 2013]. The smaller size of these non-coding transcripts compared to mRNAs is explainable by a higher frequency of poly(A) signals and depletion of splice site-like sequences in the upstream direction of PCG promoters. 5' splice sites of pre-mRNAs are bound by U1 snRNP, a component of the spliceosome, during the co-transcriptional splicing reaction thereby preventing the recognition of alternative poly(A) sites. Due to the low frequency of splice sites, upstream antisense transcripts expressed from active promoters are terminated quickly by cleavage of the nascent RNA and subsequent polyadenylation, using the same machinery as for the polyadenylation of mRNAs. Opposed to this, the higher frequency of U1 snRNP binding sites in the genic direction of a promoter suppresses PAS-mediated RNA cleavage and early polyadenylation, allowing for the expression of full-length mRNAs [Ntini et al., 2013],[Almada et al., 2013]. Although polyadenylation increases RNA stability and half-life, short polyadenylated transcripts are more prone to exosome-mediated degradation than longer transcripts, explaining the unstable nature of PROMTs [Andersson et al., 2014b].

Recently, the emergence of promoter antisense transcripts was attributed to the presence of reverse oriented core promoter elements that are similar to their counterparts in the PCG direction. At a preferred distance of 100-200 nucleotides in between divergent TSSs of bidirectional promoters, Inr-like and TATA-box-like elements have been identified at the edges of well-positioned nucleosomes and overlaying the TSSs of PROMTs [Duttke et al., 2015],[Core et al., 2014]. Unidirectional promoters that do not express upstream antisense RNAs, are depleted of reverse core promoter elements and do not show defined nucleosome positioning in the upstream promoter direction. This suggests that divergent antisense transcription initiation is not due to promiscuous transcription initiation of Pol II but due to the presence of core promoter-like elements that favor PIC assembly [Duttke et al., 2015].

It has been proposed that in evolutionary terms, divergent transcription initiation could fuel the emergence of new functional genes, by exposing the transcribed DNA strand to mutagenic alterations. Higher frequencies of C→T transitions and A→G transitions are observable in the non-template strand, thereby transcribed DNA regions can gain GT-rich sequences such as splice sites. The acquisition of splice sites would consequentially favor the origination of longer, more stable transcripts that could then acquire functionality [Wu and Sharp, 2013].

1.5.3 Transcription at enhancers

Assembly of the transcription machinery and Pol II recruitment to active enhancers in combination with transcription of associated enhancer regions has been determined by several genome-wide studies [Kim et al., 2010],[De Santa et al., 2010],[Koch et al., 2011]. Production of these transcripts, termed enhancer RNAs (eRNAs), correlates well with

characteristic enhancer modifications at chromatin such as H3K4me1, H3K27ac and p300 binding [Core et al., 2014]. eRNAs were found to be either short, non-polyadenylated and bidirectionally expressed but also long, polyadenylated and unidirectional [Kim et al., 2010],[Venters and Pugh, 2013],[Ørom et al., 2010]. For bidirectionally expressed eRNAs, the divergent TSS are separated by 110 bp on average and core promoter elements, such as TATA boxes and Inr elements, are present at similar frequencies at enhancer TSSs as in PCG promoters, thus explaining transcription initiation [Core et al., 2014],[Andersson et al., 2014a]. The expression level of eRNAs correlates with levels of mRNA synthesis at neighboring genes implying eRNA involvement in transcriptional activation [Kim et al., 2010]. Additional studies established that eRNA production precedes the culmination of target gene transcription and that target gene activation can be lost when eRNAs are specifically depleted from cells [Li et al., 2013],[Schaukowitch et al., 2014]. This functionality of eRNAs could be due to the involvement of these transcripts in enhancer-promoter looping, and/or facilitating release of paused Pol II into productive elongation by their interaction with the NELF complex [Li et al., 2013],[Schaukowitch et al., 2014]. Also, analogous to transcription from PCG promoters, only one of the bidirectionally expressed transcripts could be functional, with the other being a side product of bidirectional transcription initiation [Lam et al., 2013]. In conclusion, expression of eRNAs can be used as determinant for the activity status of enhancers, with active enhancers expressing eRNAs and non-active or poised enhancers being characterized by the H3K4me1 and/or H3K27ac histone modification patterns [Core et al., 2014].

1.5.4 Similarity of transcription initiation at promoters and enhancers

Promoters and enhancers share many characteristics in mammalian genomes although they have been classically defined as distinct transcriptional regulatory elements. Similarities include the binding of transcription factors to both elements, the existence of core promoter elements at similar frequencies and initiation of divergent transcription through assembly of distinct PICs at the paired CPEs [Core et al., 2014],[Venters and Pugh, 2013],[Andersson et al., 2014a]. Transcription at divergent TSS TSSs in promoters and enhancers, initiates within a comparable distance of 100 - 200 bp and at positioned nucleosomes flanking both TSSs [Duttke et al., 2015],[Core et al., 2014]. A classic distinction between promoter and enhancers has been drawn through the presence of different histone modifications with promoters being characterized by a higher H3K4me3/H3K4me1 ratio [Heintzman et al., 2009]. Recently, this feature has been explained by the unequal transcription levels at both elements with H3K4me3 levels scaling up to promoter levels at highly transcribed enhancers [Core et al., 2014]. These findings argue that the classical distinctions between enhancers and promoters are not sufficient to categorize these elements and that promoters and enhancers are much more similar than ever assumed.

1.6 ncRNA species encoded in the human genome

Although pervasive transcription of the genome has only recently been recognized due to the development of sensitive high throughput sequencing technologies, the existence of functional ncRNAs is established for several decades (as reviewed in [Morris and Mattick, 2014]). In general, the catalog of functional ncRNAs is growing continuously [Cech and Steitz, 2014].

1.6.1 RNAs functioning in protein synthesis

Ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) are among the longest known functional ncRNA species [Cech and Steitz, 2014]. rRNAs represent the RNA component of the eukaryotic ribosome that consists of a large and small subunit including three (5S, 5.8S, 28S) and one RNA species (18S), respectively [Alberts et al., 2002],[Johnson and Alberts, 2002]. The rRNA species in cooperation with accessory proteins mediate protein synthesis from an mRNA template, with amino acids being carried to the ribosome by tRNAs [Alberts et al., 2002],[Johnson and Alberts, 2002]. In the human genome, 544 rRNA genes and 497 tRNA genes are encoded, and additionally 2 rRNA genes and 22 tRNA genes are encoded in the human mitochondrial DNA [The GENCODE Consortium, 2014],[Derrien et al., 2012].

1.6.2 RNAs functioning in RNA processing

During eukaryotic mRNA maturation, introns are removed from primary transcripts by the spliceosome [Alberts et al., 2002],[Johnson and Alberts, 2002]. The five small nuclear RNAs (snRNAs) U1, U2, U4, U5 and U6 snRNA associate with more than 200 different proteins to form the spliceosome [Valadkhan, 2005]. Introns represent about 20% of the non-coding transcription in the genome, however they can be processed into other RNA species, e.g. small nucleolar RNAs (snoRNAs) [Cech and Steitz, 2014]. SnoRNAs direct chemical modifications including methylation and pseudouridylation of pre-rRNAs, thereby being involved in the maturation of this RNA species [Cech and Steitz, 2014]. Cajal body-associated (sca)RNAs are functionally similar to snoRNAs but located in the Cajal bodies. These mediate the modification of the spliceosomal RNAs [Cech and Steitz, 2014]. It is estimated that the human genome encodes 1,896 snRNA genes and 961 snoRNA genes [The GENCODE Consortium, 2014].

1.6.3 RNAs functioning in regulation of RNA expression

Besides the regulation by proteins, RNA expression and stability is also controlled by small RNA species. One class of these small RNAs are microRNAs (miRNAs) of ~22 nt in length that frequently base pair in the 3'UTR of mRNAs resulting in mRNA deadenylation and translational repression due to their interaction with Argonaute proteins (as

reviewed in [Ameres and Zamore, 2013]). Another class of small RNAs encoded in the human genome and functional in regulation of RNA stability are piwi-interacting RNAs (piRNAs) of ~27 nt in length. This species interacts with PIWI proteins, a subclade of Argonaute proteins, and degrades expressed transposable elements in the germline (as reviewed in [Luteijn and Ketting, 2013]). In the human genome, 4,093 miRNA genes are annotated [The GENCODE Consortium, 2014].

1.6.4 RNAs with diverse functions - long ncRNAs

Transcripts >200 nt in size that lack coding potential, as determined by the absence of large open reading frames and codon conservation, have been classified as long ncRNAs [Derrien et al., 2012],[Morris and Mattick, 2014]. These ncRNAs are functionally diverse and can control various biological processes such as X-chromosome inactivation, imprinting or transcriptional gene activation [Penny et al., 1996],[Leighton et al., 1995],[Ørom et al., 2010]. Recent studies suggest that about three times more long ncRNAs are encoded in the human genome than protein-coding genes, with estimates ranging to as much as 60,000 long ncRNA genes [Iyer et al., 2015]. As different long ncRNA data sets show moderate overlap with each other, ranging between 30-40%, it is likely that the long ncRNA transcriptome is still incompletely captured and that the numbers of long ncRNA genes might further increase [Derrien et al., 2012].

1.7 Functionality of long ncRNAs

Long ncRNAs are a heterogeneous group of transcripts, solely distinguished from other ncRNA species by their larger size of >200 nt. Due to this loose defining criterion for long ncRNAs as a class, the absence of a common functional mechanism of action is explainable.

1.7.1 Classification of long ncRNAs

The GENCODE gene annotation subclassifies long ncRNAs into five categories depending on their location and orientation with respect to the nearest PCG [Derrien et al., 2012]:

1. Intergenic long ncRNAs: do not intersect any PCG locus
2. Exonic antisense long ncRNAs: intersect an exon of a PCG locus on the opposite DNA strand
3. Intronic long ncRNAs: reside within an intron of a PCG on the same or opposite strand, but without any overlap with PCG exons
4. Overlapping long ncRNAs: contain a PCG within one of their introns on the same or opposite strand
5. Processed transcripts: do not harbor an open reading frame and do not fit in any of the other subcategories

However, no intrinsic functional differences between these ncRNA categories have yet been demonstrated [Morris and Mattick, 2014].

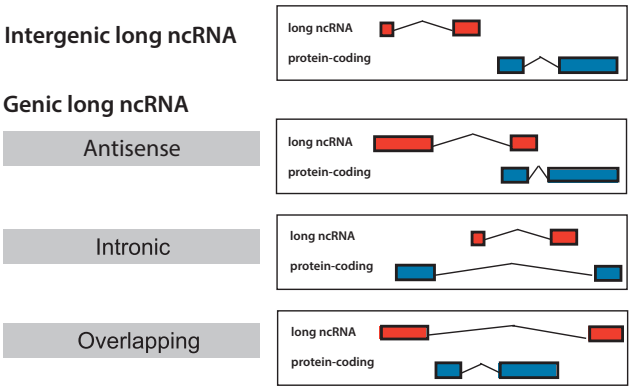

Figure 1.2: Subclassification of long ncRNAs. Long ncRNAs are classified based on their location and orientation to the closest PCG. Intergenic long ncRNAs do not intersect any PCG locus. Genic long ncRNAs intersect a PCG locus either on the same or the opposite DNA strand. Antisense long ncRNAs intersect an exon of a PCG locus on the opposite DNA strand. Intronic long ncRNAs reside within an intron of a PCG and overlapping long ncRNAs contain a PCG within one of their introns. Figure modified from [Derrien et al., 2012].

1.7.2 Characteristics of long ncRNAs

Many long ncRNA genes have a similar genomic structure as PCGs. The promoters of both long ncRNAs and PCGs show histone modifications indicative of transcriptional activity such as H3K4me2, H3K4me3, H3K9ac or H3K27ac [Derrien et al., 2012]. Long ncRNA transcripts also share characteristics with mRNAs, such as splicing, capping and polyadenylation. However, long ncRNAs overall differ from mRNAs in their lower expression level, lower number of exons and higher tissue-specificity [Cabili et al., 2011],[Derrien et al., 2012]. More precisely, long ncRNAs show on average a ~10-fold lower median maximal expression level than mRNAs [Cabili et al., 2011],[Derrien et al., 2012]. This has been suggested to be either caused by less efficient transcription of long ncRNA genes or more efficient degradation of the transcripts, e.g owing to their lack of open reading frames, rendering them prone to decay mechanisms such as the nonsense-mediated decay (NMD) [Ulitsky and Bartel, 2013]. A strong bias for two-exon transcripts has been described for long ncRNAs, however, these early studies neglected single-exon long ncRNAs in order to distinguish lowly expressed long ncRNAs from transcriptional noise [Derrien et al., 2012]. However, it is becoming increasingly clear that many long ncRNAs are unspliced single-exon transcripts and that long ncRNAs as a group are characterized by the presence of few exons [Iyer et al., 2015]. Although having few exons, long ncRNAs show high tendencies for alternative splicing [Cabili et al., 2011],[Derrien et al., 2012],[Iyer et al., 2015].

Brain and testis show especially high expression levels of long ncRNAs, with about one

third of long ncRNAs being specifically expressed in testis [Cabili et al., 2011],[Derrien et al., 2012]. Specific expression of long ncRNAs has also been observed in differentiated systems, such as T cells, muscles or breast tissue as well as in disease states such as cancer [Morris and Mattick, 2014],[Iyer et al., 2015]. Long ncRNAs can be located predominantly in the nucleus or in the cytoplasm, depending on the transcript, however, compared to mRNAs, long ncRNAs show higher enrichment in the nucleus [Ulitsky and Bartel, 2013]. Within the nucleus, these transcripts are particularly found in the chromatin-associated fraction [Derrien et al., 2012].

Exonic sequences of long ncRNA genes are on average significantly less conserved during evolution than those of PCGs, however some conservation is observable when compared to neutrally evolving sequences such as ancestral repeats [Morris and Mattick, 2014],[Ørom et al., 2010]. In general, promoters of long ncRNA genes exhibit higher conservation rates than their exonic sequences [Derrien et al., 2012],[Ørom et al., 2010]. However, a wide range of evolutionary conservation is observable for different long ncRNA species, so that ultraconserved genes exist as well as primate-specific ones [Morris and Mattick, 2014],[Necsulea et al., 2014]. In numbers, about 400 long ncRNA genes are conserved to the common ancestor of tetrapods and therefore originated more than 300 million years ago, but estimates suggest a far higher number of 11,000 primate-specific long ncRNAs [Necsulea et al., 2014]. This indicates that many long ncRNAs may have originated due to lineage-specific adaptive radiation [Morris and Mattick, 2014]. However, a lack of primary sequence conservation not necessarily indicates a lack of function, as RNA secondary structures, the main functional constraint of long ncRNAs, can also be retained by complementary base exchanges [Johnsson et al., 2014].

Many long ncRNAs likely evolved from transposable element (TE) insertions, as ~80% of long ncRNA genes contain TE sequences, such as degenerated versions as ERVs (endogenous retroviruses), LINEs (long interpersed nuclear elements), SINEs (short interpersed nuclear elements), and LTRs (long terminal repeats), and as TEs comprise about 40% of ncRNA sequences [Derrien et al., 2012],[Kelley and Rinn, 2012]. This high TE content is in contrast to protein coding gene sequences, in which TEs overlap only 6% of sequences [Kelley and Rinn, 2012].

Concerning the low expression and low evolutionary conservation levels of long ncRNAs, questions towards their functionality has been raised. However, the precision of long ncRNA expression, across tissues and the existence of alternative splice forms, as well as the many reported functions of individual long ncRNAs, argue for a functionality of many of these transcripts [Morris and Mattick, 2014],[Tsai et al., 2010],[Lai et al., 2013]. In this line of evidence, genome wide association studies showed that more than 80% of cancer-associated single nucleotide polymorphisms are found in noncoding regions that are transcribed into long ncRNAs [Cheetham et al., 2013].

1.7.3 Functions of long ncRNAs

The reported increasing numbers of long ncRNAs, due to advances in high-throughput sequencing techniques, is in opposition to the relatively few examples of functionally characterized long ncRNAs.

Early reports on ncRNA functionality date back about 20 years and describe the regulatory mechanisms of well-expressed ncRNAs such as Xist (X-inactive specific transcript) or H19 in X-chromosome inactivation and imprinting [Penny et al., 1996],[Leighton et al., 1995]. Since then the spectrum of ncRNA functional involvement has extended to various biological processes such as development, pluripotency or the p53 response pathway [Rinn et al., 2007],[Guttman et al., 2009],[Huarte et al., 2010]. A reoccurring theme is the regulation of gene expression by long ncRNAs resulting in activation or repression of target genes involved in specific processes (as reviewed in [Rinn and Chang, 2012]). As long ncRNAs display low expression levels and high degrees of tissue specificity, catalytic roles in their regulation of gene expression can be suggested.

Molecular mechanisms for long ncRNA action involves their interaction with proteins, DNA and RNA (as reviewed in [Rinn and Chang, 2012]):

In that way, long ncRNAs can act as decoys for proteins such as transcription factors, by interacting with their respective DNA-binding domain thereby modulating transcription factor docking to DNA. An example for this is the long ncRNA PANDA that interacts with the transcription factor NF-YA to reduce expression of pro-apoptotic genes under conditions of DNA damage [Hung et al., 2011].

Also, the guidance of proteins to certain genomic locations has been determined as a long ncRNA mechanism of action. This DNA targeting can occur through RNA-DNA base pairing or long ncRNA interaction with a DNA binding protein. Xist RNA interacts with the polycomb repressive complex 2 (PRC2) and recruits it to the X chromosome to establish the inactivating chromatin mark H3K27me3 throughout the dimension of this chromosome [Plath et al., 2003].

Long ncRNA functioning as scaffolds for protein complex formation has also been reported thereby bringing different proteins into physical proximity. This is exemplified by the long ncRNA HOTAIR and its binding of PRC2 and the LSD1/CoREST/REST complex, which results in H3K27 methylation and H3K4 demethylation and in silencing of the targeted HOXD gene [Tsai et al., 2010].

Long ncRNAs have additionally been identified to play a role in the establishment of chromatin structure, more specifically in enhancer-promoter interactions for example by binding to the Mediator complex [Lai et al., 2013]. Thus these transcripts can facilitate transcription initiation.

Besides interaction with proteins and DNA, base pairing of long ncRNAs with complementary RNAs allows for gene regulation [Cech and Steitz, 2014]. The long ncRNA 'antisense Uchl1' that is encoded antisense to the Uchl1 (ubiquitin carboxy-terminal hy-

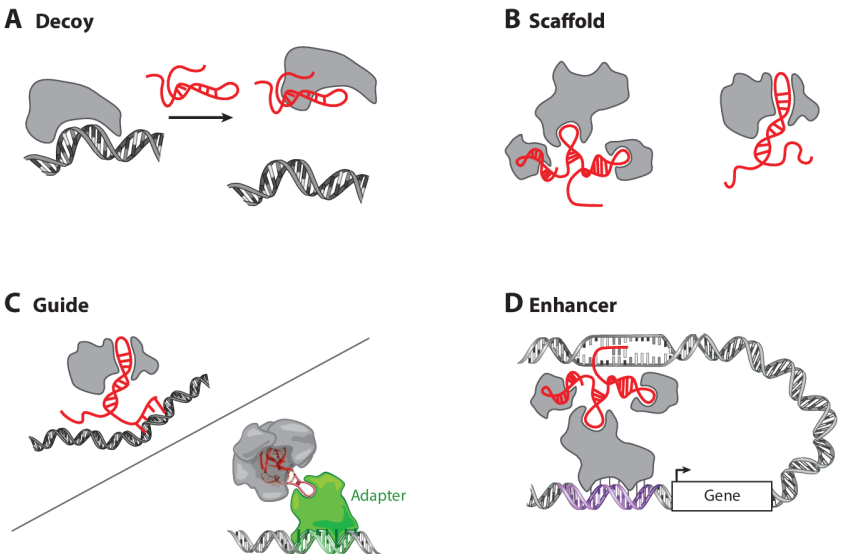

Figure 1.3: Molecular mechanisms of long ncRNA action. Long ncRNAs can act as decoys for proteins, as scaffolds for protein complex formation, as guides for proteins to be recruited to certain genomic locations and in the establishment of chromatin interactions such as in enhancer-promoter interactions. Figure modified from [Rinn and Chang, 2012].

drolase L1) gene, base pairs at the 5' end with Uchl1 mRNA. Association of Uchl1 mRNA with polysomes in the cytoplasm, resulting in increased protein synthesis and is achieved via 5' pairing with antisense Uchl1 RNA due to its possession of a SINE element [Carrieri et al., 2012].

Yet another mechanism of action for long ncRNAs is to work as competing endogenous RNAs for miRNAs. The circular RNA molecule CDR1as (cerebellar degeneration-related protein 1 transcript) harbors 63 conserved binding sites for miRNA miR-7 and binds miR-7 in neuronal tissues, thereby regulating midbrain development [Memczak et al., 2013]. This example, as well as the preceding one, extends the post-transcriptional regulative repertoire of long ncRNAs towards PCGs.

Another long ncRNA functionality is transcriptional interference, whereby not the RNA product is active but the act of its transcription serves a regulatory function [Cech and Steitz, 2014]. For example, transcriptional silencing of the Igf2r promoter by the Airn transcript only requires transcriptional overlap of the ncRNA gene with this promoter to induce its parental-specific silencing and imprinting [Latos et al., 2012].

The various functions of long ncRNAs can either be accomplished in cis (at their site of transcription) or trans (when diffusing to other loci), depending on the precise molecular mechanism and the expression level of the transcript [Cech and Steitz, 2014]. A working mechanism in trans has been assigned to several long ncRNAs interacting with chromatin regulatory proteins and includes also the translational enhancer activity of antisense Uchl1 RNA [Guttman et al., 2011],[Carrieri et al., 2012]. Long ncRNA expressed at only few copies per cell were found to function predominantly in cis, for example activating long ncRNAs that are involved in chromatin-looping and that interact with the

Mediator complex [Lai et al., 2013]. Thereby, the highly tissue-specific expression pattern of long ncRNAs is consistent with their functional role in tissue-specific regulation of chromatin structure, transcription factor binding or enhancer activity.
The cellular dependence on long ncRNAs for these diverse processes can be explained by the fact that RNA can bridge larger distances in comparison to protein chains. An α-helix of 50 amino acids extends for only 7.5 nm whereas a 50 base pair RNA arm extends for 13 nm, thereby an RNA is better suited to organize several protein binding motives. Also expression of RNAs results in decreased cellular metabolic costs compared to the production of proteins [Cech and Steitz, 2014].

1.7.4 Association of long ncRNAs with PCG promoters

Initially, research has been focused on the functional characterization of long ncRNAs originating from genomic regions that do not overlap with PCGs, as such transcripts facilitate experimental manipulation and computational analysis [Wang et al., 2011],[Ørom et al., 2010],[Guttman et al., 2009]. However, early estimates suggested that 35% of annotated long ncRNAs intersect loci of PCGs [Derrien et al., 2012]. Also, an increasing number of non-coding transcripts is detected upstream of PCG promoters (as described in Section 1.5.2). One study catalogued long ncRNAs expressed in human embryonic stem cells (hESCs) and suggested that >60% of long ncRNAs are divergently transcribed from PCGs (within a window of ±2 kb around the PCG TSSs). Half of these transcripts only consist of one exon and are nonspliced [Sigova et al., 2013]. Different numbers of PCG-associated long ncRNAs determined in these studies, can be explained by the consideration or disregard of single-exon ncRNAs as well as by varying filtering criteria [Sigova et al., 2013],[Derrien et al., 2012].
However, these results raise questions towards the correlation of long ncRNA and respective PCG expression as well as towards the functionality of these PCG-associated transcripts [Seila et al., 2008],[Preker et al., 2008].
Correlations of expressions between neighboring long ncRNAs and PCGs were yielding contradictory results, as on the one hand suggesting higher positive correlation of expression between long ncRNAs and neighboring PCGs [Derrien et al., 2012],[Ørom et al., 2010], on the other hand none such relationship was found [Cabili et al., 2011].
Another study focused on the expression of long ncRNAs around the promoters of 56 cell-cycle genes under diverse perturbation conditions, such as DNA damage or oncogenic stimuli [Hung et al., 2011]. This analysis revealed that sets of ncRNAs expressed in a genomic neighborhood are responsive to the respective stimuli and that their expression is correlated. However, the expression of long ncRNAs and their closest PCG was not found to be correlated [Hung et al., 2011].
Also, the expression of long ncRNAs, bidirectionally transcribed to PCGs in the mouse cerebral cortex, was found to be correlated with tissue-specific mRNAs. It was therefore

concluded that the function of these transcripts would be to activate their paired PCG [Uesaka et al., 2014].

The above mentioned study on the hESC transcriptome observed that PCG-associated long ncRNAs show coordinated changes in expression with their PCG partners when cells are differentiated into endoderm, suggesting that these ncRNA/PCG pairs are co-ordinately regulated. This suggests similar functional roles for bidirectionally expressed ncRNA/PCG pairs [Sigova et al., 2013].

Few studies investigated the functionality of long ncRNAs transcribed upstream of PCG promoters and indicate that the molecular mechanism of action of these transcripts is not uniform and not specific to the upstream PCG [Wang et al., 2008],[Grote et al., 2013],[Hung et al., 2011]:

Several long ncRNAs species are detected 300-1,500 bp upstream of the cyclin D1 gene and these are induced upon DNA damage signals. The transcripts recruit the RNA binding protein TLS (translocated in liposarcoma) to the cyclin D1 promoter and allow TLS interaction with CBP and with p300. Gene specific repression of cyclin D1 transcription is then achieved through the inhibition of p300 function by TLS [Wang et al., 2008].

In the mouse genome, the long ncRNA Fendrr (Fetal-lethal noncoding developmental regulatory RNA) is encoded 1,250 bp upstream of the TSS of the transcription factor-coding gene Foxf1. Fendrr and Foxf1 mRNA are co-expressed in mesodermal tissue. Fendrr binds the PRC2 complex and tethers it to its target promoters, the promotors of transcription factors Foxf1 and Pitx2, resulting in H3K27 trimethylation and transcriptional silencing of both genes [Grote et al., 2013],[Grote and Herrmann, 2013].

Additionally, the long ncRNA PANDA (P21 associated ncRNA DNA damage activated) is located 5 kb upstream of the TSS of CDKN1A (Cyclin-dependent kinase inhibitor 1A) and encoded in a divergent fashion to the CDKN1A locus. Induction of PANDA upon DNA damage is coordinated with the induction of CDKN1A mRNA and depends on the binding of p53 about 2.5 kb upstream and in the midpoint between both genes. PANDA does not regulate expression of CDKN1A but it binds to the transcription factor NF-YA, thereby interfering with NF-YA binding to DNA and with its activation of apoptotic genes. The action of the CDKN1A gene product p21 mediates cell cycle arrest upon DNA damage. As simultaneous expression of PANDA blocks apoptosis induction, a linkage between both genes is potentially advantageous for cancer progression [Hung et al., 2011].

In conclusion, the divergent encoding of two genes can allow co-regulation and co-expression of these genes, for example by a shared upstream transcriptional network. The binding of transcription factors within a shared promotor region then mediates transcription initiation of the paired genes. When both gene products are coordinately expressed, they can contribute to the same cellular response independently, as is the case of PANDA and p21 [Hung et al., 2011].

1.8 Human tumor suppressors

Through the diverse mechanisms of action of long ncRNAs, these transcripts can play regulatory roles in cancer initiation and progression, as cancer is a disease of aberrant gene expression. In this way, long ncRNAs can guide chromatin-modifying complexes, resulting in epigenetic changes and gene expression changes. Functionally characterized long ncRNAs have been found to be mis-expressed in cancers (as reviewed in [Cheetham et al., 2013]). Additionally, mutations in non-coding or regulatory sequences of the genome such as enhancers can play a causative role in the development of cancer [Lee and Young, 2013],[Lubbe et al., 2012].

Loss or inactivation of tumor suppressor genes promotes the initiation or progression of cancer and can result from genic mutations or DNA methylation, with the latter being especially effective on CpG islands containing bidirectional promoters [Sun and Yang, 2010],[Wakano et al., 2012]. Due to the resulting loss of function or loss of expression of affected tumor suppressor genes, cells lose the ability to control their proliferation [Sun and Yang, 2010]. Adding to the complex regulatory control of tumor suppressor genes, many of these genes have been found to harbor a nearby antisense RNA [Yu et al., 2008]. Also, genes strongly associated with cancer are frequently expressed from bidirectional promoters [Yang et al., 2007].

Tumor suppressors represent a diverse group of molecules that inhibit the development of cancer by four major mechanisms: inhibition of cell division, induction of apoptosis, DNA damage repair and inhibition of metastasis. The action of many tumor suppressors can be attributed to only one of these mechanisms, but others may promote more than one mechanism (as reviewed in [Sun and Yang, 2010]). Tumor suppressive functions has not solely been assigned to proteins, and also long ncRNAs can represent tumor suppressors [Zhou et al., 2012].

Several mechanisms for tumor suppression have been described with a steadily growing number of mechanisms and genes involved in tumor suppression. Thereby, suppression of cell division is the working mechanism of most tumor suppressors and exemplified by the action of the retinoblastoma protein (pRB), APC (adenomatosis polyposis coli), p15, p16, p21 or p53 [Sun and Yang, 2010]. pRB inhibits cell cycle progression by interaction with different transcription factors such as E2Fs, which themselves regulate proliferation genes [Sun and Yang, 2010],[Helin et al., 1993]. Another mechanism to inhibit cell division is that of APC, a protein that stabilizes microtubules [Green et al., 2005]. p15, p16, p21 and others negatively regulate the activity of CDKs, which for their part inhibit RB1 activity (as reviewed in [Sherr, 1996]).

Induction of apoptosis is mediated by tumor suppressors such as p53, APC or PTEN (phosphatase and tensin homolog) [Sun and Yang, 2010]. p53 for example is able to induce apoptosis via the extrinsic and intrinsic pathway, which activates a caspase cascade and promotes formation of the apoptosome, respectively [Haupt et al., 2003].

Tumor suppressors involved in DNA damage repair include MSH2 (mutS homolog 2), MLH1 (mutL homolog 1), ATM (ataxia-telangiectasia mutated gene product), BRCA1 (breast cancer gene 1) and p53 [Sun and Yang, 2010]. For example, MSH2 and MLH1 repair DNA mismatches and p53 can induce nucleotide excision repair [Seifert and Reichrath, 2006],[Adimoolam and Ford, 2003].
Finally, inhibition of metastasis is a tumor suppressive mechanism that is promoted by the action of proteins such as metastin, BRMS1 (breast cancer metastasis suppressor 1) or TIMP (tissue inhibitor of metalloproteinase) [Sun and Yang, 2010].
For this thesis, the tumor suppressive activities of RB1 and calreticulin are of particular interest.

1.8.1 Functionality of the retinoblastoma protein

RB1 was the first tumor suppressor gene to be described and it was discovered due to its determination of susceptibility to hereditary retinoblastoma, a childhood tumor of the eye [Lee et al., 1987],[Friend et al., 1986]. pRB, the RB1 gene product, belongs to the pocket domain family also including p107 and p130 proteins, that are structurally and functionally related to pRB [Burkhart and Sage, 2008].
Since its discovery, the RB1 gene has been found mutated or otherwise inactivated in a wide range of human cancers beside retinoblastoma [Viatour and Sage, 2011]. Initially, the tumor suppressive mechanism of RB1 has been attributed to its ability to restrict cell cycle progression, however, pRB also functions in cellular senescence, differentiation, apoptosis and maintenance of genomic stability [Burkhart and Sage, 2008].
The capacity of pRB to arrest cells in G1 phase of the cell cycle, thereby regulating cell cycle progression at the G1/S transition, is mediated by the unphosphorylated and active protein. Unphosphorylated pRB binds to E2F transcription factors located at the promoters of E2F regulated proliferation genes and inhibits their activity, thereby silencing expression of these genes. pRB binds preferentially to E2F1, E2F2, and E2F3 whereas p107 and p130 are most often found in complex with E2F4 and E2F5 (as reviewed in [Chinnam and Goodrich, 2011]).
Upon mitogenic signaling, pRB is phosphorylated by Cyclin/Cdk complexes, resulting in its loss of interaction with E2F factors and cell cycle genes so that cyclins are increasingly expressed and cells divide [Dynlacht et al., 1994]. Thereby, pRB can be phosphorylated by different kinases, such as cyclin-D-CDK4, cyclin-D-CDK6, cyclin-A-CDK2, cyclin-E-CDK2, CHK2 (checkpoint homologue 2) and RAF1. Cyclin/Cdk complexes are themselves regulated by cell cycle inhibitors of the INK4 and CIP/KIP families, such as p16INK4a and p12CIP1 (as reviewed in [Burkhart and Sage, 2008],[Viatour and Sage, 2011]).
Furthermore, interaction of pRB with APC/C (anaphase-promoting complex/cyclosome) that results in increased stability of the p27 cell cycle inhibitor is part of pRB's mecha-

nisms to achieve G1 cell cycle arrest [Binné et al., 2007].
The additional functions of pRB are achieved by its working as a transcriptional co-factor of various transcription factors to assemble specific protein complexes on DNA as well as being an adaptor protein to chromatin remodeling enzymes. Overall, pRB regulates the expression of specific target genes and alters chromatin structure (as reviewed in [Macaluso et al., 2006]). For example, pRB recruits histone deacetylases to chromatin, a process that results in the alteration to a repressive chromatin structure [Takaki et al., 2004].
Inactivation of pRB is mostly occurring in later stages of tumor development as loss of pRB is not advantageous during tumor initiation. This is due to its prosurvival function that are somewhat contradictory to its tumor suppressive activity [Burkhart and Sage, 2008]. When pRB function is lost, E2F1-specific target genes that are involved in apoptosis become activated, resulting in the upregulation of p53 and activation of p53-dependent apoptosis. This causes an increase in cellular death upon RB1 inactivation [Tsai et al., 1998],[Macleod et al., 1996]. As a result, pRB1 is mostly inactivated following mutation in other components of the cell death machinery (as reviewed in [Viatour and Sage, 2011]).

1.8.2 Functionality of calreticulin

Calreticulin is a multi-functional protein that mostly resides in the endoplasmic reticulum (ER) lumen, however, recently its various roles outside the ER have gained attention [Michalak et al., 2009]. Its functions range from Ca^{2+}-binding, -storage and -signaling, protein chaperoning and regulation of steroid-sensitive gene expression and also include various others (as reviewed in [Michalak et al., 2009],[Wang et al., 2012]).
In the ER, the cellular organelle crucial for maintaining calcium homeostasis, synthesis of lipids and proteins, protein folding and post-translational modifications, calreticulin plays the role of a Ca^{2+}-binding chaperone. Ca^{2+}-binding is achieved via a high-capacity, low-affinity as well as a high-affinity, low-capacity binding domain of the protein and assisted by the function of calnexin, a homologue of calreticulin. Cellular proteins synthesized into the ER, interact with several molecular chaperones including calreticulin that ensure the transport of properly folded proteins outside of the organelle and to different cellular locations (as reviewed in [Michalak et al., 2009],[Wang et al., 2012]).
Outside the ER, calreticulin is present in the cytoplasm, nucleus, extracellular matrix and at the cell membrane, participating in a variety of biological processes such as wound-healing, cell adhesion, phagocytosis and recognition of malignant cells (as reviewed in [Gold et al., 2010]).
Calreticulin is expressed at varying levels during differentiation or malignant transformation of cells, demonstrating the importance of transcriptional and post-transcriptional regulation of this housekeeping gene [Waser et al., 1997]. Calreticulin has been found

overexpressed in several tumors such as breast cancer, bladder cancer or leukemia and this altered expression correlated with higher invasiveness and poorer prognosis [Lwin et al., 2010],[Kageyama et al., 2004],[Helbling et al., 2005]. Simultaneously, exogenous increase in calreticulin levels reduce tumor growth by calreticulin's function as an inhibitor of angiogenesis [Pike et al., 1998],[Pike et al., 1999]. Somatic mutations in the calreticulin (CALR) gene have been found in >70% of patients with myeloproliferative neoplasms with nonmutated JAK2 kinase, suggesting a link between a mutant calreticulin protein and cancer [Nangalia et al., 2013],[Klampfl et al., 2013].

Cell surface exposed calreticulin has immunomodulatory activity by serving as an 'eat-me' signal to phagocytic cells, more specifically to macrophages and dendritic cells [Chao et al., 2010]. This activity is exploited by certain chemotherapeutic agents that induce the translocation of calreticulin to the cell surface and thereby involve of the immune system during cancer cell death [Obeid et al., 2007]. This so-called immunogenic cell death (ICD) is induced by anthracyclines or oxaliplatin and is one of the goals to be achieved during chemotherapy [Panaretakis et al., 2009]. Translocation of calreticulin to the cell surface occurs before other signs of apoptosis are observable, thus in a pre-apoptotic fashion [Obeid et al., 2007]. The process of ICD involves removal of cells by macrophages and can modulate cancer cell survival. Therefore it has been suggested that this programmed cell removal must be overcome by successful cancer clones in addition to their avoidance of programmed cell death (apoptosis) [Chao et al., 2012]. Indeed, increased levels of cell surface calreticulin by several cancer types, probably due to endoplasmic stress, is counterbalanced by increased exposure of CD47 on these cells, a 'don't-eat me signal' [Chao et al., 2010].

1.9 Aims of the thesis

Various long ncRNA species are recently detected in high-throughput transcriptome studies, revealing that the human genome is pervasively transcribed. A reoccurring theme is the prevalent transcription of these RNA at regulatory elements such as promoters and enhancers and their intersection with loci of PCGs. Although increasing numbers of such long ncRNAs are reported, their functionality remains mostly elusive.

This thesis aims to characterize the transcriptional regulation of long ncRNA genes expressed from bidirectional promoters shared with PCGs. Thereby, one objective is to determine the potential of co-regulation and co-expression of both gene types. Furthermore, gene regulatory effects of bidirectionally expressed long ncRNA/PCG pairs are intended to be investigated. This includes the determination of individual target genes for each of the gene partners, as well as their involvement in the regulatory circuits of the paired gene. This intends to answer the question if gene pairing entails co-functionality. A regulatory interplay is imagined to either represent a direct regulatory effect of the paired gene or a complete or partial functional overlap of both gene partners in the regulation of a biological pathway. Finally, the question should be answered if a regulatory interplay has biological significance.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Instrumentation

All devices used in this work are listed in Table 2.1.

Table 2.1: List of devices

Device	Name	Manufacturer, Country
Analytical balance	BP 61	Sartorius, Germany
Centrifuge	5430, 5810 R, MiniSpin	Eppendorf, Germany
Centrifuge	FRESCO 17	Heraeus, Germany
Chemolumineszenz imager	FUSION-SL Advance 4.2 MP	PeqLab, Germany
Electrophoresis system (DNA gels)	Mini-Sub Cell GT Cell	Bio-Rad Laboratories, USA
Electrophoresis system (protein gels)	XCell SureLock Mini-Cell	Life Technologies, USA
Flow cytometer	Cyan ADP	Beckman Coulter, USA
Flow cytometry analyzer	FACScan	Becton Dickinson, USA
Fluid aspiration system	BioChem-VacuuCenter BVC 21	Vacuubrand, Germany
Fluorescence spectrofluorometer	LUMIstar OPTIMA	BMG Labtech, Germany
Freezer	Comfort, Premium NoFrost	Liebherr, Switzerland
Heating Block	Thermomixer5436	Eppendorf, Germany
Ice machine	AF30	Scotsman Ice Systems, USA
Incubator for bacteria	Heraeus-Brutschrank B 504	Heraeus, Germany
Incubator for cell culture	Heracell CO2	Heraeus, Germany
Magnetic stirrer	TK22	Kartell Labware, Australia
Microplate luminometer	LUMIstar Omega	BMG Labtech, Germany
Microscope	Axiovert 40 CFL	Zeiss, Germany
Microscope	IXM XLS	Molecular Devices, USA
Microwave	SEVERIN 900&Grill	Severin, Germany
Multi-pipette	Multipette Xstream	Eppendorf, Germany
pH meter	HI 221	Hanna Instruments, Canada
Photometer	Ultrospec 10 Cell Density Meter	Amersham Biosciences, UK
Pipettes	PIPETMAN P2, P20, P200, P1000	Gilson, USA
Pipettor	VacuuHandControl	Vacuubrand, Germany
Power supply	Power Pac 300	Bio-Rad Laboratories, USA
qRT-PCR cycler	ABI (PRISM 7900 HT)	Life Technologies, USA
Refrigerator	ProfiLine	Liebherr, Switzerland
Rocking platform	ST5	Ingenieurbüro CAT, Germany
Shaker for bacteria culture	Innova 440	New Brunswick Scientific, Germany

Continued on next page

Table 2.1: *Continued from previous page*

Device	Name	Manufacturer, Country
Sonicator	W375	Heat Systems, USA
Spectrophotometer	NanoDrop 2000	PeqLab, Germany
Sterile bench	HERAsafe HSF 12	Heraeus, Germany
Thermocycler	Peqstar 2x gradient	PeqLab, Germany
UV transilluminator	Gel iX20 Imager	Intas, Deutschland
Vortexer	Vortex Genie 2	Scientific Industries, USA
Water bath	WNE	Memmert, Germany
Water purification system	Purelab Chorus	Elga Labwater, Germany
Wet Blotting System	Mini Trans-Blot Cell	Bio-Rad Laboratories, USA

2.1.2 Consumables

Table 2.2 lists the items routinely used in this thesis.

Table 2.2: List of consumables

Product	Manufacturer, Country
4-12% NuPAGE Bis-Tris Precast Gels	Life Technologies, USA
96-well Black/Clear Imaging Plates	BD Biosciences, USA
96-well white plates (LumiNunc)	Thermo Fisher Scientific, USA
Bottle top filter (Steritop-GV, 0.22 µm)	Merck Millipore, Germany
Cell culture plates (10 cm, 6-, 24-, 96-well)	TPP, Switzerland
Cell scraper	Sarstedt, Germany
Combitips advanced (0.1 ml, 0.5 ml)	Eppendorf, Germany
Eppendorf safe-lock micro test tubes (1.5 ml, 2 ml)	Eppendorf, Germany
Falcon Tubes (15 ml and 50 ml)	Greiner-Bio-One, Germany
Gloves	VWR International, Germany
MicroAmp Clear Adhesive Film	Life Technologies, USA
Microscope slides	Thermo Fisher Scientific, USA
Needles	BD Biosciences, USA
Optical well plates for qPCR	Life Technologies, USA
Pasteur-Pipetten	VWR International, Germany
Petri dishes	Greiner-Bio-One, Germany
Pipette tips	DeckWorks, Corning, USA
Precision Plus Protein Dual Color Marker	Bio-Rad Laboratories, USA
PVDF membrane	Bio-Rad Laboratories, USA
Serological pipetts	Sarstedt, Germany
Surgical blades	B. Braun Melsungen AG, Germany
Syringes	BD Biosciences, USA
Weighting dishes	Roth, Germany
Whatman Gel-Blotting Paper, 1.4 mm	Thermo Fisher Scientific, USA

2.1.3 Chemicals

All chemicals used in this study are listed in Table 2.3.

Table 2.3: List of chemicals

Chemical	Manufacturer, Country
1,4-Dithiothreitol (DTT)	Biomol, Germany
2-Propanol	Merck, Germany
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 1 M	AppliChem, Germany
4,6-diamidino-2-phenylindole (DAPI)	Sigma-Aldrich, USA
Acetic acid	Sigma-Aldrich, USA
Agar, Bacto	BD Biosciences, USA
Ampicillin, sodium salt	AppliChem, Germany
Bis(2-hydroxyethyl)amino-tris(hydroxy-methyl)methan	Sigma-Aldrich, USA
CellTracker Orange CMRA Dye	Life Technologies, USA
CD11b MicroBeads, human and mouse	Bergisch Gladbach, Germany
Complete, EDTA free, protease inhibitor cocktail tablets	Roche, Switzerland
Crystal violet	Alfa Aesar, USA
Deoxyadenosine triphosphate (dATP)	Life Technologies, USA
Deoxycytidine triphosphate (dCTP)	Life Technologies, USA
Deoxyguanosine triphosphate (dGTP)	Life Technologies, USA
Deoxythymidine triphosphate (dTTP)	Life Technologies, USA
Diethyldicarbonat (DEPC)	Sigma-Aldrich, USA
Ethylenediaminetetraacetic acid (EDTA), 500 mM	AppliChem, Germany
Fast SYBR Green Master Mix	Life Technologies, USA
Fetal bovine serum (FBS)	EuroClone, Italy
Fluorescent mounting medium	Dako, Germany
Formaldehyde (37%)	AppliChem, Germany
GeneRuler 100bp Plus DNA Ladder	Thermo Fisher Scientific, USA
GeneRuler 1kb DNA Ladder	Thermo Fisher Scientific, USA
Glycerol	Merck Millipore, Germany
Glycerol, BioUltra	Sigma-Aldrich, USA
Glycine	Merck, Germany
GlycoBlue Coprecipitant	Life Technologies, USA
Goat serum	Life Technologies, USA
HiPerFect Transfection Reagent	Qiagen, Germany
IGEPAL CA-630	Sigma-Aldrich, USA
LE Agarose	Biozym, Germany
Lipofectamine 2000	Life Technologies, USA
Magnesium chloride ($MgCl_2$)	Sigma-Aldrich, USA
Methanol	Merck, Germany
Milk powder	Biomol, Germany
NuPAGE MOPS SDS Running Buffer	Life Technologies, USA
Orange G	Sigma-Aldrich, USA
Paraformaldehyde	Sigma-Aldrich, USA
Penicillin-Streptomycin	Life Technologies, USA
Phosphate-buffered saline (PBS), 10x	Life Technologies, USA
Precision Plus Protein Dual Color Marker	Bio-Rad Laboratories, USA
RNASEZAP	Sigma-Aldrich, USA
RotiLoad	Roth, Germany
Sodium chloride (NaCl)	AppliChem, Germany
Sodium deoxycholate	Sigma-Aldrich, USA
Sodium dodecyl sulfate (SDS)	Promega, USA
Sodium hydroxide (NaOH)	Sigma-Aldrich, USA
Stripping buffer (for western blots)	Thermo Fisher Scientific, USA

Continued on next page

Table 2.3: *Continued from previous page*

Chemical	Manufacturer, Country
Sucrose	Sigma-Aldrich, USA
SUPERase In RNase Inhibitor	Life Technologies, USA
SuperSignal West DURA Extended Duration	Thermo Fisher Scientific, USA
SYBR Safe DNA Gel Stain	Life Technologies, USA
Tris/Borate/EDTA (TBE) buffer (10x)	AppliChem, Germany
Tris-HCl, 1M soln., pH 7.4, RNase free	Alfa Aesar, USA
Triton X-100	Sigma-Aldrich, USA
Trizma base	Sigma-Aldrich, USA
TRIzol Reagent	Life Technologies, USA
Tryptone, Bacto	BD Biosciences, USA
Tween-20	VWR International, Germany
Yeast extract, Bacto	BD Biosciences, USA

2.1.4 Buffers, solutions and media

All buffers and media were prepared with Milli-Q water or DEPC-treated Milli-Q water (in case of their use for RNA isolation) and sterile filtered (0.22 µm) or autoclaved. The buffers, solutions and media used in this work are listed in Table 2.3, 2.4 and 2.5.

Table 2.4: List of solutions

Solution	Composition
Ampicillin (1000x)	100 mg ampicillin in 1 ml water
Blocking solution	5% (w/v) milk powder, 1x PBS, 0.1% (v/v) Tween-20
DNA loading buffer (6x)	30% (v/v) glycerol, 0.25% (w/v) orange G
Glycerol buffer	20 mM Tris (pH 7.5, RNase-free), 75 mM NaCl, 0.5 mM EDTA, 50% (v/v) glycerol
LB agar	10 g tryptone, 5 g yeast extract, 10 g NaCl, 15 g agar in 1 l water (pH 7.0, adjusted with NaOH)
LB medium	10 g tryptone, 5 g yeast extract, 10 g NaCl in 1 l water (pH 7.0, adjusted with NaOH)
Lysis buffer	10 mM Tris-HCl (pH 7.5, RNase-free), 150 mM NaCl, 0.15% (v/v) IGEPAL CA-630
Nuclear lysis buffer	10 mM HEPES (pH 7.6), 7.5 mM MgCl2, 0.2 mM EDTA, 0.3 M NaCl
Ripa buffer	25 mM Tris (pH 7.6), 150 mM NaCl, 1% (v/v) IGEPAL CA-630, 1% sodium deoxycholate, 0.1% SDS
Sucrose buffer	10 mM Tris-HCl (pH 7.5, RNase-free), 150 mM NaCl, 24% (w/v) sucrose
PBST	1x PBS, 0.1% (v/v) Tween-20
Transfer buffer	3.03 g Trizma base, 14.4 g glycine, 140 ml methanol in 1 l water

Table 2.5: List of media

Media	Manufacturer, Country
DMEM, High Glucose, Pyruvate	Life Technologies, USA
RPMI 1640 Medium	Life Technologies, USA
Opti-MEM Reduced Serum Medium	Life Technologies, USA
PBS, 1x	Life Technologies, USA
Trypsin-EDTA (0.25%), phenol red	Life Technologies, USA

2.1.5 Molecular biology kits

All molecular biology kits used in this thesis are listed in Table 2.6.

Table 2.6: List of molecular biology kits

Kit	Manufacturer, Country
Bicinchoninic Acid Kit for Protein Determination	Sigma-Aldrich, USA
Dual-Glo Luciferase Assay System	Promega, USA
ExactStart Eukaryotic mRNA 5'- & 3'-RACE Kit	Epicentre, USA
Fast SYBR Green Master Mix	Life Technologies, USA
High-Capacity RNA-to-cDNA Kit	Life Technologies, USA
Mix & Go E. coli Transformation Kit & Buffer Set	Zymo Research, USA
TOPO TA Cloning Kit for Subcloning	Life Technologies, USA
QIAamp DNA Mini Kit	Qiagen, Germany
QIAGEN Plasmid Mini Kit	Qiagen, Germany
QIAprep Spin Miniprep Kit	Qiagen, Germany
QIAquick Gel Extraction Kit	Qiagen, Germany
QIAquick PCR Purification Kit	Qiagen, Germany
SuperSignal West Dura Chemiluminescent Substrate	Thermo Fisher Scientific, USA

2.1.6 Enzymes and proteins

All enzymes used in this work are listed in Table 2.7.

Table 2.7: List of enzymes and proteins

Enzyme	Manufacturer, Country
Antarctic Phosphatase	New England Biolabs, USA
BamHI	New England Biolabs, USA
BbsI	New England Biolabs, USA
BglII	New England Biolabs, USA
DNase I	New England Biolabs, USA
MluI	New England Biolabs, USA
Phusion High-Fidelity DNA Polymerase	Thermo Fisher Scientific, USA
T4 DNA Ligase	New England Biolabs, USA
Taq DNA Polymerase	Life Technologies, USA

2.1.7 Plasmids

Table 2.8 summarizes all plasmids and constructs used in this work.

Table 2.8: List of plasmids

Plasmid	Description	Manufacturer, Country
pCR2.1-TOPO	for sequencing of PCR amplified or subcloning transcript or genomic regions	Life Technologies, USA
pGL3-Basic Vector	for cloning of *Renilla reniformis* luciferase	Promega, USA
pRL-TK Vector	or amplification of *Renilla reniformis* luciferase	Promega, USA

Continued on next page

Table 2.8: *Continued from previous page*

Plasmid	Description	Manufacturer, Country
pGL3-Renilla Luc vector	for cloning of bidirectional promoter sequences	This study
pcDNA3	for overexpression of ncRNA-RB1	Invitrogen, USA
pX330-U6-Chimeric_BB-CBh-hSpCas9	for cloning of guide RNAs specific to ncRNA-RB1 genomic locus	Addgene, USA

2.1.8 Antibodies

Table 2.9 summarizes all antibodies used in this work.

Table 2.9: List of antibodies

Antibody	Manufacturer, Country
α-tubulin (DM1A), mouse mAb	Cell Signaling, USA
β-actin (8H10D10), mouse mAb	Cell Signaling, USA
Alexa Fluor 488 Goat Anti-Mouse	Life Technologies, USA
CALR (FMC 75), mouse mAb	Abcam, UK
HRP-linked, anti-mouse IgG	Cell Signaling, USA
Rb1 (4H1), mouse mAb	Cell Signaling, USA
CD11b-FITC, anti-human and mouse	Miltenyi Biotec, Germany

2.1.9 Oligonucleotides

All primers, dsiRNAs, guide RNAs and ASOs used in this thesis are listed in Tables 2.10, 2.11, 2.12, 2.13, 2.14, 2.15 and 2.16. All were obtained from Integrated DNA Technologies, Belgium.

Cloning primers

Table 2.10: List of cloning primers

Name	Sequence	Application
BRCA1_FW	5'-GGAGATCTAAGCCGCAACTGGAAGAGTA-3'	cloning of BRCA1 promoter via BglII and HindIII
BRCA1_RV	5'-GGAAGCTTACCCAGAGCAGAGGGTGAAG-3'	cloning of BRCA1 promoter via BglII and HindIII
CALR_FW	5'-AGCATCTTATCGTCCCTACCA-3'	cloning of CALR promoter into pCR2.1
CALR_RV	5'-CACGGATAGCAGCATGGC-3'	cloning of CALR promoter into pCR2.1
CCNG1_FW	5'-GGAGATCTCAGCCGATTGACCTGACC-3'	cloning of CCNG1 promoter via BglII and HindIII
CCNG1_RV	5'-GGAAGCTTGAGACAACTCGGCCCTGAT-3'	cloning of CCNG1 promoter via BglII and HindIII
FKTN_FW	5'-GGAGATCTGGTGAGGATGCGACAAGAGT-3'	cloning of FKTN promoter via BglII
FKTN_RV	5'-GGAAGCTTGAGCCTCCCGTACCTTACCT-3'	cloning of FKTN promoter via BglII
Magoh_FW	5'-GGAGATCTTGCAGTCTTGTTGCCACTTC-3'	cloning of Magoh promoter via BglII and HindIII
Magoh_RV	5'-GGAAGCTTGCCTGAACTTCCAAGAGCAA-3'	cloning of Magoh promoter via BglII and HindIII

Continued on next page

Table 2.10: *Continued from previous page*

Name	Sequence	Application
ncRNA-RB1_FW	5'-GGAAGCTTTCACGTCCGCGAGGCTCC-3'	cloning and overexpression of ncRNA-RB1 via XbaI and HindIII
ncRNA-RB1_RV	5'-GGTCTAGACATCAGACAAAGGTTGGGATT-3'	cloning and overexpression of ncRNA-RB1 via XbaI and HindIII
RB1_FW	5'-GGAAGCTTGCAACTGAGCGCCGCGTC-3'	cloning of RB1 promoter via HindIII and BglII
RB1_RV	5'-GGAGATCTAGCGCCCCAGTTCCCCAC-3'	cloning of RB1 promoter via HindIII and BglII
RL_FW	5'-GGACGCGTCACTATAGGCTAGCCACCATGA-3'	cloning of *Renilla* Luc gene from pRL-TK Vector via MluI and BamHI
RL_RV	5'-GGGGTACCTGGATCCTTATCGATTTTACCA-3'	cloning of *Renilla* Luc gene from pRL-TK Vector via MluI and BamHI

Mutation primers for RB1 promoter

Table 2.11: List of mutation primers for RB1 promoter

Name	Sequence	Application
RB1prom_2Mut_FW	5'-GGGAGCCTCGCGGACGAGCCGCCGCGGGCGGAAGT-3'	introduction of 2 bp mutation into the ncRNA INR sequence
RB1prom_2Mut_RV	5'-ACTTCCGCCCGCGGCGGCTCGTCCGCGAGGCTCCC-3'	introduction of 2 bp mutation into the ncRNA INR sequence
RB1prom_3Del_FW	5'-CCCGGGAGCCTCGCGGACGCGCCGCGGGCGGAAGT-3'	introduction of 3 bp deletion into the ncRNA INR sequence
RB1prom_3Del_RV	5'-ACTTCCGCCCGCGGCGCGTCCGCGAGGCTCCCGGG-3'	introduction of 3 bp deletion into the ncRNA INR sequence
RB1prom_9Del_FW	5'-GGAGCCTCGCGCCGCGGGCGGAAGTGA-3'	deletion of the ncRNA INR sequence
RB1prom_9Del_RV	5'-GTCACTTCCGCCCGCGGCGCGAGGCTCC-3'	deletion of the ncRNA INR sequence
RB1prom_10Del_FW	5'-CCGCGGTTGGCAGTTGCCGGGCGGGGGA-3'	deletion of the RB1 BREu sequence
RB1prom_10Del_RV	5'-TCCCCCGCCCGGCAACTGCCAACCGCGG-3'	deletion of the RB1 BREu sequence
RB1prom_TATA_FW	5'-AAGTGACGTTTTATAGCGGTTGGA-3'	introduction of a TATA box into the ncRNA INR sequence
RB1prom_TATA_RV	5'-TCCAACCGCTATAAAACGTCACTT-3'	introduction of a TATA box into the ncRNA INR sequence
RB1prom_RL_FW	5'-ACCACTGCGGACCAGTTATC-3'	anneals in the *Renilla* Luc gene
RB1prom_FL_RV	5'-GCCTTATGCAGTTGCTCTCC-3'	anneals in the Firefly Luc gene

qRT-PCR primers

Table 2.12: List of qRT-PCR primers

Name	Sequence
β-actin_FW	5'-CGACAGGATGCAGAAGGAG-3'
β-actin_RV	5'-GTACTTGCGCTCAGGAGGAG-3'
7SL_FW	5'-GTCAAAACTCCCGTGCTGAT-3'
7SL_RV	5'-GCTGGAGTGCAGTGGCTATT-3'
BRCA1-mRNA_FW	5'-CAATGGAGATAATGGCAGCA-3'
BRCA1-mRNA_RV	5'-TCCAAATTCCACGTGACTACC-3'
CALR_FW	5'-GACATGCACGGAGACTCAGA-3'

Continued on next page

Table 2.12: *Continued from previous page*

Name	Sequence
CALR_RV	5'-AGCACGTTCTTGCCCTTGTA-3'
CBX_FW	5'-AGCGCAAAGCTGATTCTGAT-3'
CBX_RV	5'-AGCCTCGTGGCTTTTCTGA-3'
CCNG1-mRNA_FW	5'-TCACCTTCCAACAATTCCTGA-3'
CCNG1-mRNA_RV	5'-AAGGTTGTGGAGAAAGGCTTC-3'
COPRS_FW	5'-GGACTCGGAGTTGAAAGCAG-3'
COPRS_RV	5'-AAATGCTCTCCTGGATGTCG-3'
CXCL5_FW	5'-GATCCAGAAGCCCCTTTTCT-3'
CXCL5_RV	5'-GAAACTTTTCCATGCGTGCT-3'
EPCAM_FW	5'-CGTCAATGCCAGTGTACTTCA-3'
EPCAM_RV	5'-TCCCAAGTTTTGAGCCATTC-3'
EZR_FW	5'-GGCTAAGGAGGAGCTGGAGA-3'
EZR_RV	5'-TGGCAGTGTATTCTGCAAGC-3'
IL6_FW	5'-CCTTCCAAAGATGGCTGAAA-3'
IL6_RV	5'-CCTCAAACTCCAAAAGACCA-3'
ncRNA-BRCA1_FW	5'-CACCGCACCTGGTCGATTAA-3'
ncRNA-BRCA1_RV	5'-GGGAGCCTTGATGTGTGCTT-3'
ncRNA-CCNG1_FW	5'-AGTGGTTCTGCCCCATCTTT-3'
ncRNA-CCNG1_RV	5'-GTGCTTTGAGAGGCCAAAGT-3'
ncRNA-PRKCQ_FW	5'-GGTGGGACTGCTTTCAACTT-3'
ncRNA-PRKCQ_RV	5'-GCTGTTATCCGTTTGCCATT-3'
ncRNA-RB1_FW1	5'-GGACGTGCTTCTACCCAGAAC-3'
ncRNA-RB1_FW2	5'-ACAAACTTGGAGCGCTGATA-3'
ncRNA-RB1_RV	5'-TCCTTCTCAGTTGACGAGTTCA-3'
preGAPDH_FW	5'-CAATGACCCCTTCATTGACC-3'
preGAPDH_FW	5'-GGCTCACCATGTAGCACTCA-3'
PRKCQ-mRNA_FW	5'-TGAGAGGTGCAGGAAGAACA-3'
PRKCQ-mRNA_RV	5'-GCCTTCCGTCTCAAATTCAT-3'
Promt-40-9_FW	5'-GGCATCTGGACTAGAATGAA-3'
Promt-40-9_RV	5'-TTGACACCGCCTAATCTTAT-3'
Promt-40-33_FW	5'-CTGGCCTAGCTAAAGTCTCA-3'
Promt-40-33_RV	5'-TCTGCTCCTAGCTCTCAGTC-3'
Promt-40-54_FW	5'-AAGGCCCCTACTTAACTCTC-3'
Promt-40-54_RV	5'-GAGTTTTGGATGGAAAATGA-3'
RB1-mRNA_FW	5'-GAGCAAGGTCTAAGGCAGGA-3'
RB1-mRNA_FW	5'-CTGGCAGTTTGTTGCTTCAG-3'
RRP40_FW	5'-TCATTGGACAGGATGGTCTG-3'
RRP40_RV	5'-CCCACTTCCTGTATGATTTCAC-3'

Sequencing primers

Table 2.13: List of sequencing primers

Name	Sequence
M13_FW	5'-TGTAAAACGACGGCCAGT-3'
M13_RV	5'-CAGGAAACAGCTATGACC-3'

Guide RNAs

Table 2.14: List of guide RNAs

Name	Sequence
ncRNA-RB1_guide1	5'-CCCTATCAGACCCCGGGATA-3'
ncRNA-RB1_guide2	5'-TGGCTTGCCTCACGTTACAA-3'

Antisense oligonucleotides

Table 2.15: List of antisense oligonucleotides

Name	Sequence
ASO_ctrl	5'-G*C*G*C*C*T*G*G*C*A*A*T*T*A*A*A*A*A-3'
ASO1_ncRNA-RB1	5'-G*G*A*C*C*C*A*C*G*C*C*A*G*G*T*T*T*C-3'
ASO2_ncRNA-RB1	5'-C*C*T*C*A*T*G*A*C*T*T*A*G*C*G*T*C*C-3'
ASO3_ncRNA-RB1	5'-G*T*T*C*T*G*G*G*T*A*G*A*A*G*C*A*C*G-3'

*: phosphorothioate-modification

Dicer substrate RNAs

Table 2.16: List of Dicer substrate RNAs

Name	Sequence
dsi1_EXOSC3_s	5'-CGUUGAGCCUGAAUGCUAGAGCGTG-3'
dsi1_ EXOSC3_as	5'-CACGCUCUAGCAUUCAGGCUCAACGGU-3'
dsi2_ EXOSC3 _s	5'-AGAAACAGACCAAAUGUGCAGGCTA-3'
dsi2_ EXOSC3_as	5'-UAGCCUGCACAUUUGGUCUGUUUCUUU-3'
dsi1_ncRNA-RB1_s	5'-GACGCUAAGUCAUGAGGAAUUAAAC-3'
dsi1_ncRNA-RB1_as	5'-GUUUAAUUCCUCAUGACUUAGCGUCCC-3'
dsi2_ncRNA-RB1_s	5'-CUGAACUCGUCAACUGAGAAGGAAA-3'
dsi2_ncRNA-RB1_as	5'-UUUCCUUCUCAGUUGACGAGUUCAGAU-3'
dsi1_RB1_s	5'-CUCCCAUGUUGCUCAAAGAACCATA-3'
dsi1_RB1_as	5'-UAUGGUUCUUUGAGCAACAUGGGAGGT-3'
dsi2_RB1_s	5'-UCCUGCUCUGGGUCCUCCUCAGGAG-3'
dsi2_RB1_as	5'-CUCCUGAGGAGGACCCAGAGCAGGACA-3'

2.1.10 Cell lines

All cell lines and the bacterial strain used in this work are listed in Tables 2.17 and 2.18.

Table 2.17: List of cell lines

Cell Line	Description
A549	Human lung carcinoma cell line; initiated in 1972 through explant culture of lung carcinomatous tissue of a 58-year-old male [Lieber et al., 1976]

Continued on next page

Table 2.17: *Continued from previous page*

Cell Line	Description
HEK293	Human embryonic kidney cell line; initiated in 1973 by transformation of cells with sheared adenovirus 5 DNA [Graham et al., 1977]. Cloning and sequencing of the adenovirus 5 insert determined that a colinear segment from nucleotides 1 to 4,344 bp is integrated into chromosome 19 (19q13.2) [Louis et al., 1997]
HeLa	Human cervical carcinoma cell line; initiated in 1951 by explant from a 31-year-old female [SCHERER et al., 1953]
HepG2	Human hepatoblastoma-derived cell line; derived from a 15-year-old adolescent male and initiated in 1979 [López-Terrada et al., 2009],[Aden et al., 1979]
IMR90	Human fibroblast cell line, derived from the lungs of a 16-week female fetus in 1975 [Nichols et al., 1977]
K562	Human chronic myelogenous leukemia cell line, established in 1970 from the pleural effusion of a 53-year-old female [Lozzio and Lozzio, 1979]
MCF7	Human breast adenocarcinoma cell line, established from the pleural effusion of a 69-year-old female in 1976 [Lippman et al., 1976]
THP-1	Human acute monocytic leukemia cell line, derived from a 1-year-old male and established in 1980 [Tsuchiya et al., 1980]
U2OS	Human osteosarcoma cell line derived in 1964 from a moderately differentiated sarcoma of the tibia of a 15-year-old female [Pontén and Saksela, 1967]

2.1.11 Bacterial strain

Table 2.18: List of bacterial strain

Bacterial Strain	Description	Manufacturer, Country
E. coli Zymo DH5-alpha	F-ϕ80lacZΔM15 Δ(lacZYA-argF)U169 deoR nupG recA1 endA1 hsdR17(rK- mK+) phoA glnV44 (supE44) thi-1 gyrA96 relA1, λ	Zymo Research, USA

2.1.12 Software

The software used in this thesis is listed in Table 2.19.

Table 2.19: List of software

Software	Reference
CellQuest Pro	Becton Dickinson, USA
SDS Software 2.2	Applied Biosystems, USA
Primer3	Untergasser et al., 2012 [Untergasser et al., 2012]
UCSC	Kent et al., 2002 [Kent et al., 2002],[Kent, 2002]
DAVID	Huang et al., 2009 [Huang et al., 2009a],[Huang et al., 2009b]
MetaXpress 5.1	Molecular Devices, USA
FlowJo	Tree Star, USA

2.2 Methods

2.2.1 Molecular biology methods

2.2.1.1 Isolation of genomic DNA

Genomic DNA from human cell lines was isolated using the QIAamp DNA Mini Kit according to the manufacturer's instructions.

2.2.1.2 Polymerase chain reaction

Polymerase chain reaction (PCR) was used to amplify genomic regions. PCR reaction mixes were prepared according to the Phusion High-Fidelity DNA Polymerase protocol and as described below (Table 2.20):

Table 2.20: Composition of PCRs

Compound	Amount
Phusion HF buffer (5x)	4 µl
dNTP mixture (10 mM of dATP, dCTP, dGTP, dTTP)	0.4 µl
DMSO	0.6 µl
Primer* mixture (FW and RV, each 10 µM)	1 µl
DNA	50-200 ng
Phusion DNA polymerase (2U/µl)	0.2 µl
ddH_2O	ad 20 µl

*: Primers used are listed in Tables 2.10, 2.11, 2.12, 2.13, 2.14, 2.15 and 2.16

The amplification was performed using the following PCR program (Table 2.21):

Table 2.21: Conditions for PCRs

Step	Temperature [°C]	Time	Cycles
Initial denaturation	98	30 s	1
Denaturation	98	5-10 s	
Annealing	T_m + 3°C	10-30 s	35
Extension	72	30 s/kb	
Final extension	72	10 min	1

2.2.1.3 Site-directed mutagenesis PCR

To introduce desired mutations or internal deletions of several base pairs into PCR fragments of the RB1 promoter sequence, two consecutive PCR reactions were carried out. Four primers were used of which two are located at the ends of the full length DNA region to be amplified, more specifically in the Firefly and *Renilla* luciferase genes. Two additional primers were designed to anneal at the site of mutation, harboring the mutations to be introduced and were overlapping to each other. These primers were annealing in the RB1 promoter sequence.

First, two PCR amplicons were generated by standard PCR (Section 2.2.1.2) using one

outer primer and one of the mutation primers, respectively. Secondly, a PCR was conducted so that the PCR mix included both previously amplified PCR fragments (1µl of each PCR reaction) that served as primers to each other due to their sequence overlap. In the 6^{th} cycle of this PCR reaction, the two outer primers were added (Table 2.21) to amplify the mutated full-length PCR fragments.

2.2.1.4 Agarose gel electrophoresis

Nucleic acids were separated by agarose gel electrophoresis, allowing to determine specificity and amount of amplified PCR products or digested plasmids (sections 2.2.1.2, 2.2.1.9). Each DNA sample was mixed with 6x DNA loading buffer and the mixture was loaded onto agarose gels (0.8-1.0% agarose in 1x TBE buffer supplemented with SybrSafe). As size markers, commercial DNA ladders were loaded. Gels were run at a constant voltage of 110 V in 1x TBE buffer for 20-40 min. DNA bands were visualized by UV illumination at 254 nm.

2.2.1.5 DNA gel extraction

DNA bands of interest were cut from agarose gels during UV illumination using a surgical blade. DNA was extracted from the gel using the QIAquick Gel Extraction Kit following the manufacturer's instructions.

2.2.1.6 Determination of nucleic acid concentration and purity

The concentration of DNA and RNA in aqueous solution was determined by light absorption at a wavelength of 260 nm using a Nanodrop spectrophotometer. An optical density (OD) of 1 corresponds to 50 µg/ml for dsDNA and to 40 µg/ml for RNA. The purity of the sample is determined by the ratio of absorbances measured at 260 and 280 nm (A260/A280) and reaches a value of ~1.8 for pure DNA 2.0 for pure RNA.

2.2.1.7 A-tailing of PCR products

PCR products that were amplified using the Phusion DNA polymerase (Section 2.2.1.2) needed to be A-tailed in order to clone these into the TOPO TA vector. To this end, the Taq polymerase was used which catalyzes the non-template directed addition of an adenine residue to the 3'-end of both DNA strands. The A-tailing reaction was performed by incubating the following mix (Table 2.22) at 72°C for 20 min and subsequent use of the QIAquick PCR Purification Kit according to the manufacturer's recommendations, thereby removing residual nucleotides.

Table 2.22: Composition of A-tailing reactions

Compound	Amount
PCR buffer (10x)	2 µl
$MgCl_2$ (50 mM)	0.6 µl
dATP (1 mM)	2 µl
Taq DNA polymerase (5 U/µl)	0.5 µl
ddH_2O	ad 20 µl

2.2.1.8 TOPO TA cloning

For subcloning or sequencing, PCR products (Section 2.2.1.2) were cloned into the pCR2.1-TOPO vector using the TOPO TA Cloning Kit for Subcloning and following the recommendations of the manufacturer. Within a few minutes, this kit allows to ligate A-tailed PCR products into the linearized pCR2.1-TOPO vector that harbors single 3'-T overhangs and a covalently bound topoisomerase I.

2.2.1.9 Restriction digest

PCR products and vectors were digested with restriction enzymes to generate compatible ends for subsequent ligation reactions. Depending on the cloning strategy, respective restriction endonucleases from New England Biolabs were used and buffers as well as incubation temperatures were chosen according to the instructions. A typical digestion reaction contained either 1-3 µg of plasmid DNA, the eluate of a gel extraction reaction (Section 2.2.1.5) or the eluate of a PCR purfication reaction (QIAquick PCR Purification Kit, Table 2.6) in case of PCR products in a final volume of 20 µl. After digestion, the DNA was purified from restriction enzymes as well as small DNA fragments using the QIAquick PCR Purification Kit according to the manufacturer's instructions. In case of larger DNA by-products, the digested DNA was size separated on an agarose gel (Section 2.2.1.4) followed by gel extraction (Section 2.2.1.5).

2.2.1.10 Dephosphorylation of vectors

To prevent the re-ligation of cohesive ends of a vector, a dephosphorylation reactions were performed. Following restriction digest and purification (Section 2.2.1.9), the digested vector was incubated with Antarctic phosphatase according to the instructions of the manufacturer. Briefly, a 20 µl reaction containing vector, buffer and enzyme was incubated at 37°C for 20 min, followed by a 10 min incubation at 65°C to heat-inactivate the enzyme.

2.2.1.11 Ligation of DNA

Digested inserts and vectors were ligated due to the presence of cohesive ends. Ligation reactions typically contained a 3-fold molar excess of insert DNA in proportion to the vec-

tor. A 10 µl ligation reaction containing T4 ligase (400 U), buffer, vector and insert was set up according to the manufacturer's instructions and incubated at room-temperature for 1 h or at 16°C overnight. The mixture was then directly used for transformation of *E. coli* cells (Section 2.2.1.13) or stored at -20°C.

2.2.1.12 Generation of chemically competent *E. coli* cells

To generate chemically competent *E. coli* cells, the Mix & Go E. coli Transformation Kit was employed. Briefly, *E. coli* Zymo DH5-a cells were plated on an agar plate without antibiotics and incubated at 37°C for 18 hours. The following day, one colony was picked and incubated in 4 ml LB medium at 37°C and 300 rpm overnight. A day culture was started by transferring 0.8 ml of overnight culture into 50 ml of ZymoBroth medium and incubation at 26°C and 300 rpm until the suspension reached an OD600 of 0.6. The culture was cooled down on ice for 10 min and cells were pelleted by centrifugation at 3,000 rpm for 10 min at 4°C. The supernatant was removed and cells were resuspended in 5 ml ice-cold 1x Wash Buffer. Cells were then re-pelleted by centrifugation at 3,000 rpm for 10 min at 4°C. The supernatant was removed and cells were resuspended in 5 ml ice-cold 1x Competent Buffer. Finally, 50 µl of suspension were aliquoted on ice into 0.5 ml reaction tubes, each, and stored at -80°C until competent cells were transformed (Section 2.2.1.13).

2.2.1.13 Transformation of *E. coli* cells

Chemically competent *E. coli* cells, generated with the Mix & Go E. coli Transformation Kit (Section 2.2.1.12), were thawed on ice, then 5 µl of ligation reaction (Section 2.2.1.11) or 200 ng of plasmid were carefully added to the cells followed by further incubation of the mixture on ice for 10 min. Cells were spread on pre-warmed agar plates containing ampicillin (100 µg/ml) and inoculated at 37°C for 18 hours.

2.2.1.14 Colony PCR

In order to test whether the cloning was successful, colonies of transformed *E. coli* cells were used in colony PCRs. To this end, a colony was resuspended in 10 µl of ddH_2O and 1 µl of this mixture was used in a PCR reaction (Section 2.2.1.2) containing the primer pairs used to amplify the desired insert or primers that annealed in the vector backbone.

2.2.1.15 Bacterial culture

Following successful cloning, plasmids were amplified by culturing positive bacterial clones in 4 ml LB medium containing ampicillin (100 µg/ml) at 37°C for and 300 rpm for 18 hours.

2.2.1.16 Plasmid isolation and analytical digest

To extract and purify plasmid DNA from bacterial cultures on a small scale, the QIAGEN Plasmid Mini Kit was used according to the protocol of the manufacturer, if plasmids were subsequently transfected into human cells (Section 2.2.2.2). The QIAprep Spin Miniprep Kit was used, if plasmids were subsequently used for sequencing or in further cloning reactions.
To test whether plasmids contained desired inserts, restriction digestions followed by agarose gel electrophoresis were performed (Section 2.2.1.5).

2.2.1.17 Sanger sequencing

Sequences of cloned inserts which showed the correct size in agarose gel electrophoresis were verified by Sanger sequencing. 5 µl of plasmid DNA (at 80-100 ng/µl) or 5 µl of PCR product (at 20-80 ng/µl) were mixed with 5 µl of primers (at 5 µM), which either annealed in the plasmid backbone or on the amplified PCR fragment, and send to GATC Biotech AG (Germany) for sequencing.

2.2.1.18 Diethylpyrocarbonate-treatment of dH_2O

To inactivate RNases, dH_2O was treated with DEPC before use in RNA isolation procedures. 1 ml of DEPC was added to 1 l of dH_2O and the solution was shaken vigorously to bring the DEPC into solution. Following incubation at 37°C overnight, the solution was autoclaved to remove traces of DEPC.

2.2.1.19 RNA isolation of human cells

Trizol reagent was used to isolate RNA from human cells in culture. Before isolation, growth medium was removed from the cells and cells were rinsed with 1x PBS. 1 ml Trizol reagent was directly transferred into the cell culture dish or on the cell pellet thereby lysing cells. RNA was then isolated according to the manufacturer's recommendation. To help precipitation and for visibility of small RNA amounts, 1 µl of GlycoBlue was added to the sample when precipitating RNA by adding isopropanol during the procedure. Following isolation, RNA was dissolved in 30-50 µl of DEPC-treated water. The obtained RNA was stored at -80°C.

2.2.1.20 DNase I treatment of RNA

Preceding reverse transcription (Section 2.2.1.21) or use in qRT-PCR reactions (Section 2.2.1.22), isolated RNA was treated with DNase I to remove traces of co-purified DNA. The DNase I reaction was performed by incubating the following mixture (Table 2.23) at 37°C for 15 min.

Table 2.23: Composition of DNase I reactions

Compound	Amount
DNase I Reaction Buffer (10x)	2 µl
DNase I	1 µl
RNA	500-4000 ng
DEPC-treated ddH_2O	ad 20 µl

Subsequently, the DNase I enzyme was inactivated by adding 2.2 µl EDTA (50 mM) to the mixture, thorough pipetting and incubation for 10 min at 75°C.

2.2.1.21 Complementary DNA synthesis

For reverse transcription of RNA, the High-Capacity RNA-to-cDNA Kit was used. The kit includes the MultiScribe MuLV reverse transcriptase, dNTPs, a mixture of random octamers and oligo-dT-16 oligonucleotides to generate the first strand of the complementary DNA (cDNA). Per reaction, up to 2 µg of freshly isolated RNA or 9 µl of DNase I-treated RNA was used and reverse transcribed according to the manufacturer's protocol. The cDNA was diluted with 4 volumes of dH_2O before used in subsequent applications.

2.2.1.22 Quantitative real-time PCR

The Fast SYBR Green Master Mix was used for quantitative real-time PCR (qRT-PCR) applications. To achieve and monitor DNA synthesis from cDNA, this master mix includes the AmpliTaq Fast DNA Polymerase, SYBR Green I, dNTPs, Uracil-DNA Glycosylase and the ROX dye as passive internal reference. PCR products are monitored by measuring the increase in fluorescence caused by the binding of SYBR Green dye to double-stranded DNA. Primers used in qRT-PCRs were designed so that the resulting PCR product was 80-120 bp in size and originated from amplification of two exonic sequences, whenever possible. Using Primer3 software, primers with a melting temperarure (T_m) close to 60°C were chosen. qRT-PCRs were set up in 384-well plates with a final reaction volume of 10 µl per well (Table 2.24).

Table 2.24: Composition of qRT-PCRs

Compound	Amount
Fast SYBR Green Master Mix	5 µl
Primer* mixture (FW and RV, each 10 µM)	0.5 µl
cDNA	2-3 µl
ddH_2O	ad 10 µl

*: Primers used are listed in Tables 2.10, 2.11, 2.12, 2.13, 2.14, 2.15 and 2.16

The amplification was performed using the following 2-step PCR program on a PRISM 7900 HT (ABI) cycler using the standard setting and including a melting curve profile (Table 2.25):

Table 2.25: Conditions for qRT-PCRs

Step	Temperature [°C]	Time	Cycles
Initial denaturation	95	20 s	1
Denaturation	95	1 s	
Annealing & Extension	60	20 s	40
Melting	95	15 s	
Melting	60	15 s	1

All reactions were performed in triplicates. To control for primer specificity and contamination of the reaction mixture, minus-reverse transcriptase controls were included. The SDS2.2 software was used to determine the threshold cycle (Ct) independently for each primer pair. The fold-difference in expression for each gene of interest was determined using the comparative C_t Method (2-(Δ)(Δ)C_t method).

2.2.1.23 Next generation sequencing

RNA sequencing libraries were prepared from total RNA with the TruSeq RNA Sample Preparation Kit v2 at the MPIMG sequencing facility. Sequencing was performed on a HiSeq 2000 instrument using paired-end sequencing (2x50bp).

2.2.2 Cell culture methods

2.2.2.1 Culture of human cell lines

The human cell lines A549, HEK293, HeLa, HepG2, MCF7 and U2OS were cultured at 37°C with 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM) containing 10% heat inactivated fetal bovine serum (FBS), 2 mM L-glutamine, 50 µg/ml penicillin and 50 µg/ml streptomycin. Adherent cultures at ~80% confluence were routinely split 1:5 in 10 cm culture dishes as follows: Cells were washed in pre-warmed PBS, 1 ml Trypsin-EDTA (0.25%) was added to the dishes that were placed at 37°C for 5 minutes. After cells were detached from the dishes, 5 ml pre-warmed culture medium was added and the cells transferred to a 50 ml falcon tube. Cells were spun down at 200 xg and plated in new dishes with fresh culture medium. The K562 cell line was cultured under the same conditions but in RPMI 1640 Medium. K562 cells were split every second day to a dilution of 300,000 cells/ml by transferring an appropriate volume of cells to new culture medium. This was preceded by spinning the cells down once a week, removal of exhausted culture medium and resuspension in fresh medium.

2.2.2.2 Transfection of human cells

Lipofectamine 2000 was used to transfect A459 cells, whereas HiPerFect Transfection Reagent was used for transfections of U2OS cells. One day before transfection, cells were plated to reach 60-80% confluence the following day. Depending on the RNA amount needed for further analysis, cells were transfected in 6-well plates (e.g. for qRT-PCR analysis (Section 2.2.1.22) and next generation sequencing (Section 2.2.1.23)) or in 10 cm dishes (e.g. for cellular fractionation (Section 2.2.2.5)). Before transfection, complete culture medium was exchanged to antibiotics-free growth medium. Lipofectamine 2000 was gently diluted in Opti-MEM reduced serum medium and incubated for 5 min before transferred to a reaction tube containing nucleic acids diluted in Opti-MEM. The mixture was incubated for 20 min to allow the nucleic acid - lipid complex to form and then pipetted onto the cells. Table 2.26 reactions gives an overview on the volumes used for respective transfections.

Table 2.26: Composition of human cell transfections

Compound	Amount for transfection of a 6 well	Amount for transfection of a 10 cm dish
Growth medium	800 µl	4 ml
Opti-MEM	2x 100 µl	2x 500 µl
dsiRNA, ASO/plasmid	0.1-50 nM final conc./10-1000 ng	10 nM final conc. of dsiRNA
Lipofectamine 2000	2 µl	10 µl

DsiRNA transfections carried out with HiPerFect were done in 6-well plates. Per well, 200 µl of Opti-MEM was mixed with the dsiRNA (final conc. of 10 nM) and with 6 µl of HiPerFect. The mixture was incubated at RT for 10 min and carefully pipetted on U2OS cells, which were previously seeded in 800 µl complete growth medium.
24 hours post transfections, 1 ml or 5 ml of complete culture medium was added per well of a six well plate or per 10 cm dish, respectively. Initially, the cell lines were transfected with fluorescent dsiRNAs to control for transfection efficiencies.

2.2.2.3 Luciferase assay

For luciferase assays, 10,000 HEK293 or A549 cells were plated in 200 µl complete growth medium per well of a 96-well white plate one day before transfection. Prior to transfection, complete growth medium was removed and replaced with 50 µl antibiotics-free growth medium per well.
200 ng of pGL3-bidirectional promoter plasmid were transfected per well. To this end, 0.5 µl Lipofectamine 2000 was gently mixed with 25 µl Opti-MEM and incubated for 5 min before transferred to a reaction tube containing luciferase plasmids diluted with Opti-MEM to a final volume of 25 µl. The 50 µl reaction was gently mixed and incubated for 20 min before pipetted onto the cells. Transfections of all plasmids were performed in triplicates. The medium was removed 24 hours post transfection and replaced by 25

µl of 1x PBS. 25 µl of Dual-Glo reagent was then added to each well, and the plate was incubated for 10 min on a shaking platform before determination of Firefly luciferase activity using a microplate luminometer. 25 µl of Stop & Glo reagent was further added, incubated for 10 min, followed by determination of *Renilla reniformis* luciferase activity. As a control, the pGL3-bidirectional plasmid not containing any promoter was transfected. All determined promoter-specific Firefly and *Renilla reniformis* luciferase activities were normalized to the respective luciferase activities of the promoter-less plasmid to control for variations between biological replicates.

2.2.2.4 Preparation of cellular extracts

For preparation of cellular extracts, cells were rinsed with PBS, plates were placed on ice and 100-200 µl of RIPA buffer supplemented with proteinase inhibitors was added per well of a 6-well plate. Cells were scraped off the culture surface and lysates were transferred into reaction tubes for sonication (10 bursts, medium intensity). Tubes were then spun for 5 min at 14,000 g and 4°C and the supernatant was transferred into fresh reaction tubes. Samples were kept on ice during this procedure and extracts were snap frozen in liquid nitrogen and kept at -20°C until used for SDS PAGE (Section 2.2.3.2) immediately after preparation.

2.2.2.5 Fractionation of cells

Cells grown on 10 cm dishes were trypsinized, resuspended in complete growth medium, spun for 5 min at 200 g followed by a washing step with PBS and a repeated spin. Cell pellets were lysed in 400 µl lysis buffer. The samples were incubated on ice for 5 min and carefully pipetted up and down 3-4 times before layered on top of 1 ml sucrose buffer and spinning for 10 min at 3,000 g and 4°C. The supernatant was taken into a fresh reaction tube as the 'cytoplasmic fraction'. The residual pellet was gently rinsed with PBS-EDTA (1 mM) before taken up in 500 µl glycerol buffer. An equal volume of nuclear lysis buffer was added, the sample was vortexed for 2 s followed by a 2 min incubation on ice and spinning for 2 min at 14,000 g and 4°C. The supernatant was transferred into a fresh reaction tube as 'nucleoplasmic fraction'. The residual pellet represented the 'chromatin fraction' and was rinsed with PBS-EDTA (1 mM) before taken up in 1 ml Trizol reagent using a syringe and needle. For RNA isolation (Section 2.2.1.19), 200 µl of cytoplasmic and nucleoplasmic fractions were taken up in 1 ml Trizol reagent. All buffers used for fractionation were ice-cold and freshly supplemented with 1 mM DTT, proteinase inhibitors and RNase inhibitor.

2.2.2.6 Generation of stable cell lines using CRISPR

The type II prokaryotic CRISPR (clustered regularly interspaced short palindromic repeats)/Cas adaptive immune system has recently been engineered to be used for genome

editing strategies in eukaryotes by inducing cleavage at genomic loci of interest [Cong et al., 2013]. The pX330 plasmid, expressing a human codon-optimized Cas9 and chimeric guide RNA, was used to clone DNA oligos that served as guide RNAs into the BbsI site. GuideRNA sequences were obtained using the CRISPR Design Tool [Massachusetts Institute of Technology, 2014], and were chosen when their score was >90 and their off-target score <1. U2OS and A549 cells were transfected with 1 µg of pX330 plasmid per 6-well plate (Section 2.2.2.2) and genomic DNA was isolated (Section 2.2.1.1) from a fraction of cells 48 h post transfection to determine successful genomic deletion across the cell population by PCR (Section 2.2.1.2). Subsequently, single cells were seeded by serial dilution into 96-well plates to grow colonies. Successful homozygous genomic deletions across these single-clone colonies was verified by PCR (Section 2.2.1.2).

2.2.3 Protein biochemical methods

2.2.3.1 Determination of protein concentration

Protein concentrations of cellular extracts were determined using the Bicinchoninic Acid Kit for Protein Determination according to the manufacturer's instructions. This assay is based on the reduction of Cu^{2+} to Cu^{1+} by the amino acids cysteine, cystine, tryptophan, tyrosine and the peptide bond under alkaline conditions. Bicinchoninic acid then forms a stable purple-blue complex with Cu^{1+} that is monitored as a measure of the sample's protein amount.
For each experiment, a standard rank was prepared in parallel using bovine serum albumin. Each sample was pipetted in duplicates and at two different dilutions. Absorbances were measured at 562 nm using a microplate reader. Absorbances of the standard rank were used to plot a linear regression curve and calculate protein concentrations of samples.

2.2.3.2 SDS polyacrylamide gel electrophoresis

Protein samples were analyzed by SDS Polyacrylamide gel electrophoresis (SDS-PAGE), which allows to resolve proteins according to their molecular weight. Before loading onto 4-12% Bis-Tris gels, cellular extracts were mixed with Roti Load buffer and incubated at 95°C for 5 min to denature proteins. Protein samples were loaded alongside a protein ladder as size marker and gels were then run in 1x MOPS buffer at 180 V for 50 min.

2.2.3.3 Western blot

Following SDS-PAGE, protein gels (Section 2.2.3.2) were transferred onto polyvinylidene difluoride (PVDF) membranes for 90 min at 120 V and 4°C in transfer buffer. Membranes were blocked in blocking solution for 1 h with slight agitation. Subsequently, membranes were incubated for 1 h at RT or overnight at 4°C with primary antibodies diluted in

blocking solution at the following dilutions: anti-actin - 1:5000, anti-Calr - 1:3000, anti-RB1 - 1:2000 and anti-tubulin - 1:5000.
The membrane was washed three times in PBST for 5 min on a rocking platform at high speed. Horseradish peroxidase-conjugated secondary antibodies specific to primary antibodies were diluted 1:10,000 in blocking solution and incubated on membranes for 1 h at RT. The membrane was again washed three times in PBST for 5 min on a rocking platform at high speed. Blots were developed using enhanced chemiluminescence (ECL) HRP substrate and signals were visualized using a chemolumineszenz imager.

2.2.3.4 Immunocytochemistry

10,000 U2OS cells were seeded in poly-L-lysine-coated 96-well black/clear imaging plates and subjected to dsiRNA-mediated knock-down of ncRNA-RB1 (Section 2.2.2.2). After forty-eight hours, the cells were fixed with 3.7% paraformaldehyde containing Hoechst 33342 for 15 min. Cells were washed twice with PBST for 5 min and permeabilized with 0.1% Triton X-100 for 10 min. Cells were washed twice with PBST for 5 min and blocked in PBS including 2% FBS for 1 h followed by staining with anti-CALR antibody (1:300) for 30 min in blocking buffer. Cells were washed three times with PBST for 5 min before incubation with a fluorochrome-conjugated secondary antibody diluted in blocking buffer (1:400) for 1 h in the dark. Cells were washed three times in PBST for 5 min before mounting of the coverslips onto microscope slides using fluorescent mounting medium. For longer storage, slides were kept at 4°C and protected from light.

2.2.3.5 Cell surface immunocytochemistry and flow cytometry

U2OS cells were treated with dsiRNAs against ncRNA-RB1 or scrambled control dsiRNAs for 36 h and subjected to 2 μM MTX for 12 h. Cells were collected, washed twice with PBS and thereafter incubated with anti-CALR antibody diluted in cold blocking buffer (2% FBS in PBS) for 30 min on ice. Following two washing steps with cold PBS, cells were incubated with Alexa Fluor 488-conjugated secondary antibody in blocking buffer for 30 min on ice. Thereafter, cells were washed in cold PBS, PI was added to the final concentration of 1 μg/ml and samples were analyzed by means of a FACScan cytofluorometer. Isotype-matched IgG antibodies were used as a negative staining control, and the analysis was limited to living (PI-) cells. Data were statistically evaluated by means of the Cell Quest Software package.

2.2.3.6 Macrophage uptake assay

U2OS cells were treated with dsiRNAs against ncRNA-RB1 or scrambled control dsiRNAs for 36 h, subjected to 2 μM MTX for 12 h to induce cell surface exposure CALR and finally stained with orange cell tracker. Peripheral blood mononuclear cells (PBMCs) were isolated from freshly harvested blood samples of healthy volunteers by means of Fi-

coll density gradient centrifugation. Macrophages were purified using a CD11b-positive selection and labeled with CD11b-FITC antibody following the manufacturer's protocol. Fluorescein isothiocyanate (FITC)-labeled macrophages were incubated for 2 h at 37°C with orange cell tracker-stained U20S cells. Cells were recovered and stained with the viability marker 4,6-diamidino-2-phenylindole and immediately acquired on a Cyan ADP flow cytometer. The uptake of apoptotic U2OS cells stained with cell tracker orange by CD11b-FITC-positive macrophages was analyzed using the FlowJo software.

2.2.4 Computational methods

2.2.4.1 Filtering of ncRNA/PCG pairs

The GENCODE V19 annotation (Dec 2013) of long non-coding RNA genes was used to compute the interdistance between each long ncRNA and its closest PCG encoded on the opposite DNA strand. Initially, the distance between both TSSs was required to be between -2 kb and +2 kb. To further define bidirectional long ncRNA/PCG promoters, the interdistance between the TSSs for each pair of long ncRNA/PCG was reduced to a window size of $\geq$0 to $\leq$500 bp, thereby excluding anti-sense overlapping transcript pairs. Furthermore, long ncRNAs that do not overlap a PCG (in any region of the gene) were extracted to yield a final set of 1,107 bidirectionally expressed long ncRNAs.

2.2.4.2 Conservation analysis of promoters

For promoter conservation analysis, the promoter regions of bidirectionally encoded long ncRNA/PCG pairs were defined as those 700 bp regions surrounding the annotated lncRNA TSS (+500 bp upstream and -200 bp downstream). Promoters regions of long ncRNA genes in general and protein-coding genes were defined in the same way. For comparison, intergenic regions of 700 bp were extracted from random locations in the genome and masked for repetitive regions. As the number of long ncRNA bidirectional promoters was 1,107, in order to be fair in the comparison, 10 promoter sets of comparable size were considered for each of the other promoter classes and the results were averaged over 10 datasets. Position-wise conservation scores were computed from the PhastCons vertebrate conservation track from UCSC. An average conservation score was then computed for each of the 700 bp regions, and the distribution of these scores plotted for each class.

2.2.4.3 Analysis of next generation sequencing data

Next generation sequencing data were subjected to the quality control (QC) using defined metrics on an automated quality-control pipeline that combines published tools as FastQC [Andrews, 2014] with in-house standardized methods. All reads that passed quality metrics were mapped to the latest human genome build (UCSC hg19). Sequencing

duplicated reads were removed using PicardTools [Broad Institute, 2014]. For alignment, the most recent version of Bowtie2 [Langmead and Salzberg, 2012] and TopHat2 [Kim et al., 2013] were used. Read counting was performed using Cufflinks2 [Roberts et al., 2011].

2.2.4.4 Determination of immunofluorescence intensity

Following immunocytochemistry of CALR stained U2OS cells (Section 2.2.3.5), nine view fields per well were acquired by means of a Molecular Devices IXM XL automated microscope and images were analyzed for cytoplasmic fluorescence intensity by using the MetaXpress 5.1 software.

2.2.5 Statistical data analysis

Experiments were carried out in minimally three independent replicates. Statistical analyses were performed using two-tailed Student's t-test.

Chapter 3

Results

Statement of contributions

Parts of this project have been published as:

Musahl A.S., Huang X., Rusakiewicz S., Ntini E., Marsico A., Kroemer G., Kepp O., Ørom U.A. (2015). A long non-coding RNA links calreticulin-mediated immunogenic cell removal to RB1 transcription. Oncogene. [Musahl et al., 2015]

The work presented here is the result of collaborative projects. Experiments were performed by myself unless otherwise noted:
Computational analysis to determine the association of ncRNAs with protein coding genes was performed by Evgenia Ntini. Conservation analysis of promoter sequences of PCGs, ncRNAs and ncRNA/PCG pairs was done by Annalisa Marsico. Deep sequencing of ncRNA-RB1 and RB1 depleted RNA samples was performed in the MPIMG sequencing facility. Deep sequencing data were mapped and filtered by Ruping Sun and Marcus W. Albrecht.
Staining of cell surface exposed calreticulin and quantification was performed by Xing Huang. Macrophage uptake assays were done by Xing Huang and Sylvie Rusakiewicz from Guido Kroemer's lab.
The manuscript for the paper was mainly written by Ulf Andersson Ørom and myself.

3.1 Long ncRNA/PCG pairs encoded in the human genome

3.1.1 Association of long ncRNA genes with PCGs

In order to determine the association of long ncRNA genes with PCGs in the human genome, the GENCODE V19 annotation was used and the distance between the TSS of each long ncRNA and the TSS of the closest PCG encoded on the opposite DNA strand was computed (Figure 3.1 A) [Derrien et al., 2012]. When a distance of ±2 kb separating the TSSs of ncRNA genes and neighboring PCGs was allowed, 3,891 long ncRNA of a total of 22,831 annotated long ncRNA genes were detected as PCG-associated. This distance has previously been used to describe the association of long ncRNAs with PCG promoters [Sigova et al., 2013]. In relative terms, 17% of long ncRNA genes were found to be associated with PCG promoters.

The generated data set included 3,891 polyadenylated long ncRNAs with distinct Ensemble transcript IDs that were associated with 2,077 PCGs, meaning that every pair consisted of a distinct long ncRNA transcript but several pairs had the associated PCG in common. Among the ncRNA/PCG pairs, 1,633 (42%) of long ncRNAs are expressed in an overlapping fashion to their PCG partner, when considering TSS interdistances of -2,000 bp to -1 bp. Correspondingly, 2,258 (58%) of the gene pairs are expressed in a non-overlapping fashion, thus their TSSs were 0 bp to +2,000 bp separated from each other.

To get insight into the distribution of distances between the TSSs of divergently encoded long ncRNA/PCG pairs, the frequency of distances was plotted by a density distribution. This revealed a density peak at an interdistance of +100 bp to +200 bp, suggesting that transcription initiation is majorly occurring within very small distances and that these ncRNA genes/PCG pairs are non-overlapping (Figure 3.1 B).

As an architectural feature of the human genome, the divergent organization of PCGs has been described and genes were suggested to be expressed from a bidirectional promoter when both transcription start sites were separated by less than 1 kb of intervening sequence [Adachi and Lieber, 2002]. When assigning the expression of ncRNA/PCG pairs from bidirectional promoters (TSSs separated by 0 bp to +1 kb), it was found that almost half (1,898 ncRNA/PCG pairs, 48.8%) of the ncRNA/PCG pairs of the data set share a bidirectional promoter with their PCG partner.

3.1.2 Functional categories of PCGs expressed from bidirectional long ncRNA/PCG promoters

To assign functional attributes to the set of PCGs expressed from a bidirectional ncRNA/PCG promoter (1,898 ncRNA/PCG pairs), the functional annotation tool of DAVID was used [Huang et al., 2009a],[Huang et al., 2009b]. Gene ontology analyses revealed that the associated proteins are enriched for biological processes such as transcription and its

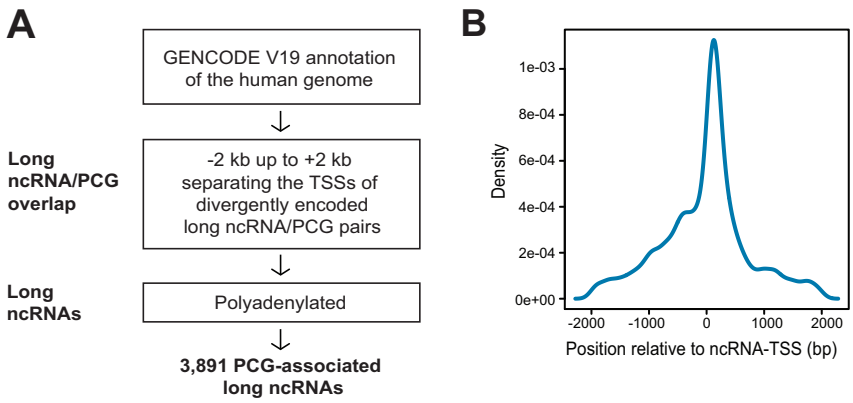

Figure 3.1: Association of long ncRNAs with PCGs in the human genome. A. Criteria for promoters and long ncRNAs to filter for 3,891 PCG-associated long ncRNAs in the GENCODE V19 annotation of the human genome. **B.** Density plot showing the distribution of distances between TSSs of divergently encoded PCG-associated long ncRNAs (±2 kb window).

regulation, RNA transport and RNA localization (Figure 3.2 A). Additionally, pathway association showed that 26 of the proteins are cancer-related, representing the most enriched pathway in numeric terms (Figure 3.2 B). For example, the tumor suppressor RB1 (retinoblastoma-1) [Manning and Dyson, 2011], BRCA-1 (breast-cancer associated-1) [Mavaddat et al., 2013], NFκB (nuclear factor NF-kappa-B p100 subunit) [Hoesel and Schmid, 2013], GSK3β (glycogen synthase kinase-3 beta) [Luo, 2009], and PI3K (phosphatidylinositol 4,5-bisphosphate 3-kinase) [Fruman and Rommel, 2014] are among the cancer-related genes bidirectionally paired with a long ncRNA.

A

GO annotation	Count	%	P-Value	Q-Value
Transcription	157	15.3	2.5×10^{-5}	7.4×10^{-2}
Regulation of transcription	186	18.1	4.6×10^{-5}	6.8×10^{-2}
mRNA transport	16	1.6	6.8×10^{-5}	6.7×10^{-2}
RNA transport	17	1.7	6.9×10^{-5}	5.1×10^{-2}
Establishment of RNA localization	17	1.7	6.9×10^{-5}	5.1×10^{-2}

B

Pathway association	Count	%	P-Value	Q-Value
Pathways in cancer	26	2.5	3.5×10^{-2}	4.7×10^{-1}
Insulin signaling pathway	15	1.5	1.0×10^{-2}	5.7×10^{-1}
Tight junction	14	1.4	2.2×10^{-2}	4.5×10^{-1}
Ubiquitin mediated proteolysis	13	1.3	5.4×10^{-2}	5.8×10^{-1}
Adipocytokine signaling pathway	12	1.2	6.1×10^{-4}	9.3×10^{-2}

Figure 3.2: Functional categories of PCGs associated with long ncRNAs. A. Five most enriched gene ontology (GO) categories of PCGs associated with long ncRNAs. **B.** Five most enriched biological pathways of PCGs associated with long ncRNAs. Count: Number of candidate PCGs included in the corresponding GO category. %: Percentage of candidate PCGs of the total number of PCGs included in the corresponding GO category. P-Value: Calculated using Fisher' Exact Test. Q-Value: Calculated using Benjamini-Hochberg correction.

3.1.3 Final set of bidirectionally expressed long ncRNA/PCG pairs

To simplify all subsequent bioinformatics and laboratory analysis, long ncRNAs within the bidirectional data set of 1,898 ncRNA/PCG pairs that overlap other annotated genes, were removed. Also, the set was narrowed down by selecting for TSS pairs separated by 0 bp to +500 bp (Figure 3.4 A,B), with regard to the major enrichment of small bidirectional promoter sizes. This final data set consisted of 1,107 bidirectionally expressed ncRNA/PCG pairs still including 16 the cancer-related genes (Figure 3.3) (Appendix Table A1).

Genename	Protein
BRCA1	Breast-cancer associated
CEBPA	CCAAT/enhancer binding protein alpha
FN1	Fibronectin
FZD10	Frizzled-10
FZD4	Frizzled-4
GSK3B	Glycogen synthase kinase-3 beta
IKBKB	Inhibitor of nuclear factor kappa-B kinase subunit beta
ITGA3	Integrin alpha3
NFKB1	Nuclear factor NFkappa-B p100 subunit
PDGFA	Platelet-derived growth factor subunit A
PIK3CA	Phosphatidylinositol 4,5-bisphosphate 3-kinase
PIK3R5	Phosphoinositide 3-kinase regulatory subunit 5
PTGS2	Prostaglandin G/H synthase 2
RB1	Retinoblastoma 1
STK4	Serine/threonine-protein kinase 4
SLC2A1	Solute carrier family 2, facilitated glucose transporter member 1

Figure 3.3: Cancer-related genes in the final set of 1,107 bidirectionally expressed ncRNA/PCG.

3.1.4 Conservation of bidirectional ncRNA/PCG promoters

As the promoters of long ncRNA genes have been found to be more conserved than neutrally evolving sequences but less than PCG promoters [Derrien et al., 2012], conservation rates of the final set of 1,107 bidirectional ncRNA/PCG promoters were compared to the conservation of unidirectional long ncRNA promoters and PCG promoters [Derrien et al., 2012]. To this end, conservation scores for promoters of the three classes as well as for random non-genic genomic locations were computed from the PhastCons vertebrate conservation track of the UCSC genome browser [Siepel et al., 2005]. The distribution of these scores was visualized by a density plot (Figure 3.4 C). This analysis confirmed that long ncRNA promoters are only slightly more conserved than random regions in the genome, whereas many PCG promoters showed considerably higher rates of conservation than do long ncRNA promoters. When comparing the conservations of PCG promoters and bidirectional ncRNA/PCG promoters, significantly higher conservation rates were determined for the latter ($p = 6.5 * 10^{-5}$), pointing to an evolutionary selection of this head-to head orientation of ncRNAs and PCGs.

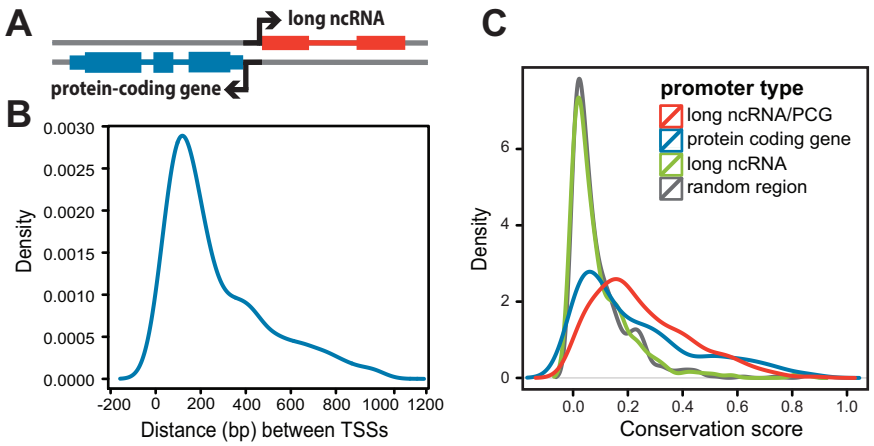

Figure 3.4: Expression of long ncRNAs and PCGs from bidirectional promoters. **A.** Schematic genomic arrangement of bidirectional long ncRNA/PCG pairs. **B.** Density plot showing the distribution of distances between TSSs of divergently encoded long ncRNA/PCG pairs (1 kb window), for those long ncRNAs not overlapping a PCG throughout their genomic region. **C.** Density plot showing the distribution of conservation scores for promoters of bidirectional ncRNA/PCG pairs (as included in the final data set), PCGs, long ncRNAs and random regions. Bidirectional ncRNA/PCG promoters exhibit significantly higher conservation rates than the average PCG promoters ($p = 6.5 * 10^{-5}$).

3.2 Expression of long ncRNA/PCG pairs from bidirectional promoters

3.2.1 Polymerase II occupancy at bidirectional promoters

Pol II ChIPseq data as provided by the ENCODE project (UCSC browser tracks: HEK ChIP Seq: wgEncodeSydhTfbsHek293Pol2StdSig, HEPG2 ChIP Seq: wgEncodeSydhTfbsHepg2Pol2IggrabSig and K562 ChIP Seq: gEncodeSydhTfbsK562Pol2StdSig) was visualized in the UCSC genome browser for the three cell lines HEK293, HepG2 and K562 to determine Pol II occupancy at bidirectional long ncRNA/PCG promoters. Figure 3.5 shows the representative Pol II ChIPseq signals at the promoters of ncRNA-BRCA1/BRCA1, ncRNA-CCNG1/CCNG1, ncRNA-PRKCQ/PRKCQ and ncRNA-RB1/RB1. At each of the promoters examined, a clear Pol II signal was observed across all cell lines, indicative of the assembly of transcription initiation complexes at these promoters and thus expression of the associated genes. Also, for two gene pairs, ncRNA-BRCA1/BRCA1 and ncRNA-CCNG1/CCNG1, a clear double peak of Pol II signal is observable whereas at the promoters of the other two pairs (ncRNA-RB1/RB1, ncRNA-PRKCQ/PRKCQ) a strong major peak as well as a weaker second Pol II peak were visible. This suggests that all four representative bidirectional promoters enable expression of both paired genes in the three cell lines.

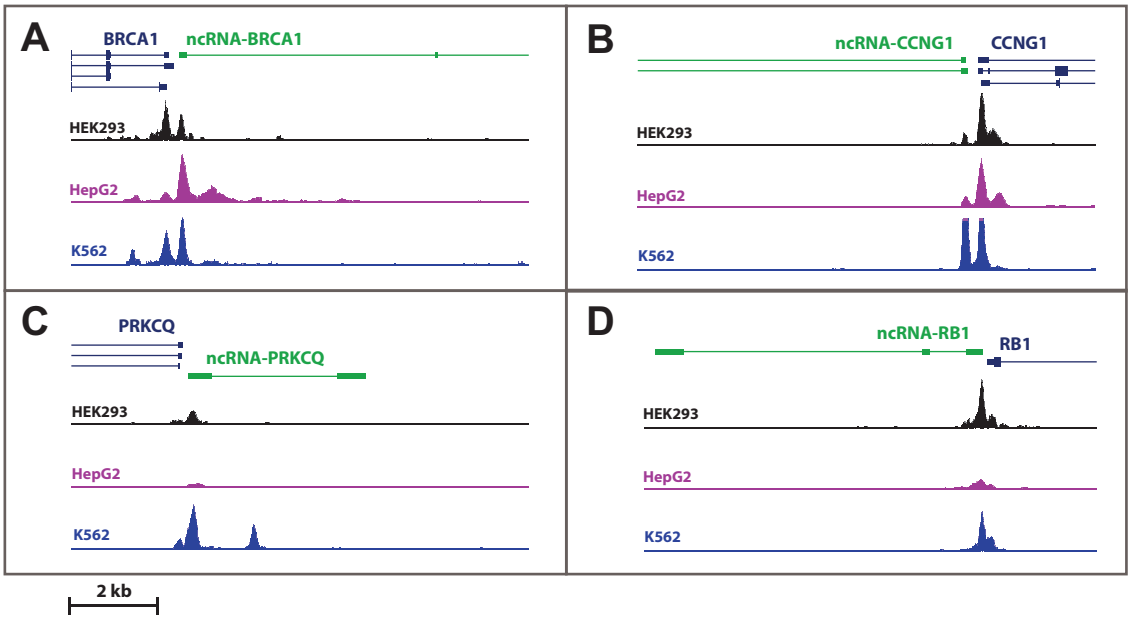

Figure 3.5: Polymerase II ChIP-seq Signal at bidirectional promoters. UCSC genome browser visualization of Pol II ChIP-seq data provided by the ENCODE project. Pol II occupancy at the four bidirectional promoters of **A.** ncRNA-BRCA1/BRCA1, **B.** ncRNA-CCNG1/CCNG1, **C.** ncRNA-PRKCQ/PRKCQ and **D.** ncRNA-RB1/RB1 in the three cell lines HEK293, HepG2 and K562, are observable as double peaks. Peak height was scaled to 300 in all cell lines. wgEncodeSydhTfbsHek293Pol2StdSig, wgEncodeSydhTfbsHepg2Pol2IggrabSig and gEncodeSydhTfbsK562Pol2StdSig tracks were used for the three cell lines.

3.2.2 Capacity of long ncRNA/PCG promoters to initiate transcription bidirectionally

To assess the potential of long ncRNA/PCG promoters to initiate transcription in both directions, five candidate promoters were tested for their bidirectional activity in a promoter reporter assay. Classically, unidirectional promoter activity is determined by cloning a promoter of interest upstream of a luciferase gene. In order to simultaneously monitor promoter activity in both directions, a second luciferase gene, *Renilla* luciferase, was inserted into the Firefly-containing pGL3 basic vector so that the orientation of both luciferases was head-to-head (Figure 3.6 A). In between the two luciferases, the promoters of BRCA1, CCNG1, FKTN, MAGOH and RB1 genes were inserted, by amplification and cloning of intergenic regions separating the TSSs of each respective ncRNA/PCG pair. Only few nucleotides of genic sequence downstream of each TSS were included. All promoters were inserted in the same orientation into the bidirectional reporter vector, thereby Firefly luciferase activity represented the promoter strength in the PCG direction and *Renilla reniformis* luciferase activity represented the promoter strength relevant to expression of the long ncRNA. This assay allowed to simultaneously determine the capacity of the candidate bidirectional promoters to initiate transcription in two directions. All candidate bidirectional promoters as well as the two supposedly unidirectional promoters of GAPDH and TK genes, showed high Firefly Luciferase activity indicative

of their potential to mediate transcription into the PCG direction (Figure 3.6 B). In contrast, solely the five bidirectional ncRNA/PCG promoters induced expression of the *Renilla reniformis* luciferase, which is representative for their potential to initiate transcription of a long ncRNA gene when residing in their genomic context.
To determine the significance of this observation, the ratio between *Renilla reniformis* and Firefly luciferase activities was calculated for each of the assayed promoter:

Table 3.1: Calculation of Promoter Activity Ratios

Promoter		**Ratio**
BRCA1	=	0.85
CCNG1	=	0.30
FKTN	=	0.22
MAGOH	=	0.38
RB1	=	0.43
GAPDH	=	0.01
TK	=	0.07

A value close to 1 was indicative of a high bidirectional potential of the promoter, whereas a low value implied that transcription initiation was mostly unidirectional. The resulting values for candidate bidirectional promoters were found to be significantly higher than for control promoters ($p \leq 0.001$), suggesting that these promoters mediate bidirectional transcription initiation. However, as this reporter assay takes the promoters out of their genomic context, signals for transcription elongation and termination within the associated genes are not considered.

3.2.3 *In vivo* expression levels of ncRNA/PCG pairs

Subsequently *in vivo* expression levels of ncRNA/mRNA pairs transcribed from bidirectional promoters were measured. To this end, RNA from nine human cell lines, routinely cultured in the lab, was isolated and relative expression levels of four transcript pairs, expressed from BRCA1, CCNG1, PRKCQ and RB1 promoters, were determined by qRT-PCR. All transcripts could be detected in the cell lines assayed ($C_t \leq 30$), however relative transcript abundances varied across cell lines for each of the transcripts indicative of the cell lines' diverse origins (Figure 3.6 F).
Although reporter assay studies suggested that promoter activities in both directions were in the same range, *in vivo* transcript levels of ncRNAs and paired mRNAs were differing by one (ncRNA-BRCA1/BRCA1 mRNA pair) or two (ncRNA-CCNG1/CCNG1 mRNA, ncRNA-RB1/RB1 mRNA pairs) orders of magnitude within each of the cell lines assayed. Only for the ncRNA-PRKCQ/PRKCQ mRNA pair, transcript abundances were

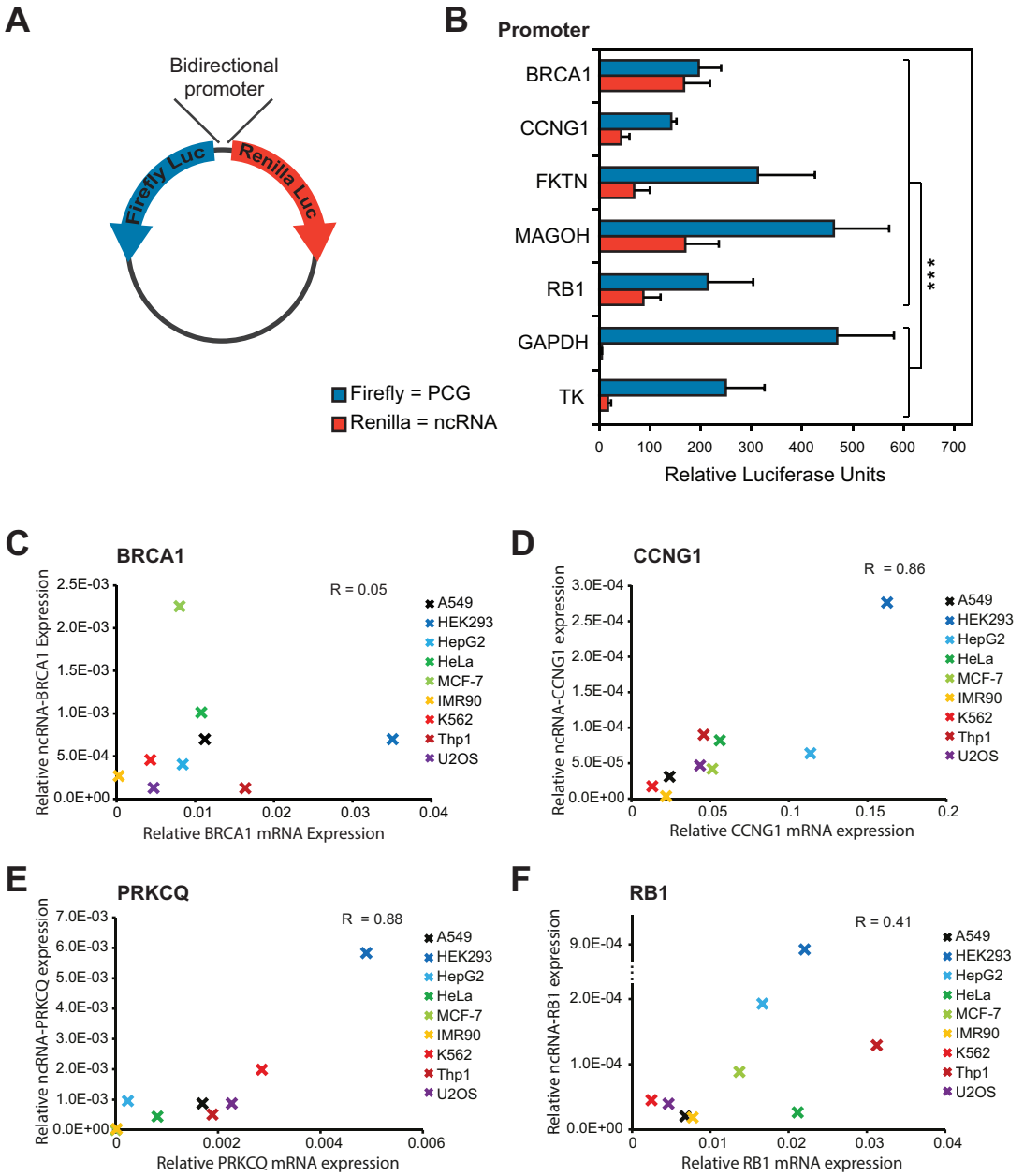

Figure 3.6: Bidirectional promoter activity and expression levels of ncRNA/PCG pairs. **A.** Schematic of the reporter vector to determine bidirectional promoter activity. Promoters were inserted so that Firefly luciferase activity represents the promoter strength in the PCG direction and *Renilla* luciferase activity represents the promoter strength in the ncRNA direction. **B.** Bidirectional promoter activities of CCNG1, FKTN, MAGOH and RB1 candidate bidirectional promoters and GAPDH and TK as control promoters are represented by the relative luciferase units for both Firefly and *Renilla* luciferases. **C-F:** Relative expression levels of paired transcripts. Transcripts expressed from **C.** BRCA1, **D.** CCNG1, **E.** PRKCQ and **F.** RB1 promoters as determined by qRT–PCR. The y axis represents ncRNA expression level and the x-axis the mRNA expression level in nine cell lines (A549, HEK293, HeLa, HepG2, IMR90, K562, MCF7, Thp1 and U2OS). Expression values are presented relative to the value of actin as reference gene. Mean values ± s.d. are shown, n≥3 replicates. $***p \leq 0.05$.

similar within each cell line.

Determination of the correlation coefficient R of the expression levels for each ncRNA/mRNA

pair across the nine cell lines revealed various degrees of correlation. A high correlation of endogenous ncRNA and mRNA expression was observed for the ncRNA-CCNG1/CCNG1 mRNA ($R_{CCNG1} = 0.86$) and ncRNA-PRKCQ/PRKCQ mRNA ($R_{PRKCQ} = 0.88$) transcript pairs, whereas the correlation for ncRNA-RB1/RB1 mRNA ($R_{RB1} = 0.41$) was less pronounced and for ncRNA-BRCA1/BRCA1 mRNA ($R_{BRCA1} = 0.05$) not observable. The absence of correlation for transcript pairs expressed from bidirectional promoters indicates that regulatory mechanisms are involved additional to the regulation of transcription initiation.

3.2.4 Regulation of long ncRNA expression by the exosome complex

The exosome complex as the major eukaryotic 3'-5' exoribonuclease, has been described to control PCG promoter directionality by degrading promoter upstream antisense transcripts (PROMTs) [Preker et al., 2008]. To determine the extent of exosome involvement in the regulation of expression levels of the long ncRNAs bidirectionally expressed from PCG promoters, its core component Rrp40 was depleted from A549 cells. To this end, two dicer-substrate RNAs (dsiRNAs) against the human exosome homologue EXOSC3 were used.

DsiRNAs differ from siRNAs in their length, being 27mer and 21mer duplex RNAs, respectively. While siRNAs are designed to mimic Dicer cleavage products, dsiRNAs are bound and processed by Dicer which supports their incorporation into RNA-induced silencing complexes. This results in increased potency of dsiRNAs at lower transfection concentrations.

Knock-down of EXOSC3 mRNA was highly effective at a final dsiRNA concentration of 50 nM, yielding knock-down efficiencies of >70% for the more effective dsiRNA (Figure 3.7 A). As a control, a dsiRNA was used that was not targeting any RNA of the human transcriptome. The effect of exosome depletion on different transcripts was varying but resulted in an up to 3-fold stabilization of long ncRNAs (e.g. for ncRNA-RB1) (Figure 3.7 B). For long ncRNA transcripts that have been found to be lowly expressed in A459 cells (e.g. ncRNA-BRCA1, ncRNA-RB1), the effect of EXOSC3 knock-down was more pronounced than for higher expressed ncRNA species (e.g. ncRNA-PRKCQ). For comparison, stabilization of three different PROMT species was measured (PROMT 40-9, PROMT 40-33, PROMT 40-54). Two of the three PROMTs were found to be stabilized about 2.5-fold. Their increased abundance is in accordance with previous reports, however stabilization of up to 70-fold have been reported for these ncRNA species [Preker et al., 2008].

Reduction in exosome activity results in comparable effects on bidirectionally expressed ncRNAs and PROMTS, suggesting that the exosome complex regulates expression levels of different transcript species expressed upstream of PCG promoters. The major difference between the assessed species is their degree of splicing, with PROMTS staying

unspliced and the assayed long ncRNA consisting of several exons (Figure 3.5).

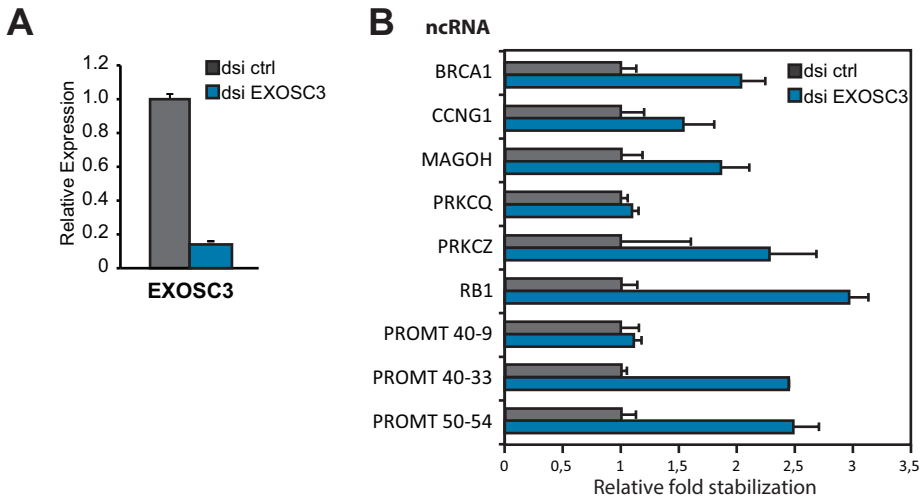

Figure 3.7: Knock-down of the EXOSC3 component of the human exosome complex and stabilization of long ncRNA species. A. Knock-down of the EXOSC3 component of the human exosome complex in A549 cells using one dsiRNA at a final concentration of 50 nM. As a control, A549 cells were transfected with a non-targeting control dsiRNA. **B.** Stabilization of long ncRNA species expressed from bidirectional ncRNA/PCG promoters and of PROMTs upon EXOSC3 knock-down. Expression values are presented relative to the value of actin as reference gene. Mean values ± s.d. are shown, n≥3 replicates.

3.3 The bidirectional ncRNA-RB1/RB1 promoter

3.3.1 Effect of mutations on ncRNA-RB1/RB1 promoter activity

Among the candidate ncRNA/PCG promoters, the RB1 promoter has been intensively studied on the sequence level as well as for the function of its expressed gene product [Gill et al., 1994],[Gill et al., 1994]. By using the RB1 promoter sequence which resides in the bidirectional luciferase reporter vector, the effect of sequence alterations on the expression strength in the RB1 mRNA direction as well as in the ncRNA-RB1 direction were investigated. The cloned RB1 promoter sequence of 198 bp contained the annotated 114 bp-long promoter region separating the TSSs of ncRNA-RB1 and RB1 (according to GENCODE V19 annotation), and at the same time it included 61 bp of the ncRNA-RB1 sequence and 23 bp of the RB1 5'UTR.

3.3.1.1 Mutations within transcription factor binding sites

Within the RB1 promoter sequence, binding sites of at least five transcription factors that are responsible for regulated expression of the RB1 gene have been identified: ATF, p53, E2F1, E4TF1 and Sp1 [Sakai et al., 1991],[Shiio et al., 1992],[Shan et al., 1994],[Savoysky et al., 1994]. ATF, SP1 and E4TF1 are activators of the RB1 gene, and p53 and E2F1 (in complex with pRB1 itself) act as repressors [Sakai et al., 1991],[Shiio et al., 1992],[Gill

et al., 1994],[Sowa et al., 1997],[Hamel et al., 1992]. An *in vitro* study on RB1 promoter activity established that deleting a 17-bp region overlapping the binding sites of SP1, ATF and E2F1 completely abrogates promoter activity in the RB1 direction [Sakai et al., 1991]. Furthermore, two point-mutations within the binding sites of the activating transcription factors ATF and SP1 have been identified by their negative effects on RB1 expression and for being causal in the development of retinoblastoma (Figure 3.8 A) [Gill et al., 1994].

To determine the negative effect of these three mutations on the bidirectional RB1 promoter activity, the RB1 promoter within the bidirectional reporter vector was mutated. The 17 bp deletion resulted in a drastically reduced promoter activity into the RB1 direction as well as into the ncRNA-RB1 direction as represented by the decrease in Firefly and *Renilla reniformis* activities when compared to the wild-type promoter activity (Figure 3.8 B). Also, the single nucleotide G→A transition in the SP1 binding site and the G →T transition in the ATF binding site resulted in significant reductions of the bidirectional promoter activity, equally affecting both transcriptional directions (Figure 3.8 B). This reduction of divergent promoter activity, suggests that transcription initiation in the two directions of the promoter is co-regulated.

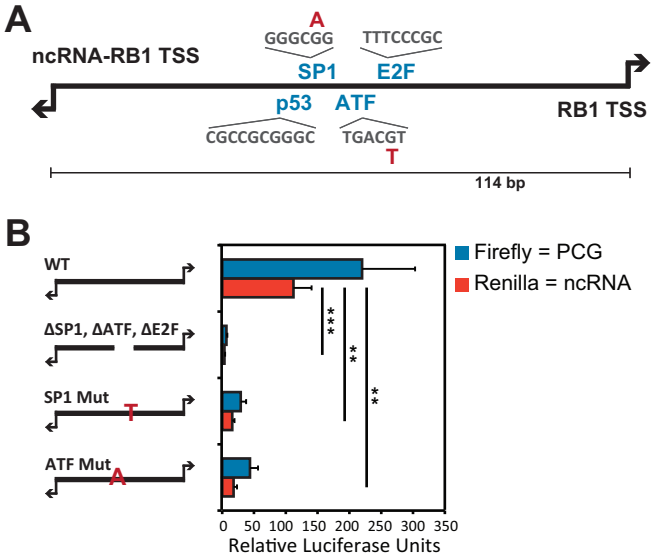

Figure 3.8: The ncRNA-RB1/RB1 promoter and its bidirectional activity. A. Schematic of the ncRNA-RB1 promoter. ncRNA-RB1 and RB1 TSSs are separated by a 114 bp promoter region that encodes the binding sites of ATF, E2F, p53 and SP1 transcription factors. Two retinoblastoma-associated point mutations in the ATF (G→T, red) and SP1 binding sites (G→A, red) are shown. **B.** Bidirectional RB1 promoter activities of the wild-type promoter (WT), the promoter with deleted SP1, ATF and E2F binding sites (ΔSP1, ΔATF and ΔE2F), the RB1 promoter with mutated SP1-binding site (SP1 Mut) and the RB1 promoter with mutated ATF-binding site (ATF Mut) are represented by relative luciferase units for both Firefly and *Renilla* luciferases. The mean values ± s.d. are shown, n≥3 replicates. $**p \leq 0.01$, $***p \leq 0.005$.

3.3.1.2 Mutations within core promoter elements

Furthermore, the possibility to selectively interfere with transcription initiation in only one direction of the bidirectional ncRNA-RB1/RB1 promoter, thereby uncoupling both promoter directions, was tested. The 114 bp RB1 promoter sequence has been characterized as GC-rich and to lack a TATA-box, two features often observed in metazoan bidirectional promoters. However, there is also no CCAAT box present, a sequence element enriched in bidirectional promoters [Trinklein et al., 2004].
To test for the possibility that bidirectional transcription initiation is regulated by individual CPEs encoded on either DNA strand, the RB1 promoter sequence was examined for the presence of BREu, BREd and Inr elements in both directions. When determining the location of the two core promoters, up to three mismatches (MM) towards the consensus sequence of each CPE were considered as well as the appropriate spacing of the CPEs towards each other. In the RB1 direction of the promoter degenerated BREu (2MM), BREd (3MM) as well as INR (1MM) elements were located in appropriate spacing (Figure 3.9 A). In the ncRNA-RB1 direction of the promoter, the same elements were found: BREu (3MM), BREd (3MM) as well as Inr (1MM). A TATA box could not be located in any direction. The observed elements also fit with the CAGE signal observed at the RB1 promoter, as provided by the ENCODE consortium (Figure 3.9 B) (UCSC browser tracks: CAGE A549: wgEncodeRikenCageA549Cell and CAGE HEPG2: wgEncodeRikenCageHepg2Cell). In order to interfere with transcription initiation from the promoter solely in one direction, each of the two potential core promoters was mutated individually.
In the direction of ncRNA-RB1, two point mutations were simultaneously introduced into the Inr element, thereby generating a total of three mismatches towards the Inr consensus sequence (Inr 2mut) (Figure 3.9 C). Also, also a 3 base pair deletion was generated in the Inr element (Inr 3del) as well as a complete deletion of the Inr element was conducted (Inr 9del). *Renilla* luciferase activity was activity was reduced by ~50% by all three mutations. However, this reduction was also observed for Firefly luciferase activity, demonstrating that both promoter directions are equally affected.
The RB1 direction of the promoter was deleted of its BREu element as well as of three additional base pairs (BREu 10del), thereby also interfering with the spacing of the CPEs. However, Firefly and *Renilla* luciferase activities were not affected (Figure 3.9 C).
Additionally, a TATA box element was specifically introduced in the promoter direction of ncRNA-RB1, reasoning that it might increase promoter activity solely in the ncRNA-RB1 direction, as the TATA box has been described for its regulation of promoter directionality [Core et al., 2012]. Introduction of this mutation (TATA cons) resulted in a significant increase of promoter activity in both reporter gene directions, however the increase of *Renilla* activity was more pronounced (Figure 3.9 C).

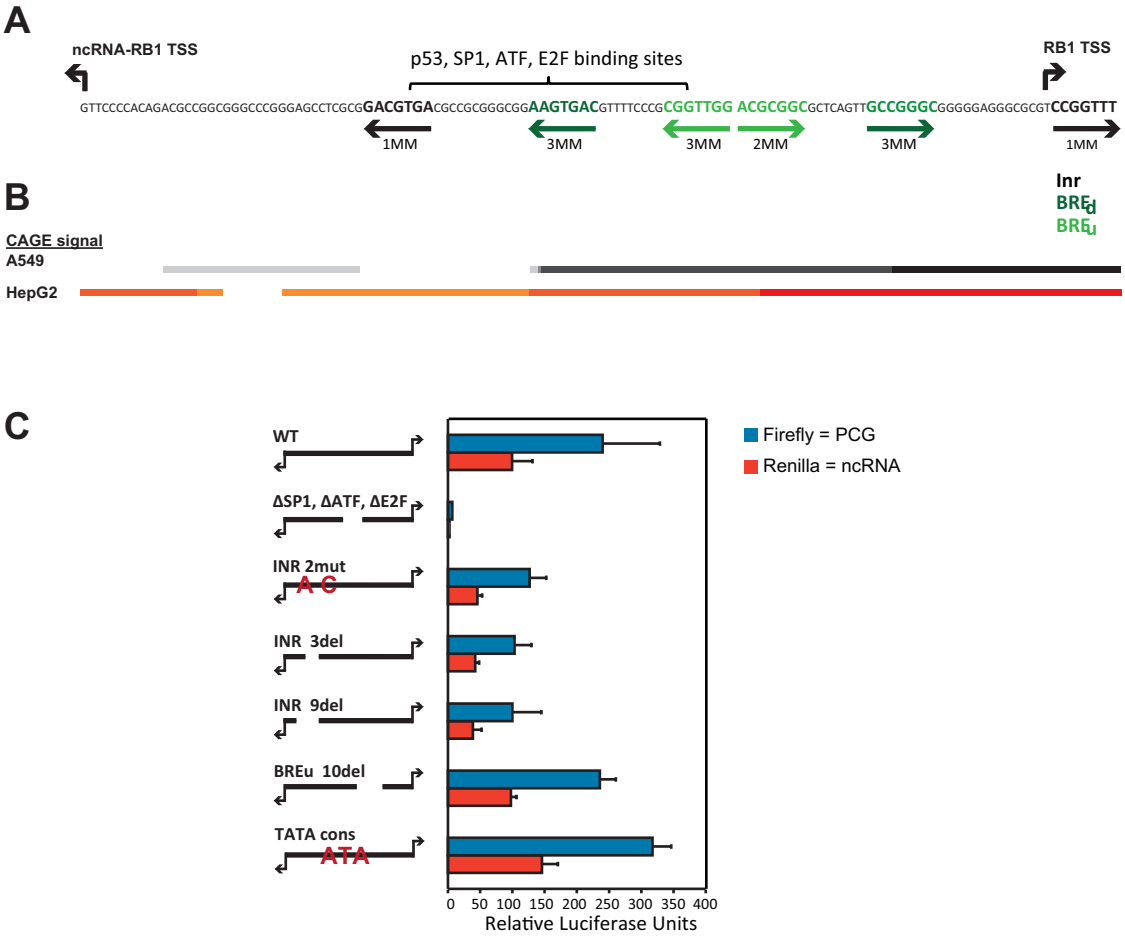

Figure 3.9: Core promoter elements within the ncRNA-RB1/RB1 promoter. A. Depiction of the 114 bp promoter sequence separating the TSSs of the ncRNA-RB1 and RB1 genes (black arrows) including ATF, E2F, p53 and SP1 transcription factors binding sites. Inr, BREu and BREd core promoter elements are illustrated in black, light green and dark green, respectively. Their directionality in the promoter is shown by arrows in the respective colors and number of mismatches (MM) towards their consensus sequence are indicated. **B.** CAGE signal at the ncRNA-RB1 and RB1 promoter in A549 (grey) and HepG2 cells (orange) as provided by the ENCODE consortium (wgEncodeRikenCageA549Cell, wgEncodeRikenCageHepg2Cell). Darker colors correspond to higher signal intensity observed at the respective sites. **C.** Bidirectional RB1 promoter activities of the wild-type promoter (WT) and the promoter with deleted SP1, ATF and E2F binding sites (ΔSP1, ΔATF and ΔE2F) are shown as well as the bidirectional promoter activity upon introduction of CPE mutations into the ncRNA-RB1 direction of the promoter: introduction of 2 point mutations into the Inr sequence (INR 2mut), of a 3-nt deletion into the Inr sequence (INR 3del) and of a complete deletion of the Inr sequence (INR 9del). Bidirectional promoter activity upon deletion of the BREu in the RB1 direction (BREu 10del) and upon introduction of a TATA box in the ncRNA-RB1 direction (TATA cons) is also depicted. All activities are represented as relative luciferase units for both Firefly and Renilla luciferases. The mean values ± s.d. are shown, n≥3 replicates.

3.4 Transcript characteristics of ncRNA-RB1

3.4.1 Transcript structure

According to the GENCODE V19 annotation, ncRNA-RB1 is encoded in a 7.1 kb genomic region and annotated to consist of two isoforms (Figure 3.10 A).
A maximum transcript length of 1.14 kb can be inferred from this annotation when adding the length of all three predicted exons. To verify the transcript structure, a region spanning exon 1 to exon 3 was amplified by PCR using primers that anneal within these two exons and cDNA from A549 and HEK293 cells. Sequencing of the resulting PCR products led to the identification of four isoforms with splice variants for exon 1, exclusion of exon 2 and inclusion of an additional fourth exon (located between exon 2 and 3) in addition to the two annotated splice forms (Figure 3.10 A). A maximum transcript length of 1.29 kb was then calculated. The presence of multiple splice forms for ncRNA-RB1 hints that this long ncRNA could have of diverse functions.

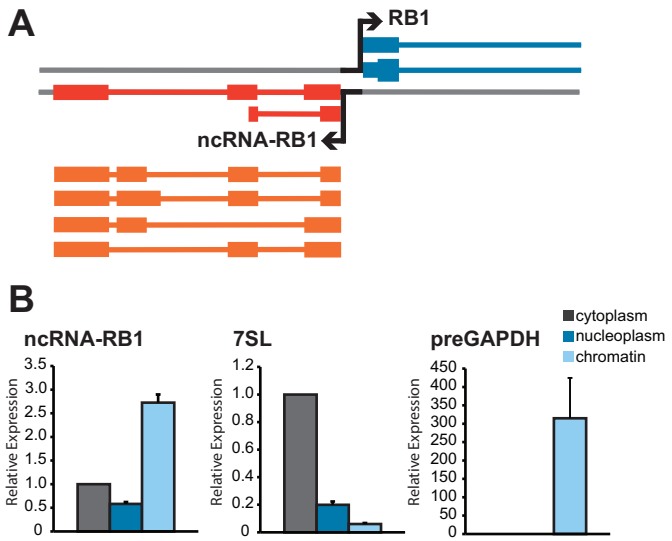

Figure 3.10: Transcript structure and cellular localization of ncRNA-RB1. A. Schematic of the ncRNA-RB1 genomic locus in the human genome. Two splice forms of ncRNA-RB1 are annotated in the GENCODE V19 annotation (depicted in dark orange). Four additional splice forms were detected in PCRs using cDNA of A549 and HEK293 cells (depicted in light orange). **B.** Localization of transcripts in cellular fractions from A549 cells. The relative expression of ncRNA-RB1 within the cellular fractions was determined by qRT–PCR. Relative expressions of 7SL and pre-GAPDH transcripts were determined as control for the purity of cellular fractions. The mean values ± s.d. are shown, n=3 replicates.

3.4.2 Cellular localization

To get insight into the cellular distribution of ncRNA-RB1, A549 and HeLa cells were chemically separated into cytoplasmic, nuclear and chromatin fractions. Following RNA isolation and reverse transcription, the relative abundance of ncRNA-RB1 in each fraction was determined by qRT-PCR. For this analysis, equal amounts of RNA per fraction

were used in reverse transcription reactions. To control for the purity of cellular fractions, relative levels of 7SL and pre-GAPDH RNA were measured as marker for cytoplasmic and chromatin fractions, respectively. 7SL (signal recognition particle RNA) is predominantly localized in the cytoplasm, due to its location at the endoplasmic reticulum membrane, whereas pre-GAPDH, being the unspliced form of GAPDH mRNA, is exclusively found in the chromatin-associated RNA fraction. NcRNA-RB1 was identified to be predominantly localized in the chromatin-associated fractions in both cell lines (Figure 3.10 B), suggesting a function in chromatin organization or transcriptional gene regulation for the ncRNA. Independent of the splice form of ncRNA-RB1, the predominant localization at chromatin of the ncRNA-RB1 was confirmed (data not shown).

3.5 Gene regulatory effects of ncRNA-RB1

To get insight into the gene regulatory functions of ncRNA-RB1 and its involvement in the regulation of RB1 expression or function, cellular levels of the transcript were modified experimentally. This approach was followed by interrogation of the effect on individual target genes or on all cellular genes by taking advantage of high-throughput methods.

3.5.1 Cellular depletion of ncRNA-RB1 and RB1 using dsiRNAs

To deplete ncRNA-RB1 from A549 cells, two dsiRNAs targeting the short first and second exons of the ncRNA, were used reasoning that most ncRNA-RB1 isoforms share these exons and will be targeted (Figure 3.10 A).
Both dsiRNAs were separately transfected into A549 cells at increasing concentrations. Knock-down efficiency of ncRNA-RB1 was determined by harvesting cells 24 hours after transfection for RNA isolation and qRT-PCR. An increasing knock-down efficiency was observed at higher dsiRNA concentrations up to 10 nM, yielding about 75% depletion for dsiRNA 1 and >90% depletion for dsiRNA 2 in the initial screening (Figure 3.11 A, data not shown). Similarly, cells were increasingly depleted from RB1 mRNA when titrating dsiRNA concentrations up to 10 nM (Figure 3.11 D, data not shown). The utilized dsiRNAs against RB1 mRNA targeted the exons 2 and 4, thereby depleting most annotated splice forms of the transcript. Knock-down efficiencies scaled up to 50% for dsiRNA1 and >80% for dsiRNA2 when compared to control transfected cells. As a consequence, only dsiRNA2 agains RB1 mRNA was used in subsequent experiments. Western blot analysis for pRB1 following knock-down of RB1 mRNA confirmed a drastic reduction also on RB1 protein levels (Figure 3.11 E).

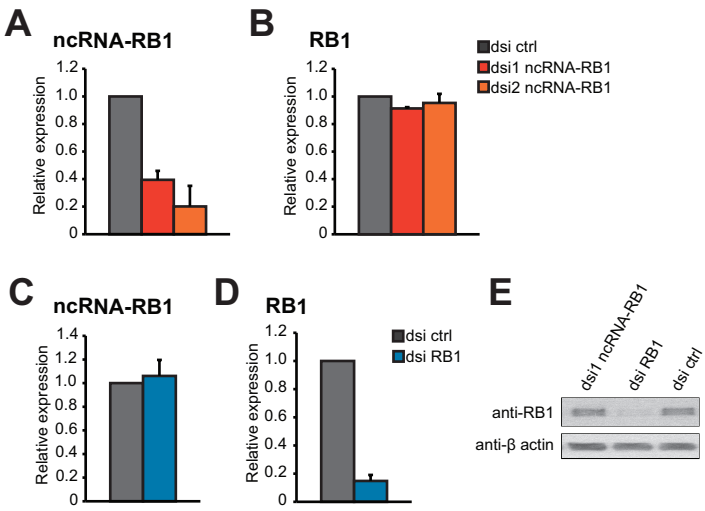

Figure 3.11: Cellular depletion of ncRNA-RB1 and RB1 mRNA. A. Knock-down of ncRNA-RB1 in A549 cells using two different dsiRNAs against ncRNA-RB1 (10 nM). As control, A549 cells were transfected with a non-targeting dsiRNA. **B.** Determination of RB1 expression levels following knock-down of ncRNA-RB1 using two dsiRNAs. **C.** Determination of ncRNA-RB1 expression levels following dsiRNA-mediated knock-down of RB1 mRNA. **D.** Knock-down of RB1 mRNA in A549 cells using one dsiRNA. As control, A549 cells were transfected with a non-targeting dsiRNA. Expression levels are presented relative to the value of actin as reference gene as determined by qRT–PCR. The mean values ± s.d. are shown, n≥3 replicates.**E.** Western blot analysis to determine RB1 protein levels following dsiRNA-mediated knock-down of RB1 mRNA and of ncRNA-RB1 in A549 cells. β-actin protein levels are depicted as loading control.

3.5.2 Effect of ncRNA-RB1 knock-down on RB1 expression levels

Several studies have suggested that long ncRNAs, transcribed upstream of PCGs function in positively or negatively regulating the expression of their PCG partner [Wang et al., 2008],[Grote et al., 2013]. To address the impact of ncRNA-RB1 depletion on RB1 mRNA levels, A549 cells were separately transfected with both dsiRNAs against ncRNA-RB1. Subsequent qRT-PCR and western blot analysis revealed that reduction in ncRNA-RB1 transcript levels did not affect RB1 mRNA or protein levels (Figure 3.11 A,B,E). Likewise, knock-down of RB1 mRNA had no effect on ncRNA-RB1 expression levels (Figure 3.11 C,D). These results indicate that both paired genes, ncRNA-RB1 and RB1, are not regulating each other's expression but might rather contribute to independent regulatory functions. Even a complete knock-out of ncRNA-RB1, by deletion of the locus from the genome, did not affect expression of RB1 mRNA (as described in Section 3.5.7.3) (Figure 3.17 A,B).

3.5.3 Effect of ncRNA-RB1 and RB1 knock-down on the cellular transcriptome

To gain insight into the regulatory functions carried out by ncRNA-RB1 and to further infer the involvement of this ncRNA within regulatory circuits of RB1, both transcripts

were individually depleted from A549 cells using the most effective dsiRNA for each of the transcripts (dsiRNA2 against ncRNA-RB1 and dsiRNA2 against RB1 mRNA). Cellular RNA was isolated 24 hours post knock-down, enriched for polyadenylated RNA and subjected to paired-end RNA-sequencing using a Illumina HiSeq 2000 instrument. Following quality control of sequencing data, reads were mapped to the genome and RPKM (reads per kilobase per million reads) values were determined for all transcripts annotated in the human genome. Genes exhibiting RPKM values <0.1 were considered to be not expressed and excluded from subsequent analysis.

When assign potential target genes of ncRNA-RB1 and RB1, genes were considered to be differentially expressed upon knock-down, when the differences in RPKM values between control and knock-down samples was 2-fold. 226 protein-coding genes were determined to be either up- or down-regulated upon knock-down of RB1 and gene ontology analysis using DAVID detected enrichment in for these genes in biological processes such as cell signaling, programmed cell death or cell growth, in agreement with known functions of pRB1 (Appendix Tables B1,B2 and B3) [Huang et al., 2009a],[Huang et al., 2009b]. Knock-down of ncRNA-RB1 resulted in the differential (increased or reduced) expression of 200 genes that did not show enrichment for any particular gene ontology term. Of the 226 and 200 potential target genes of RB1 and ncRNA-RB1, respectively, 68 (~ 30%) were overlapping, suggesting a function of both genes in common biological pathways (Figure 3.12 A).

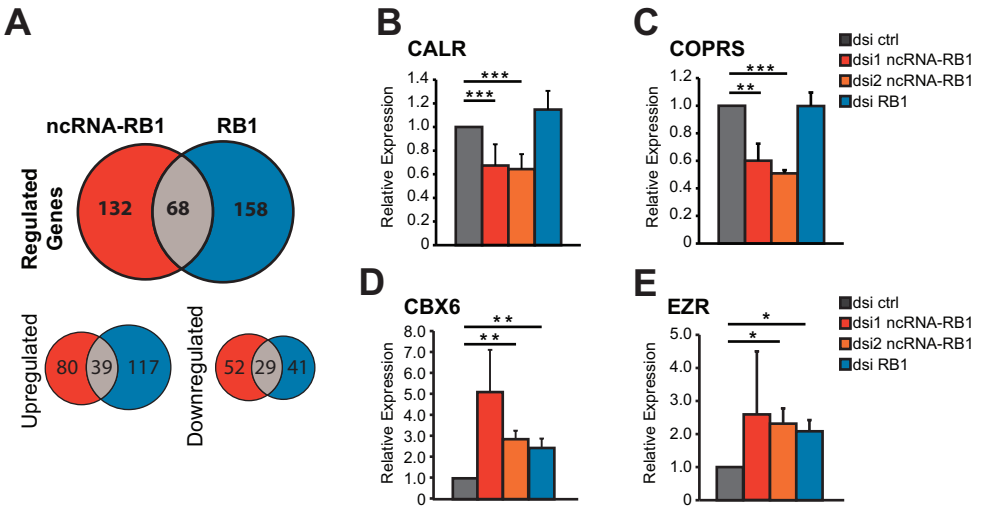

Figure 3.12: ncRNA-RB1 and RB1 regulate distinct and common target genes. A. Venn diagram depicting the number of regulated genes upon knock-down of ncRNA-RB1 (red) and RB1 (blue) determined as described in Section 3.5.3. Upper diagram shows the total number of regulated genes for each knock-down. The genes commonly regulated by ncRNA-RB1 and RB1 are depicted in grey. Lower diagrams show number of up- and down-regulated genes. **B-E.** Determination of relative expression levels of the ncRNA-RB1 and RB1 target genes **B** CALR, **C** COPRS, **D** CBX6 and **E** EZR following knock-down of ncRNA-RB1 and knock-down of RB1 mRNA. Expression levels are presented relative to the value of actin as reference gene as determined by qRT–PCR. The mean values ± s.d. are shown, n≥3 replicates. $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.005$.

3.5.4 Confirmation of ncRNA target genes by qRT-PCR

Among the ncRNA-RB1 target genes were the ER protein CALR (calreticulin) [Michalak et al., 2009], the chromatin modifying protein CBX6 (chromobox protein homolog 6) [Vandamme et al., 2011], the histone binding protein COPRS (coordinator of PRMT5) [Lacroix et al., 2008], the cytokines CXCL5 (chemokine (C-X-C motif) ligand 5) [Chandrasekar et al., 2003] and IL6 (interleukin 6) [Akira and Kishimoto, 1992], the epithelial cell adhesion molecule EPCAM (epithelial cell adhesion molecule) [Armstrong and Eck, 2003] and epithelial protein EZR (ezrin) [Krieg and Hunter, 1992].
Using qRT-PCR following knock-down of ncRNA-RB1, the genes CALR, CBX6, COPRS and EZR could be confirmed as ncRNA-RB1 targets (Figure 3.12 B,C,D,E). Two of these genes, CBX6 and EZR are regulated by both ncRNA-RB1 and RB1, as determined when depleting RB1 mRNA, whereas CALR and COPRS are only regulated by the ncRNA-RB1.

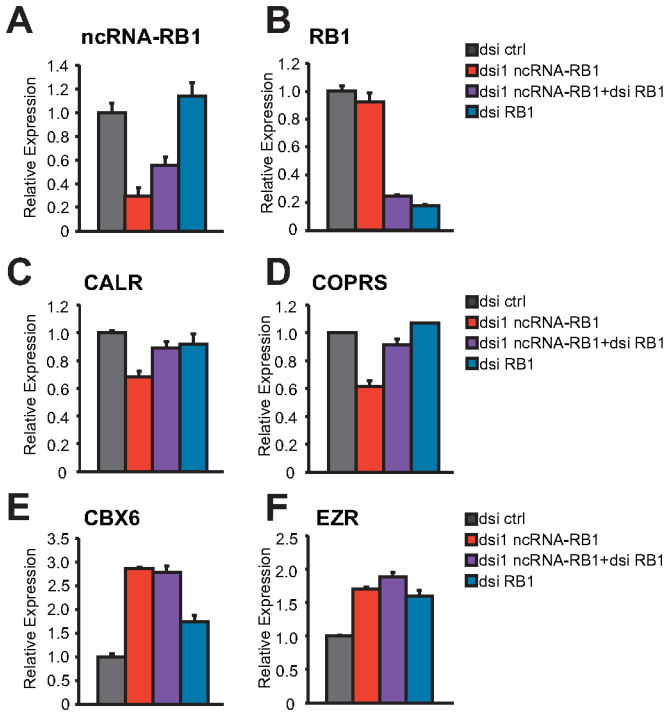

Figure 3.13: Simultaneous depletion of ncRNA-RB1 and RB1 mRNA. A. Determination of ncRNA-RB1 expression levels following knock-down of ncRNA-RB1 and/or RB1 mRNA in A549 cells using one dsiRNA (10 nM) against each transcript. As a control a non-targeting dsiRNA was used. **B.** Determination of RB1 mRNA expression levels following knock-down of ncRNA-RB1 and/or RB1 mRNA. **C-D.** Expression levels of the ncRNA-RB1 target genes **D** CALR and **D** COPRS following knock-down of ncRNA-RB1 and/or RB1 mRNA. **E-F.** Expression levels of the ncRNA-RB1 and RB1 shared target genes **E** CBX6 and **F** EZR following knock-down of ncRNA-RB1 and/or RB1 mRNA. Expression levels are presented relative to the value of actin as reference gene as determined by qRT–PCR. The mean values ± s.d. are shown, n≥3 replicates.

3.5.5 Double knock-down of ncRNA-RB1 and RB1

To get insight into the regulatory interplay of ncRNA-RB1 and RB1, both transcripts were simultaneously depleted from A549 cells and the effect on common and unique target genes was determined. Simultaneous depletion of ncRNA-RB1 and RB1 supported the notion that both genes are not regulating each other's expression as no additional reduction of ncRNA-RB1 levels was observable upon RB1 depletion and vice versa (Figure 3.13 A,B). Also, the target genes specific to ncRNA-RB1, as exemplified by CALR and COPRS did not show a more pronounced decrease in expression upon double knock-down of ncRNA-RB1 and RB1 (Figure 3.13 C,D). Most interestingly, simultaneous depletion of ncRNA and RB1 also had no additive effect on the shared target genes CBX6 and EZR (Figure 3.13 E,F). As the effect of ncRNA-RB1 and RB1 knock-down on common target genes was not additive, both genes are unlikely to target different pathways that independently affect expression levels of these genes.

3.5.6 Overexpression of ncRNA-RB1

Overexpression of ncRNA-RB1 could confirm the regulatory effects of the transcript on its target genes. To this end, the spliced ncRNA-RB1 transcript including exon 1, 2 and 3 was amplified from cDNA and cloned into the pCDNA3 expression vector. A549 cells were transfected with increasing amounts of the plasmid and the increase in ncRNA-RB1 levels was confirmed by qRT-PCR. This confirmed up to 130-fold induction in ncRNA-RB1 transcript levels (Figure 3.14 A). The effect of ncRNA-RB1 overexpression was assayed on the target genes CALR, CBX6 and EZR (Figure 3.14 B,C,D). However, no effect was observed on most of these genes with CALR showing slight but not significant increased expression levels. As a control, expression levels of GAPDH were determined, that were also unaffected by the overexpression of ncRNA-RB1 (Figure 3.14 E).

3.5.7 Calreticulin as a ncRNA-RB1 target gene

Among the numerous target genes of ncRNA-RB1, CALR has been described as a tumor suppressor gene thereby bridging back to the function of the ncRNA's genomically paired gene RB1 also being a tumor suppressor. Additionally, CALR showed the largest change in expression upon depletion of ncRNA-RB1 when considering RPKM values, meaning that its transcript levels are most highly affected in absolute terms. Also, its regulation was found to be specific to ncRNA-RB1.
To further investigate the involvement of ncRNA-RB1 within the regulatory circuits of RB1, CALR was chosen to be further confirmed as a target gene of ncRNA-RB1.

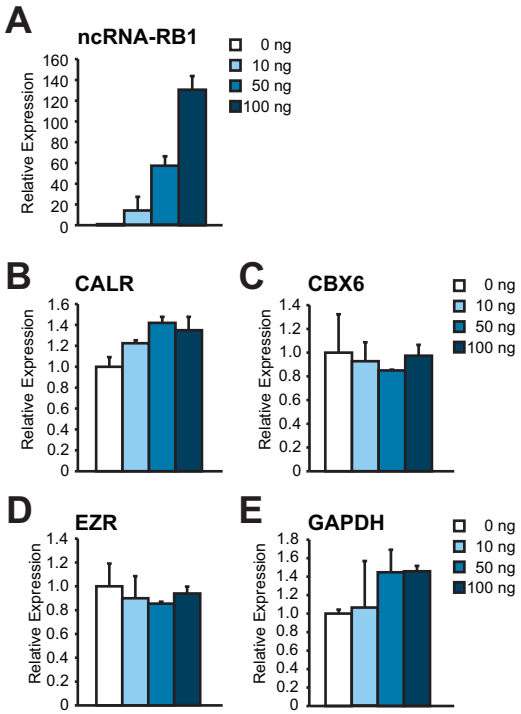

Figure 3.14: Overexpression of ncRNA-RB1. A. Determination of ncRNA-RB1 expression levels following transfection of A549 cells with varying amounts (0-100ng) of the pCDN3 expression plasmid containing the spliced ncRNA-RB1 transcript sequence including exon 1, 2 and 3. **B-D.** Expression levels of the ncRNA-RB1 target genes **B** CALR, **C** CBX6 and **D** EZR following overexpression of ncRNA-RB1. **E.** Determination of expression levels of GAPDH, as a non-related control gene, following overexpression of ncRNA-RB1.

3.5.7.1 Regulation of calreticulin protein levels

In order to determine the effect of ncRNA-RB1 depletion on CALR protein levels, western blot analysis was conducted 48 hours post knock-down of ncRNA-RB1 in A549 cells. This revealed a reduction in CALR protein levels of ~80% as estimated by titrating the amount of cellular protein of control transfected cells (Figure 3.15). Overall, the regulatory effect of ncRNA-RB1 on CALR is even more pronounced on the protein level than on mRNA level.

3.5.7.2 Knock-down of ncRNA-RB1 by antisense oligonucleotides

The use of siRNAs to knock-down transcripts can result in the determination of several false-positive target genes due to off-target effects of this method. To determine false-positives, several siRNAs (with differing seed sequences) against the transcript of interest can be used. Additionally, a different cellular degradation pathway can be employed in order to deplete the transcript such as the use of antisense oligonucleotides (ASOs). While dsiRNAs are incorporated into Argonaute proteins with the help of Dicer which subsequently results in degradation of the target mRNA or its translational inhibition, ASOs pair directly with their complementary RNA and induce RNaseH mediated decay

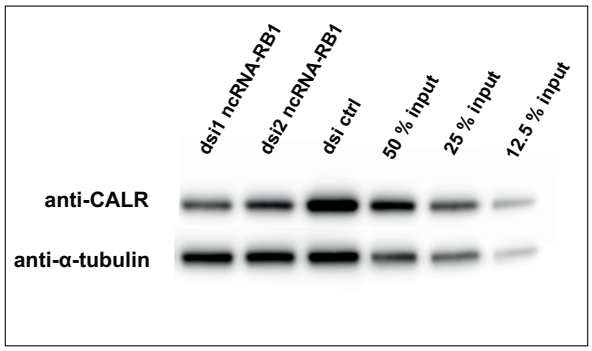

Figure 3.15: Regulation of CALR protein levels by ncRNA-RB1. Western blot analysis to determine CALR protein levels following dsiRNA-mediated knock-down of ncRNA-RB1 (10 nM) in A549 cells. Input protein levels were titrated (12.5-50%) to estimate the effect of ncRNA-RB1 depletion on CALR protein levels. α-tubulin protein levels are depicted as loading control.

of their target RNA.

To further confirm CALR as a target gene of ncRNA-RB1, A549 cells were individually transfected with three different ASOs, designed to target ncRNA-RB1 in its most common exons 1 and 2. As a control, an ASO targeting a non-transcribed genomic region was used.

The potential of ASOs to reduce cellular ncRNA-RB1 levels was most effective at a final concentration of 30 nM (Figure 3.16 A). The levels of ncRNA-RB1 were decreased by up to 70% (for ASO3), which was similar to the effect achieved by dsiRNA mediated knock-down (Figure 3.11 A). Reduction of CALR mRNA levels were confirmed in the ASO knock-down samples (Figure 3.16 B), supporting a regulatory interplay between ncRNA-RB1 and CALR gene.

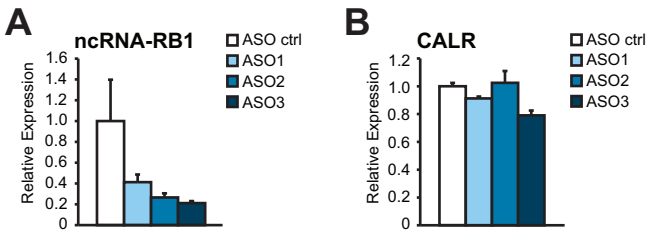

Figure 3.16: Depletion of ncRNA-RB1 using antisense oligonucleotides. A. Determination of ncRNA-RB1 expression levels following transfection of A549 cells with three different ASOs (30 nM). As a control, a non-targeting ASO was used. **B.** CALR expression levels following depletion of ncRNA-RB1 using ASOs. Expression levels are presented relative to the value of actin as reference gene as determined by qRT–PCR. The mean values±s.d. are shown, n≥3 replicates.

3.5.7.3 Knock-out of ncRNA-RB1 by genome editing using CRISPR

Only recently, the type II prokaryotic CRISPR (clustered regularly interspaced short palindromic repeats)/Cas adaptive immune system has been adapted to be used a versatile tool for genome editing in eukaryotic organisms [Cong et al., 2013].

To assess the effect of a permanent inactivation of ncRNA-RB1 on cellular CALR levels,

a stable ncRNA-RB1 knock-out cell line was generated using the CRIPSR technology. To this end, the *Streptococcus pyogenes* Cas9 protein together with a chimeric guide RNA (substituting tracrRNA and crRNA), which mediates targeting of the enzyme to a defined genomic locus, was overexpressed in A549 and U2OS cells by transfection of the pX330 plasmid. After verifying that a ~8 kb genomic region was deleted at a certain ratio within the transfected cell population, cells were individual cells were grown to form colonies. Each cell population derived from these colonies was screened by PCR for its homologous deletion of the ncRNA-RB1 locus. This was followed by qRT-PCR analysis on the ncRNA-RB1 expression levels. One clonal knock-out cell line was thereby confirmed for the U2OS cell line (Figure 3.17 A).
Determination of CALR mRNA levels in the knock-out cell line showed no reduction when compared to control cells (Figure 3.17 B). Therefore, this knock-out strategy did not lend evidence that ncRNA-RB1 is a long term transcriptional activator of the CALR gene.

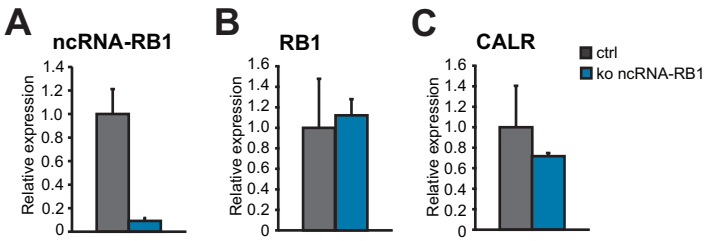

Figure 3.17: Knock-out of the ncRNA-RB1 gene. A. Determination of expression levels of ncRNA-RB1 following knock-out (ko) of the ncRNA-RB1 gene in U2OS cells by genome editing using CRISPR. As a control, WT U2OS cells were used. **B.** Determination of RB1 mRNA levels in ncRNA-RB1 knock-out and control cells. **C.** Determination of CALR mRNA levels in ncRNA-RB1 knock-out and control cells. Expression levels are presented relative to the value of actin as reference gene as determined by qRT–PCR. The mean values ± s.d. are shown, n≥3 replicates.

3.5.7.4 Knock-down of ncRNA-RB1 in U2OS cells

To control for cell line-specific effects of CALR regulation by ncRNA-RB1, the ncRNA was depleted in the U2OS cell line using dsiRNAs (Figure 3.18 A). qRT-PCR analysis confirmed that ncRNA-RB1 acts as an activator of CALR also in this osteosarcoma cell line, as reducing the levels of the ncRNA decreased CALR mRNA levels by ~75% (Figure 3.18 B). This decrease was even more pronounced than in A549 cells, probably also due to their higher steady-state levels of ncRNA-RB1 (Figure 3.6 C). This suggests that the U2OS cell line constitutes a good model to study the biological significance of CALR regulation by ncRNA-RB1.
Reductions in intracellular CALR protein levels in U2OS cells following depletion of ncRNA-RB1 (using dsiRNA1 and dsiRNA2) were confirmed by immunofluorescent stainings 48 hours post transfection. The visualization of intracellular CALR protein confirmed a preponderant localization in the endoplasmic reticulum (Figure 3.18 C). Changes

in cytoplasmic fluorescence, caused by decreased CALR protein levels, were quantified using automated image analysis. A reduction in CALR protein levels of >60% was thereby determined (Figure 3.18 D).

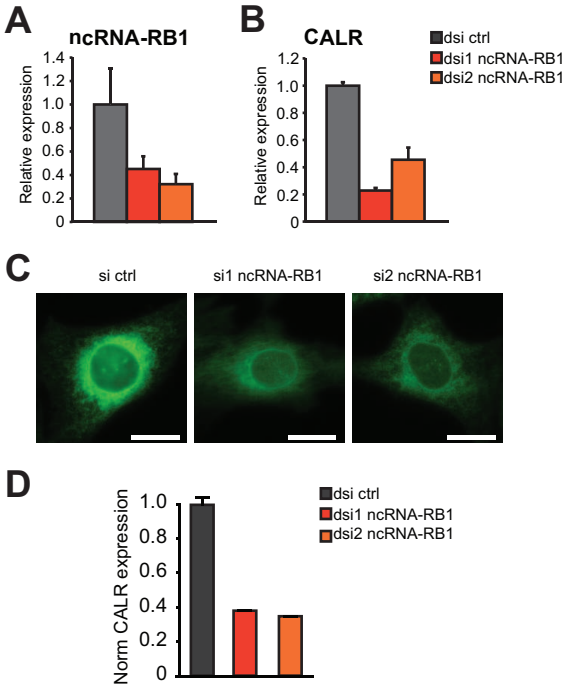

Figure 3.18: Depletion of ncRNA-RB1 in U2OS cells. A. Knock-down of ncRNA-RB1 in U2OS cells using two dsiRNAs against ncRNA-RB1 (10 nM). As a control, U2OS cells were transfected with a non-targeting dsiRNA. **B.** Determination of CALR expression levels following knock-down of ncRNA-RB1. Expression levels are presented relative to the value of actin as reference gene as determined by qRT-PCR. **C.** Representative images of CALR protein expression (green) in U2OS cells treated with dsiRNAs against ncRNA-RB1 or scrambled control siRNAs as obtained by immunohistochemistry. The scale bar equals 10 µm. **D.** Determination of CALR protein expression following depletion of ncRNA-RB1 in U2OS cells. Nine view fields per condition were analyzed by automated image segmentation and the fluorescence intensity of CALR immunostaining was normalized to dsiRNA controls.

3.6 Mechanistic insights into the regulation of calreticulin by ncRNA-RB1

3.6.1 Expression changes of calreticulin across cellular fractions

The change in relative CALR mRNA levels across cellular fractions upon ncRNA-RB1 depletion can elucidate whether regulation of CALR by ncRNA-RB1 occurs at the transcriptional level or post-transcriptionally. A major change of CALR transcript levels in chromatin fractions is indicative of a transcriptional regulation by ncRNA-RB1 as this cellular fraction captures the ongoing transcription. A major relative change in cytoplasmic fractions suggests that CALR is regulated post-transcriptionally.

NcRNA-RB1 depleted A549 cells as well as control cells were separated into cytoplasm,

nucleoplasm and chromatin factions, followed by RNA isolation of each fraction. Besides ncRNA-RB1, RB1 mRNA and CALR mRNA levels, 7SL and pre-GAPDH transcript levels were determined by qRT-PCR to control for purity of the fractions. Also, GAPDH mRNA was measured across all fractions and conditions, to control for non-specific effects due to depletion of ncRNA-RB1. Also, relative GAPDH levels were used to normalize different experiments.

Interestingly, relative CALR mRNA levels were highest in the nucleoplasm and only second most abundant in the cytoplasmic fraction when considering the number of transcripts per microgram of RNA in control cells.

Depletion of ncRNA-RB1 (using dsiRNA1 and dsiRNA2) was most efficient in the cytoplasm and nucleoplasm, which is in accordance with previous studies identifying less effective siRNA mediated knock-down outside of the cytoplasm (Figure 3.19 A). Reductions in cellular ncRNA-RB1 levels, tended to reduce CALR mRNA levels in cytoplasmic and nucleoplasmic fractions (Figure 3.19 B). However, CALR mRNA levels were only significantly decreased in the chromatin-associated RNA fraction.

Expression levels of GPADH mRNA did were not affected across the cellular fractions upon ncRNA-RB1 depletion (Figure 3.19 C).

These result could be confirmed when depleting ncRNA-RB1 using ASOs and investigating the effect on CALR across cellular fractions (data not shown).

In summary, ncRNA-RB1 is likely to play a role in transcriptional regulation of CALR expression, as nascent transcript levels are only significantly reduced upon depletion of the ncRNA.

3.6.2 Regulation of the calreticulin promoter by ncRNA-RB1

To determine whether ncRNA-RB1 regulates transcription of the CALR gene at its promoter, the CALR promoter was cloned into a luciferase reporter plasmid upstream of the promoter-less firefly luciferase gene. The construct was transfected into A549 cells and simultaneously, cells were depleted from ncRNA-RB (using dsiRNA1 and dsiRNA2). Luciferase expression levels of the CALR promoter plasmid as well as of a control plasmid harboring the GAPDH promoter, were measured 24 hours post transfection. No differences in luciferase activities were observed for any of the two promoter plasmids (Figure 3.20).

This result is inconclusive, as it is possible that the regulation of CALR by ncRNA-RB1 cannot be recapitulated using a reporter assay.

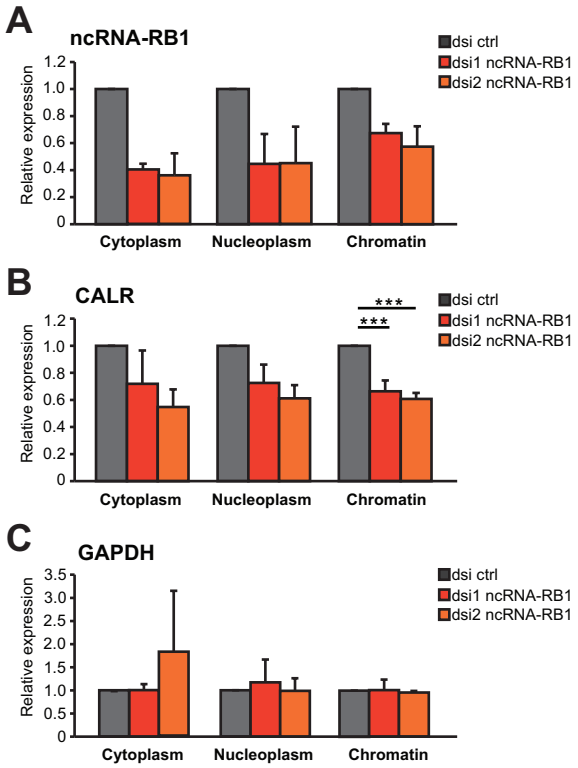

Figure 3.19: Expression changes of calreticulin across cellular fractions. A. Knock-down efficiency of ncRNA-RB1 in cytoplasmic, nucleoplasmic and chromatin fractions of A549 cells using two dsiRNAs against ncRNA-RB1 (10 nM). **B.** CALR expression levels in cytoplasmic, nucleoplasmic and chromatin fractions following knock-down of ncRNA-RB1. **C.** GAPDH expression levels in cellular fractions following knock-down of ncRNA-RB1. The relative expression of ncRNA-RB1, CALR and GAPDH within the cellular fractions was determined by qRT–PCR and normalized to actin as a reference gene. The mean values ± s.d. are shown, n=3 replicates $***p \leq 0.005$.

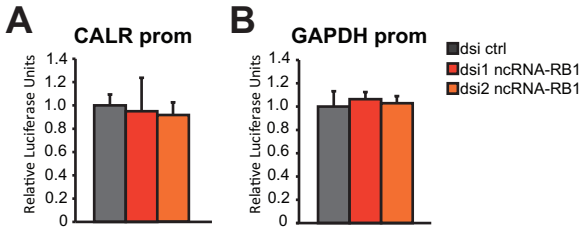

Figure 3.20: CALR promoter activity in ncRNA-RB1 depleted cells. A. CALR promoter activity as determined by the relative Firefly luciferase activity levels in A549 cells upon knock-down of ncRNA-RB1 using two dsiRNAs (10 nM). **B.** Determination of GAPDH promoter activity as a control upon knock-down of ncRNA-RB1. The mean values ± s.d. are shown, n≥3 replicates.

3.7 Consequences of calreticulin regulation by ncRNA-RB1

3.7.1 Impairment of cell-surface exposure of calreticulin

Depletion of CALR has previously been shown to not only reduce cytoplasmic CALR protein levels but also affect the cell's potential to expose CALR on its surface upon treatment with certain chemotherapeutic agents [Obeid et al., 2007]. Among other agents,

mitoxantrone (MTX) causes cell surface translocation of CALR thereby inducing immunogenic cell death [Obeid et al., 2007].

To address whether the decrease in CALR protein levels reduces translocation of CALR to the cell surface, extracellular CALR (ecto-CALR) levels were analyzed upon knock-down of the ncRNA. To this end, U2OS cells were transfected with each of the two dsiRNAs against ncRNA-RB1 or a scrambled control siRNA and treated with 2 µM MTX 48 hours post knock-down. Subsequently, immunocytochemistry was performed on living cells using non-permeabilizing conditions to avoid staining of intracellular CALR. Ecto-CALR levels were then evaluated by flow-cytometry. Ecto-CALR levels were found to be reduced upon MTX treatment under ncRNA-RB1 knock-down conditions when compared to control cells (Figure 3.21 A). Dead cells were excluded by counterstaining with propidium iodide and mean fluorescence intensity as well as the percent of CALR-positive cells were determined (Figure 3.21 B).

The result indicates that translocation of CALR is impaired and is in accordance with the idea that globally reduced CALR levels could affect many aspects of CALR function. However, steady state ecto-CALR levels in uninduced cells are not different between knock-down and control cells, suggesting that the turnover rate of extracellular CALR is low (Figure 3.21 A).

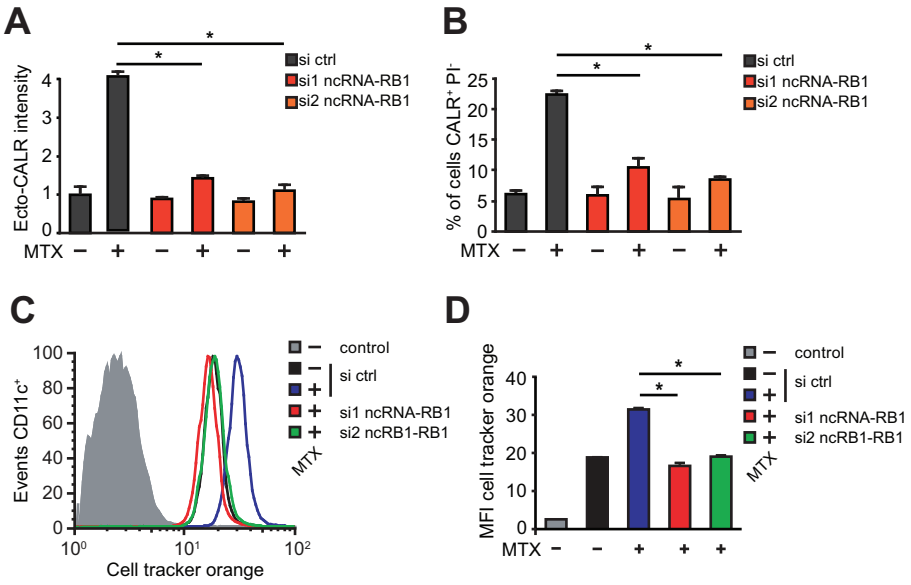

Figure 3.21: Cell-surface exposure of calreticulin and immunogenic cell removal of ncRNA-RB1 depleted cells. A-B. U2OS cells treated with dsiRNAs against ncRNA-RB1 or scrambled control siRNAs were subjected to 2 µM MTX for 12 h and surface-exposed CALR was evaluated by immunohistochemistry and subsequent flow cytometry. Dead cells were excluded by counterstaining with propidium iodide and mean fluorescence intensity as well as the percent of CALR-positive cells were determined. Statistical analysis was conducted between the MTX-treated samples. **C-D.** FITC-labeled macrophages were coincubated for 2 h with celltracker orange stained U2OS cells treated with dsiRNAs and subjected to 2 µM MTX for 12 h as indicated. The celltracker fluorescence intensity of CD11b-FITC-labeled peripheral blood mononuclear cells was measured as an indicator for tumor cell uptake. Mean values ± s.d. are shown, n=3 replicates $*p \leq 0.05$.

3.7.2 Effect on immunogenic cell removal

As cell surface exposure of CALR during apoptosis allows for the occurrence of immunogenic cell death, it is suggestive that cellular alterations in ecto-CALR levels directly interfere with phagocytosis by macrophages. To test this hypothesis, an *in vitro* phagocytosis assay was designed: U2OS cells depleted of ncRNA-RB1 were treated with 2 μM MTX for 12 h and stained using orange cell tracker. Simultaneously, macrophages were isolated from human blood samples and labelled with FITC-conjugated antibodies. Subsequently, macrophages were incubated with pre-treated U2OS cells. Cell fluorescences were acquired on a flow cytometer, allowing to distinguish between macrophages that have engulfed U2OS cells, those that did not and free U2OS cells. As a result, ncRNA-RB1 depletion reduced the uptake of MTX-treated U2OS cells by human macrophages in this *in vitro* phagocytosis assay arguing for a regulation of immunogenic cell removal by ncRNA-RB1 (Figure 3.21 C,D).

Chapter 4

Discussion

Transcription of the human genome is pervasive and by far not restricted to the generation of mRNAs. Although an increasing number of ncRNAs, transcribed upstream of PCG promoters, is detected in genome-wide studies, their functionality has largely not been elucidated.
The present thesis aims to determine and characterize the association of long ncRNA genes with divergently encoded PCGs in the human genome. The extent of co-regulation and co-expression of such long ncRNA/PCG pairs is investigated and for a representative ncRNA/PCG pair, the ncRNA-RB1 and RB1 genes, a functional link is established.

4.1 Transcriptional characteristics of long ncRNA/PCG pairs

Depending on the study, the model system used and applied filters, different sets of promoter associated long ncRNAs are determined [Derrien et al., 2012],[Sigova et al., 2013]. The present study employed the comprehensive and constantly updated GENCODE annotation of human long ncRNAs and PCGs to investigate the association of both gene types in the human genome. A large fraction of long ncRNAs was found to be divergently encoded in the genomic neighborhood of PCGs, more precisely 3,891 out of 22,831 annotated long ncRNA genes (Figure 3.1 A). Preceding this PhD thesis, different studies have investigated the genomic association of long ncRNAs with PCGs in the human genome. Estimates for the intersection of these gene types range from 35% up to 60% [Derrien et al., 2012],[Sigova et al., 2013]. This study determined that 17% of long ncRNA genes are associated with divergently encoded PCGs in the human genome. Although the numbers differ, they emphasize the commonness of this genomic arrangement and suggest a regulatory importance.
Among the 3,891 ncRNA/PCG pairs, 42% were encoded in an overlapping fashion and 49% were found to be expressed from a shared bidirectional promoter, indicating a high potential for transcriptional interference and co-regulation of the paired genes, respectively.

The high number of long ncRNA/PCG pairs, expressed from a bidirectional promoter demonstrated that transcription of long ncRNAs and PCGs from shared promoter sequences is favored by the cell. Furthermore, the bidirectional promoters of ncRNA/PCG pairs were revealed to be on average only few hundred base pairs in size (Figure 3.1 B), and thereby markedly smaller than 1 kb, being the defining criterion of promoter bidirectionality [Trinklein et al., 2004]. This bias for small promoter sizes fuelled the question if pairing of TSSs is advantageous to the cell, for example to allow co-regulation of these long ncRNA/PCG pairs.

4.1.1 Co-regulation of long ncRNA/PCG pairs

Narrowing down the initial data set by filtering for bidirectional ncRNA/PCG promoters of $\leq$500 bp in size, should facilitate to answer this question as the study was consequently focused on ncRNA/PCG pairs expressed from over-represented small promoters (Figure 3.4B). Conservation analysis for this final set of promoters revealed that bidirectional promoters from the defined set were highly conserved and that they exceeded the conservation levels of the average protein-coding gene promoter (Figure 3.4 C). This observation implied that regulatory elements such as transcription factor binding sites, are preserved within these ncRNA/PCG promoters, possibly controlling expression in both promoter directions and thereby of the paired genes. The high conservation rate therefore also suggests that pairing of both genes is conserved across species. As the existence of orthologous ncRNA genes is hard to assess due to low levels of primary sequence conservation, it was not a focus of the present study. However, studies focusing on conservation of bidirectional PCG organization, indicate that such genomic arrangement is often conserved between human and mouse [Adachi and Lieber, 2002],[Trinklein et al., 2004]. High conservation rates of ncRNA/PCG promoters could therefore indicate that in other species these promoters are able to drive divergent transcription initiation.
In support of a co-regulation that results in co-expression of paired genes, two Pol II peaks were observed at analyzed candidate ncRNA/PCG promoters when visualizing Encode ChIP data (Figure 3.5). This implies that for the investigated bidirectional promoters, both genes are expressed in a coordinated and regulated manner. Pol II enrichment is usually observed 20-60 nucleotides inside the genes where the enzyme is paused to resume elongative transcription upon controlled pause release [Muse et al., 2007].
By use of a bidirectional reporter assay, it is demonstrated that candidate bidirectional ncRNA/PCG promoters are capable of initiating transcription in divergent directions, allowing for co-expression of the paired genes (Figure 3.6 B). The assay also illustrated that the promoter strength in bidirectional promoters is high and comparable in both directions. This clearly distinguished candidate bidirectional promoters from supposedly unidirectional control promoters, with the latter favoring transcription initiation in one direction. Although this assay is well suited to determine bidirectional promoter

activity, it does not provide an accurate quantitative assessment of the two promotor directions, since it relies on expression of two reporter genes with different expression efficiencies. However, the study was only focused on determining the potential of promoters for divergent gene expression, thereby confirming bidirectional nature of candidate long ncRNA/PCG promoters. In conclusion, these results were suggesting that elements in the promoter sequence convey bidirectional promoter activity.

Such sequence elements can be transcription factor binding sites or CPEs and mutational studies demonstrated that both types of elements indeed affect transcription initiation in both directions. By using the ncRNA-RB1/RB1 promoter as a model, it is shown that disruption of individual transcription factor binding sites had the potential to reduce promoter activity drastically and equally in both directions (Figure 3.8 B). The dramatic effect of single base pair mutations on bidirectional promoter activity suggests that transcription factor binding is mediating PIC assembly at both TSSs of the promoter. Binding of specific transcription factors to promoter regions is known to precede PIC assembly and recruitment of GTFs, as transcription factors are involved in generating an open chromatin structure, e.g. by interacting with coactivator complexes.

Assuming that PIC assembly is regulated successively, it was considered whether transcription initiation at the paired TSSs can be uncoupled by mutating CPEs present in each of the promoter directions. For the ncRNA-RB1/RB1 promoter, CPEs were found at both promoter ends and to be arranged in a non-overlapping fashion, although the promoter sequence is only 114 bp in size (Figure 3.9 A). Mutations in these potential CPEs reduced bidirectional promoter activity by ~50% (Figure 3.9 C). Interestingly, transcription initiation in the two promoter directions was again equally effected, indicating that bidirectional promoter activity relies on the integrity of all CPEs. Only introduction of a TATA-box into the GC-rich ncRNA-RB1/RB1 promoter sequence shifted promoter activity towards the ncRNA direction. As previously described, TATA boxes appear to regulate promoter directionality [Core et al., 2012].

As the observed effect of CPE mutations on bidirectional promoter activity is less pronounced compared to the mutational effect of transcription factor binding sites, it can be assumed that CPEs either act redundantly or that the predicted elements do not match those used during PIC assembly.

In conclusion, the presented results imply that the candidate bidirectional promoters are able to drive transcription initiation into the PCG and ncRNA-RB1 direction, that these activities are coupled and dependent on the underlying promoter sequence. This emphasizes a transcriptional co-regulation of the investigated genomically paired long ncRNA genes and PCGs.

4.1.2 Co-expression of long ncRNA/PCG pairs

In contrast to these established promoter characteristics, in vivo expression levels of long ncRNAs and paired mRNAs deviated greatly (Figure 3.6 C). Also, expression levels of two out of four candidate transcript pairs were not correlated when measured across different cell lines. This suggests that further regulatory mechanisms are active in vivo, potentially affecting transcriptional elongation and/or post-transcriptional stability of the paired RNAs. In addition, the chromatin context regarding nucleosome occupancy and histone modification certainly differs between a vector and an actual genomic region.

Considering a post-transcriptional regulation, differences in expression levels could be attributable to the exosome complex as has been described for the unstable promoter upstream transcripts (PROMTs) [Preker et al., 2008]. To determine exosome involvement in the regulation of long ncRNA transcript abundance, this major eukaryotic 3'-5' exoribonuclease was depleted from cells. The observed stabilization of long ncRNAs suggests that the exosome complex is involved in degrading transcripts of this class (Figure 3.7 B). Especially low abundant bidirectionally expressed long ncRNAs were found to be stabilized upon exosome depletion.

It has been suggested that early polyadenylation of PROMTs, due to the increased frequency auf poly(A) signals and depletion of splice sites in the upstream direction of the promoter, recruits the exosome complex. In contrast, long ncRNAs of the examined bidirectional ncRNA/mRNA pairs in this thesis are to a great extent spliced. Interestingly, determination of BRCA1, RB1 and PRKCQ mRNA levels following depletion of the cellular exosome showed an increase in abundance of 1.5- to 2-fold. In addition to the observed increase in promoter upstream transcript levels, stabilization of mRNAs by about ~1.5-fold upon exosome depletion has also been described previously [Preker et al., 2008]. This result emphasizes that the exosome regulates stability of many cellular RNA species independent of their expression levels.

In conclusion, genomic encoding of long ncRNA and PCGs as birectional gene pairs appears to be favored by the cell allowing co-expression of such genes. But at the same time, posttranscriptional regulatory mechanisms fine-tune and uncouple the transcriptional output of such bidirectionally expressed gene pairs.

4.2 Functionality of ncRNA-RB1

4.2.1 Regulatory link between ncRNA-RB1 and RB1

As it was found that expression of long ncRNA/PCG pairs frequently occurs from promoters of cancer-related genes (Figure 3.3), one such gene pair was chosen for further analysis.

Characterization of the ncRNA-RB1 transcript, as a representative bidirectionally ex-

pressed ncRNA, revealed that ncRNA-RB1 is lowly expressed and has at least four splice forms in addition to the two annotated ones (Figure 3.10 A). This complexity suggests that expression of the ncRNA is actively regulated and not a by-product of transcription initiation at the RB1 promoter. Additionally, the gene could carry out several regulatory functions as its secondary structure changes as a result of exons inclusion. The predominant nuclear localization and low expression of the transcript further suggests that it could be involved in gene regulatory processes in the nucleus (Figure 3.10 B).
Regarding the functionality of long ncRNAs and respective PCGs, it is demonstrated by the example of ncRNA-RB1 and RB1 that both gene products are not involved in the direct regulation of the partner gene (Figure 3.11). The same is true for the ncRNA-PRKCQ/PRKCQ gene pair, in which both genes do not have transcriptional regulatory effects towards each other (data not shown). These results indicate that promoter-associated long ncRNAs likely fulfill separate functions apart from directly regulating their closest PCG, although such a regulatory interplay has previously been suggested [Uesaka et al., 2014].
In this regard, ncRNA-RB1 and RB1 were found to have to a great extent distinct target genes, but also have a significant overlap in target genes, suggesting a regulatory interplay between the genes (Figure 3.12 A). Regarding the overlap in the genes regulated between ncRNA-RB1 and RB1, it implies the involvement of ncRNA-RB1 and RB1 in a common biological pathway or the common control of an upstream regulator.
With regard to the second possibility, the large number of potential target genes as determined by RNA sequencing suggests that several of these genes are secondary targets. The regulation of a gene such as a transcription factor will entail the differential expression of several other genes. Such secondary targets could be determined in large scale by whole transcriptome analysis at different time points following knock-down of respective genes.
To further investigate a regulatory interplay between ncRNA-RB1 and RB1, both genes were simultaneously depleted from cells. However, common target genes of ncRNA-RB1 and RB1 did not display an additive regulatory effect for the double knock-down when compared to individual depletion of ncRNA-RB1 and RB1 (Figure 3.13). This would be expected for a scenario in which both genes target distinct biological pathways, thereby independently affecting the expression of a shared set of genes.

4.2.2 Gene regulation by ncRNA-RB1

This study provides evidence for a regulation of the CALR gene by ncRNA-RB1, and shows that this regulation affects transcriptional as well as protein levels of CALR (Figures 3.12 B,3.15) and is mediated at the transcriptional level (Figure 3.19 B). Among the ncRNA-RB1 specific target genes, CALR was most highly affected in absolute terms upon depletion of the ncRNA. The significance of this observation becomes clear when

considering the high cellular abundance of this endoplasmic reticulum-sessile protein and its function as a major Ca^{2+}-binding protein. Importantly, CALR has been described as a tumor suppressor gene as has RB1 [Pike et al., 1998],[Pike et al., 1999].
By ncRNA-RB1 acting as a transcriptional activator of the cancer-relevant CALR gene, a regulatory interplay of this ncRNA with the tumor suppressor RB1 is further underlined (Figure 3.12 B). In this scenario, a shared biological function of ncRNA-RB1 and RB1 also extends to their regulation of individual target genes. The idea that co-transcription could entail co-functionality of paired gene products is supported by the observation that divergent PCG/PCG pairs in the human genome (around 10% of all PCGs) are enriched in specific functional categories, such as DNA repair and the regulation of cell cycle and metabolism [Wakano et al., 2012].
Determination of CALR as a ncRNA-RB1 target gene showed that CALR protein levels were even more strongly affected than its transcript levels following depletion of the ncRNA (Figure 3.15). Also, this regulatory interplay could be confirmed by different ncRNA-RB1 depletion techniques (use of dsiRNAs and ASOs) (Figures 3.12 B,3.16 B), as well as across different cell lines (A549 and U2OS) (Figures 3.12 B,3.18 B). However, the generated knock-out cell line for ncRNA-RB1 revealed that the regulatory effect on CALR could not be long term one, suggesting that additional activating mechanisms in the cell can compensate for the loss of the ncRNA (Figure 3.16 C).
With regard to the gene regulatory mechanism mediated by ncRNA-RB1, the low expression levels of the transcript as well as its specific enrichment in the chromatin-associated cellular fraction, implied a transcriptional regulation of its target genes (Figure 3.10 B). Cellular fractionation of ncRNA-RB1 depleted cells allowed to confirm the transcriptional regulation of the CALR gene by ncRNA-RB1, demonstrating a specific reduction in actively transcribed CALR mRNA levels (Figure 3.19 B). This implies that activating transcription factors or co-factors are not efficiently recruited to the CALR gene in the absence of ncRNA-RB1.
Interestingly, ncRNA-RB1 is transcribed from chromosome 13, whereas CALR is encoded on chromosome 19. Such a regulation in trans has been suggested to require higher expression levels of ncRNAs, when assuming their diffusion to the locus of action [Cech and Steitz, 2014]. Direct interaction of both chromosomal loci could reconcile this conflict. However, examination of existing Pol II ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag Sequencing) data of K562 and MCF-7 cells as provided by the ENCODE consortium did not show any long range interactions emanating from the ncRNA-RB1 locus.
Importantly, transcriptional regulation of the CALR gene was found to extend to the cell surface translocation of calreticulin, thereby drastically reducing the induction of ecto-CALR protein levels (Figure 3.21 A).

4.2.3 Biological implications of the functional link between ncRNA-RB1 and RB1

The CALR protein has been shown to serve as an 'eat-me' signal for phagocytic cells when exposed to the plasma membrane [Chao et al., 2010],[Obeid et al., 2007]. The difference in ecto-CALR levels upon ncRNA-RB1 knock-down and induction of CALR translocation was reflected in the reduced uptake of cancer cells by human macrophages (Figure 3.21 C,D). This assigns an important role to ncRNA-RB1 in the regulation of immunogenic cell death.

It has been proposed recently that tumorigenic cells not only need to disable cell-intrinsic death programs (such as apoptosis) but they must also avoid programmed cell removal by phagocytes [Chao et al., 2012]. One of the aims of chemotherapy is therefore to elicit an anticancer immune response by inducing the uptake of dying cancer cells by phagocytes. The possibility that neoplastic cells can simultaneously abrogate expression of RB1 and ncRNA-RB1, caused by RB1 promoter mutation or methylation, has implications for cancer progression. Importantly, hypermethylation of the RB1 promoter is frequently (~15%) observed in human tumors and has been shown do drastically reduce expression of RB1 [Livide et al., 2012],[Ohtani-Fujita et al., 1993].

In later stages of tumor development, when RB1 is lost to achieve uncontrolled proliferation, it could be advantageous for the cell to simultaneously reduce extracellular CALR levels in order to escape macrophage recognition. However, increased ecto-CALR levels have been observed in several cancer types and are indicative of cellular stress (endoplasmic stress) experienced as a result of malignant transformation of the cell. To evade programmed cell removal, these cancer cells have been observed to upregulate extracellular CD47, an anti-phagocytic protein [Tsai and Discher, 2008]. RB1 promoter methylation during cancer progression could result the short-term down regulation of CALR levels due to co-inactivation of ncRNA-RB1. This would lend a time window to the cells to increase extracellular CD47 levels. If this is achieved previously to increases of ecto-CALR, cells are likely to avoid recognition by macrophages.

4.3 Comparison of non-coding transcription at regulatory elements

The next section of the discussion intends to compare bidirectional transcription initiation observed for ncRNA/PCG pairs to the divergent transcription initiation at regulatory elements which has been found to occur widespread across the genome. Also, the molecular basis for such correlated bidirectional transcription initiation as well as its functionality will be discussed.

4.3.1 Transcription initiation at promoters and enhancers

Transcription of divergent ncRNAs is not restricted to promoters, but has been detected at other regulatory elements such as enhancers and transcription termination sites [Kim et al., 2010],[De Santa et al., 2010],[Kapranov et al., 2007]. Although initially suggested, this non-coding transcription appears not to be the result of inappropriate Pol II transcription initiation due to low nucleosome occupancy at these loci. Instead, the presence of CPEs within both DNA strands that mediate the assembly of distinct PICs and promote divergent transcription initiation, argues for a regulated non-coding transcriptional output from these promoters and enhancers [Duttke et al., 2015],[Venters and Pugh, 2013],[Andersson et al., 2014a]. A similarity of divergent transcription initiation at regulatory elements, resulting in expression of two ncRNAs, one ncRNA and an mRNA or even two mRNAs, as in classical bidirectional promoters, is a short intervening sequence in between the paired TSSs. As described in the present thesis, transcription of long ncRNA/mRNA pairs annotated in the human genome is majorly initiated from a promoter size of 100-200 bp (Figure 3.1 B). In another study, the median distance between divergent TSSs of long ncRNAs and PCGs was found to be ~300 bp [Sigova et al., 2013]. Reports on bidirectional promoters expressing two PCGs pointed out that most of the paired TSSs were less than 300 bp apart [Adachi and Lieber, 2002]. Along these lines, the intervening sequence between the TSS of PCGs and any upstream initiated transcription was found to be between 110 and 200 bp, depending on the study [Duttke et al., 2015],[Core et al., 2014]. The same is true for bidirectional transcription at enhancer elements, that has been reported to initiate at the small distance of 110 bp [Core et al., 2014],[Andersson et al., 2014a].

Besides the presence of core promoter-like elements found upstream of those PCG promoters featuring divergent transcription initiation, assembly of two independent PICs has also been observed [Duttke et al., 2015],[Venters and Pugh, 2013]. This indicates that general transcription factors are actively positioned at both promoter sites. As most promoters that drive bidirectional transcription initiation are extremely short, the question arises how two distinct PICs can assemble within this restricted distance. At every TSS, the PIC contacts up to 50 bp of upstream and downstream sequence leaving little space in between both complexes as well as for the binding of additional activating transcription factors [Coulombe and Burton, 1999]. Different hypotheses can reconcile this observation: Expression of paired transcripts might not occur simultaneously at the same promoter. Instead, at any given time only a subset of cells within the population may initiate transcription in one or the other direction. On the other hand, it is possible that both PICs are assembled sequentially, with one complex forming and initiating Pol II upstream, followed by assembly of a downstream PIC and Pol II initiation. In support of this scenario, few general transcription factors stay associated with the TSS following recruitment and transcription initiation by Pol II [Yudkovsky et al., 2000]. This idea is

also supported by the fact that Pol II often pauses downstream of the TSS providing an additional regulatory switch past transcription initiation [Venters and Pugh, 2013].

4.3.2 Molecular basis for correlated bidirectional transcription initiation

Studies showed that bidirectional promoter activity can depend on the integrity of both TSSs and the activity of regulatory elements within the promoter region, such as transcription factor binding sites [Trinklein et al., 2004].
It is conceivable that binding of transcription factors within a small bidirectional promoter region can affect the assembly of paired PICs, explaining activation or repression of transcription in both directions. A study by Core et al. investigated the binding of transcription factors within DNA sequences featuring divergent transcription initiation [Core et al., 2014]. In divergent TSS regions of promoters and enhancers, different positional modes for transcription factor binding were observed: Central binders were found to have activating roles on PIC assembly at both TSSs and repressive factors tended to bind TSS-proximal, thereby preventing PIC assembly at the respective TSS and allowing anti-correlated expression of divergently expressed transcripts. The small promoter size has also been suggested to allow binding of only few factors at the same time, with several neighboring binding sites competing for transcription factor binding. This model is suitable to explain selection of multiple TSSs within one core promoter region, as binding of different transcription factors might influence initiation site selection by Pol II [Core et al., 2014].

4.3.3 Functionality of transcription at regulatory elements

Although numerous studies provide evidence for regulated transcription initiation upstream of PCG promoters and during enhancer activation, questions remain regarding the functionality of this transcriptional output. One possibility is that the act of transcription blocks negative influences on the promoter, such as the spread of repressive chromatin, allowing for steady PCG expression [Seila et al., 2009].
On the other hand, the act of transcription increases the chance for mutagenic alterations in the coding strand, thereby allowing for the acquisition of splice sites and consequently expression of longer transcripts. These transcripts can then eventually acquire functionality [Wu and Sharp, 2013]. Along these lines, bidirectional promoters have been described to be important for the emergence of species-specific transcripts, as novel transcripts have been found enriched upstream of PCGs [Gotea et al., 2013]. Additionally, promoter bidirectionality facilitates integration of transposable elements due to the open chromatin environment during the transcriptional processes. Thereby novel bidirectional gene pairs can originate, as domestication of transposable elements is important for the generation of new genes. Domesticated elements upstream of PCG promoters will conse-

quently share regulative capacity with the neighboring PCG [Kalitsis and Saffery, 2009]. These studies support the idea that the act of transcription promotes the acquisition of new genes. Thus, different transcript species could represent evolutionary stages of this process from an unspliced, early polyadenylated transcript to a spliced long non-coding RNA and eventually even to a functional PCG. In this regard, expression of bidirectional long ncRNAs could be low as they are in the process of acquiring functionality and as the cell has to control the output of deleterious transcripts.
This theory could also explain why bidirectionally expressed long ncRNAs are extensively spliced, as observed for long ncRNA-RB1 (Figure 3.10A). Not all splice forms might be functional, but when expressing various splice forms, a functional one could emerge in the cell.

4.4 Outlook

This study determines that divergent transcription initiation, commonly observed at regulatory elements in the cell, also occurs at bidirectional long ncRNA/PCG promoters. Also, a functional interplay between the bidirectionally expressed ncRNA-RB1 and RB1 genes is established by demonstrating that both genes exhibit tumor suppressive activity. Also, the long ncRNA-RB1 is found to have individual regulatory functions, such as towards the CALR gene, but also likely some regulatory potential shared with RB1. Nevertheless there are unresolved questions on the regulatory interaction of ncRNA-RB1 and RB1, resulting in an overlapping set of target genes, as well as on the regulatory mechanism of the CALR gene by ncRNA-RB1.
To determine an interaction between ncRNA-RB1 and pRB, that could result the tethering of RB1 to genomic loci targeted by the ncRNA and thereby a gene regulatory overlap, an RNA immunoprecipitation could be performed. By precipitating pRB using an antibody and isolation of associated RNAs this method could allow to map such an RNA-protein interaction in vivo. So far, an RNA-binding domain of the RB1 protein has not been identified. However, the protein also does not contain any commonly recognized DNA-binding or protein-interacting domain [Burkhart and Sage, 2008].
Additionally, methods such as Chirp (chromatin isolation by RNA purification), CHART (capture hybridization analysis of RNA targets) and RAP (RNA antisense purification) allow to pull down long ncRNA-associated proteins as well as associated chromatin by using biotinylated oligonucleotides [Chu et al., 2011],[Simon et al., 2011],[Engreitz et al., 2013]. These methods are very insightful, but generally require relatively high expression levels of the ncRNA to distinguish RNA specific signals from background.
In conclusion, employment of one of these methods could potentially reveal an interaction of ncRNA-RB1 with the CALR promoter or genomic locus, as well as the interplay of the ncRNA with transcriptional regulators and/or with pRB.

References

N. Adachi and M. R. Lieber. Bidirectional gene organization: a common architectural feature of the human genome. *Cell*, 109(7):807–809, Jun 2002.

K. Adelman and J. T. Lis. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet*, 13(10):720–731, Oct 2012. doi: 10.1038/nrg3293.

D. P. Aden, A. Fogel, S. Plotkin, I. Damjanov, and B. B. Knowles. Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line. *Nature*, 282(5739): 615–616, Dec 1979.

S. Adimoolam and J. M. Ford. p53 and regulation of DNA damage recognition during nucleotide excision repair. *DNA Repair (Amst)*, 2(9):947–954, Sep 2003.

A. Akalin, D. Fredman, E. Arner, X. Dong, J. C. Bryne, H. Suzuki, C. O. Daub, Y. Hayashizaki, and B. Lenhard. Transcriptional features of genomic regulatory blocks. *Genome Biol*, 10(4):R38, 2009. doi: 10.1186/gb-2009-10-4-r38.

S. Akira and T. Kishimoto. IL-6 and NF-IL6 in acute-phase response and viral infection. *Immunol Rev*, 127:25–50, Jun 1992.

B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell, Fourth Edition*. Garland Science, 2002. ISBN 0815332181.

A. E. Almada, X. Wu, A. J. Kriz, C. B. Burge, and P. A. Sharp. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature*, 499(7458):360–363, Jul 2013. doi: 10.1038/nature12349.

T. Amano, T. Sagai, H. Tanabe, Y. Mizushina, H. Nakazawa, and T. Shiroishi. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell*, 16(1):47–57, Jan 2009. doi: 10.1016/j.devcel.2008.11.011.

S. L. Ameres and P. D. Zamore. Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol*, 14(8):475–488, Aug 2013. doi: 10.1038/nrm3611.

R. Andersson, C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer,

D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jørgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Müller, F. A. N. T. O. M. C. , A. R. R. Forrest, P. Carninci, M. Rehli, and A. Sandelin. An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493):455–461, Mar 2014a. doi: 10.1038/nature12787.

R. Andersson, P. Refsing Andersen, E. Valen, L. J. Core, J. Bornholdt, M. Boyd, T. Heick Jensen, and A. Sandelin. Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun*, 5:5336, 2014b. doi: 10.1038/ncomms6336.

S. Andrews. FastQC - A quality control tool for high throughput sequence data. 2014. `http://www.bioinformatics.babraham.ac.uk/projects/fastqc`.

A. Armstrong and S. L. Eck. EpCAM: A new therapeutic target for an old cancer antigen. *Cancer Biol Ther*, 2(4):320–326, 2003.

J. Banerji, S. Rusconi, and W. Schaffner. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27(2 Pt 1):299–308, Dec 1981.

A. J. Bannister, P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire, and T. Kouzarides. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, 410(6824):120–124, Mar 2001. doi: 10.1038/35065138.

J. Bednar, R. A. Horowitz, S. A. Grigoryev, L. M. Carruthers, J. C. Hansen, A. J. Koster, and C. L. Woodcock. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci U S A*, 95(24):14173–14178, Nov 1998.

A. C. Bell, A. G. West, and G. Felsenfeld. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, 98(3):387–396, Aug 1999.

E. Bengal, O. Flores, A. Krauskopf, D. Reinberg, and Y. Aloni. Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II. *Mol Cell Biol*, 11(3):1195–1206, Mar 1991.

B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, and E. S. Lander. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2):315–326, Apr 2006. doi: 10.1016/j.cell.2006.02.041.

S. R. Bhaumik, E. Smith, and A. Shilatifard. Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol*, 14(11):1008–1016, Nov 2007. doi: 10.1038/nsmb1337.

U. K. Binné, M. K. Classon, F. A. Dick, W. Wei, M. Rape, W. G. Kaelin, Jr, A. M. Näär, and N. J. Dyson. Retinoblastoma protein and anaphase-promoting complex physically interact and functionally cooperate during cell-cycle exit. *Nat Cell Biol*, 9(2):225–232, Feb 2007. doi: 10.1038/ncb1532.

Broad Institute. Picard - A set of tools (in Java) for working with next generation sequencing data in the BAM format. 2014. `http://broadinstitute.github.io/picard`.

M. Bulger and M. Groudine. Functional and mechanistic diversity of distal transcription enhancers. *Cell*, 144(3):327–339, Feb 2011. doi: 10.1016/j.cell.2011.01.024.

D. L. Burkhart and J. Sage. Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nat Rev Cancer*, 8(9):671–682, Sep 2008. doi: 10.1038/nrc2399.

M. N. Cabili, C. Trapnell, L. Goff, M. Koziol, B. Tazon-Vega, A. Regev, and J. L. Rinn. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*, 25(18):1915–1927, Sep 2011. doi: 10.1101/gad.17446611.

P. Carninci, A. Sandelin, B. Lenhard, S. Katayama, K. Shimokawa, J. Ponjavic, C. A. M. Semple, M. S. Taylor, P. G. Engström, M. C. Frith, A. R. R. Forrest, W. B. Alkema, S. L. Tan, C. Plessy, R. Kodzius, T. Ravasi, T. Kasukawa, S. Fukuda, M. Kanamori-Katayama, Y. Kitazume, H. Kawaji, C. Kai, M. Nakamura, H. Konno, K. Nakano, S. Mottagui-Tabar, P. Arner, A. Chesi, S. Gustincich, F. Persichetti, H. Suzuki, S. M. Grimmond, C. A. Wells, V. Orlando, C. Wahlestedt, E. T. Liu, M. Harbers, J. Kawai, V. B. Bajic, D. A. Hume, and Y. Hayashizaki. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet*, 38(6):626–635, Jun 2006. doi: 10.1038/ng1789.

C. Carrieri, L. Cimatti, M. Biagioli, A. Beugnet, S. Zucchelli, S. Fedele, E. Pesce, I. Ferrer, L. Collavin, C. Santoro, A. R. R. Forrest, P. Carninci, S. Biffo, E. Stupka, and S. Gustincich. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*, 491(7424):454–457, Nov 2012. doi: 10.1038/nature11508.

T. R. Cech and J. A. Steitz. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*, 157(1):77–94, Mar 2014. doi: 10.1016/j.cell.2014.03.008.

B. Chandrasekar, P. C. Melby, H. M. Sarau, M. Raveendran, R. P. Perla, F. M. Marelli-Berg, N. O. Dulin, and I. S. Singh. Chemokine-cytokine cross-talk. The ELR+ CXC

chemokine LIX (CXCL5) amplifies a proinflammatory cytokine response via a phosphatidylinositol 3-kinase-NF-kappa B pathway. *J Biol Chem*, 278(7):4675–4686, Feb 2003. doi: 10.1074/jbc.M207006200.

M. P. Chao, S. Jaiswal, R. Weissman-Tsukamoto, A. A. Alizadeh, A. J. Gentles, J. Volkmer, K. Weiskopf, S. B. Willingham, T. Raveh, C. Y. Park, R. Majeti, and I. L. Weissman. Calreticulin is the dominant pro-phagocytic signal on multiple human cancers and is counterbalanced by CD47. *Sci Transl Med*, 2(63):63ra94, Dec 2010. doi: 10.1126/scitranslmed.3001375.

M. P. Chao, R. Majeti, and I. L. Weissman. Programmed cell removal: a new obstacle in the road to developing cancer. *Nat Rev Cancer*, 12(1):58–67, Jan 2012. doi: 10.1038/nrc3171.

S. W. Cheetham, F. Gruhl, J. S. Mattick, and M. E. Dinger. Long noncoding RNAs and the genetics of cancer. *Br J Cancer*, 108(12):2419–2425, Jun 2013. doi: 10.1038/bjc.2013.233.

B. Cheng, T. Li, P. B. Rahl, T. E. Adamson, N. B. Loudas, J. Guo, K. Varzavand, J. J. Cooper, X. Hu, A. Gnatt, R. A. Young, and D. H. Price. Functional association of Gdown1 with RNA polymerase II poised on human genes. *Mol Cell*, 45(1):38–50, Jan 2012. doi: 10.1016/j.molcel.2011.10.022.

A. C. M. Cheung and P. Cramer. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, 471(7337):249–253, Mar 2011. doi: 10.1038/nature09785.

A. C. M. Cheung and P. Cramer. A movie of RNA polymerase II transcription. *Cell*, 149(7):1431–1437, Jun 2012. doi: 10.1016/j.cell.2012.06.006.

M. Chinnam and D. W. Goodrich. RB1, development, and cancer. *Curr Top Dev Biol*, 94:129–169, 2011. doi: 10.1016/B978-0-12-380916-2.00005-X.

C. Chu, K. Qu, F. L. Zhong, S. E. Artandi, and H. Y. Chang. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*, 44(4):667–678, Nov 2011. doi: 10.1016/j.molcel.2011.08.027.

L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121):819–823, Feb 2013. doi: 10.1126/science.1231143.

J. Corden, B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon. Promoter sequences of eukaryotic protein-coding genes. *Science*, 209(4463):1406–1414, Sep 1980.

L. J. Core, J. J. Waterfall, and J. T. Lis. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 322(5909):1845–1848, Dec 2008. doi: 10.1126/science.1162228.

L. J. Core, J. J. Waterfall, D. A. Gilchrist, D. C. Fargo, H. Kwak, K. Adelman, and J. T. Lis. Defining the status of RNA polymerase at promoters. *Cell Rep*, 2(4):1025–1035, Oct 2012. doi: 10.1016/j.celrep.2012.08.034.

L. J. Core, A. L. Martins, C. G. Danko, C. T. Waters, A. Siepel, and J. T. Lis. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet*, 46(12):1311–1320, Dec 2014. doi: 10.1038/ng.3142.

B. Coulombe and Z. F. Burton. DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms. *Microbiol Mol Biol Rev*, 63(2):457–478, Jun 1999.

M. P. Creyghton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, and R. Jaenisch. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*, 107(50):21931–21936, Dec 2010. doi: 10.1073/pnas.1016071107.

F. Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, Aug 1970.

F. H. CRICK. On protein synthesis. *Symp Soc Exp Biol*, 12:138–163, 1958.

N. Czudnochowski, C. A. Bösken, and M. Geyer. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat Commun*, 3:842, 2012. doi: 10.1038/ncomms1846.

W. de Laat, P. Klous, J. Kooren, D. Noordermeer, R.-J. Palstra, M. Simonis, E. Splinter, and F. Grosveld. Three-dimensional organization of gene expression in erythroid cells. *Curr Top Dev Biol*, 82:117–139, 2008. doi: 10.1016/S0070-2153(07)00005-1.

F. De Santa, I. Barozzi, F. Mietton, S. Ghisletti, S. Polletti, B. K. Tusi, H. Muller, J. Ragoussis, C.-L. Wei, and G. Natoli. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol*, 8(5):e1000384, May 2010. doi: 10.1371/journal.pbio.1000384.

A. M. Deaton and A. Bird. CpG islands and the regulation of transcription. *Genes Dev*, 25(10):1010–1022, May 2011. doi: 10.1101/gad.2037511.

T. Derrien, R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D. G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan,

T. Lassmann, P. Carninci, J. B. Brown, L. Lipovich, J. M. Gonzalez, M. Thomas, C. A. Davis, R. Shiekhattar, T. R. Gingeras, T. J. Hubbard, C. Notredame, J. Harrow, and R. Guigó. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*, 22(9):1775–1789, Sep 2012. doi: 10.1101/gr.132159.111.

S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L.-H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, and T. R. Gingeras. Landscape of transcription in human cells. *Nature*, 489(7414):101–108, Sep 2012. doi: 10.1038/nature11233.

S. H. C. Duttke, S. A. Lacadie, M. M. Ibrahim, C. K. Glass, D. L. Corcoran, C. Benner, S. Heinz, J. T. Kadonaga, and U. Ohler. Human promoters are intrinsically directional. *Mol Cell*, 57(4):674–684, Feb 2015. doi: 10.1016/j.molcel.2014.12.029.

B. D. Dynlacht, O. Flores, J. A. Lees, and E. Harlow. Differential regulation of E2F transactivation by cyclin/cdk2 complexes. *Genes Dev*, 8(15):1772–1786, Aug 1994.

S. C. Elgin. The formation and function of DNase I hypersensitive sites in the process of gene activation. *J Biol Chem*, 263(36):19259–19262, Dec 1988.

R. Elkon, A. P. Ugalde, and R. Agami. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet*, 14(7):496–506, Jul 2013. doi: 10.1038/nrg3482.

E.N.C.O.D.E. Project Consortium, E. Birney, J. A. Stamatoyannopoulos, A. Dutta, R. Guigó, T. R. Gingeras, E. H. Margulies, Z. Weng, M. Snyder, E. T. Dermitzakis, R. E. Thurman, M. S. Kuehn, C. M. Taylor, S. Neph, C. M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J. A. Greenbaum, R. M. Andrews, P. Flicek, P. J. Boyle, H. Cao, N. P. Carter, G. K. Clelland, S. Davis, N. Day, P. Dhami, S. C. Dillon, M. O. Dorschner, H. Fiegler, P. G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K. D. James, B. E. Johnson, E. M. Johnson, T. T. Frum, E. R. Rosenzweig, N. Karnani,

K. Lee, G. C. Lefebvre, P. A. Navas, F. Neri, S. C. J. Parker, P. J. Sabo, R. Sandstrom, A. Shafer, D. Vetrie, M. Weaver, S. Wilcox, M. Yu, F. S. Collins, J. Dekker, J. D. Lieb, T. D. Tullius, G. E. Crawford, S. Sunyaev, W. S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A. Sandelin, I. L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H. A. Hirsch, E. A. Sekinger, J. Lagarde, J. F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermüller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korbel, O. Emanuelsson, J. S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M. C. Dickson, D. J. Thomas, M. T. Weirauch, J. Gilbert, J. Drenkow, I. Bell, X. Zhao, K. G. Srinivasan, W.-K. Sung, H. S. Ooi, K. P. Chiu, S. Foissac, T. Alioto, M. Brent, L. Pachter, M. L. Tress, A. Valencia, S. W. Choo, C. Y. Choo, C. Ucla, C. Manzano, C. Wyss, E. Cheung, T. G. Clark, J. B. Brown, M. Ganesh, S. Patel, H. Tammana, J. Chrast, C. N. Henrichsen, C. Kai, J. Kawai, U. Nagalakshmi, J. Wu, Z. Lian, J. Lian, P. Newburger, X. Zhang, P. Bickel, J. S. Mattick, P. Carninci, Y. Hayashizaki, S. Weissman, T. Hubbard, R. M. Myers, J. Rogers, P. F. Stadler, T. M. Lowe, C.-L. Wei, Y. Ruan, K. Struhl, M. Gerstein, S. E. Antonarakis, Y. Fu, E. D. Green, U. Karaöz, A. Siepel, J. Taylor, L. A. Liefer, K. A. Wetterstrand, P. J. Good, E. A. Feingold, M. S. Guyer, G. M. Cooper, G. Asimenos, C. N. Dewey, M. Hou, S. Nikolaev, J. I. Montoya-Burgos, A. Löytynoja, S. Whelan, F. Pardi, T. Massingham, H. Huang, N. R. Zhang, I. Holmes, J. C. Mullikin, A. Ureta-Vidal, B. Paten, M. Seringhaus, D. Church, K. Rosenbloom, W. J. Kent, E. A. Stone, N. I. S. C. C. S. P. , B. C. o. M. H. G. S. C. , W. U. G. S. C. , B. I. , C. H. O. R. I. , S. Batzoglou, N. Goldman, R. C. Hardison, D. Haussler, W. Miller, A. Sidow, N. D. Trinklein, Z. D. Zhang, L. Barrera, R. Stuart, D. C. King, A. Ameur, S. Enroth, M. C. Bieda, J. Kim, A. A. Bhinge, N. Jiang, J. Liu, F. Yao, V. B. Vega, C. W. H. Lee, P. Ng, A. Shahab, A. Yang, Z. Moqtaderi, Z. Zhu, X. Xu, S. Squazzo, M. J. Oberley, D. Inman, M. A. Singer, T. A. Richmond, K. J. Munn, A. Rada-Iglesias, O. Wallerman, J. Komorowski, J. C. Fowler, P. Couttet, A. W. Bruce, O. M. Dovey, P. D. Ellis, C. F. Langford, D. A. Nix, G. Euskirchen, S. Hartman, A. E. Urban, P. Kraus, S. Van Calcar, N. Heintzman, T. H. Kim, K. Wang, C. Qu, G. Hon, R. Luna, C. K. Glass, M. G. Rosenfeld, S. F. Aldred, S. J. Cooper, A. Halees, J. M. Lin, H. P. Shulha, X. Zhang, M. Xu, J. N. S. Haidar, Y. Yu, Y. Ruan, V. R. Iyer, R. D. Green, C. Wadelius, P. J. Farnham, B. Ren, R. A. Harte, A. S. Hinrichs, H. Trumbower, H. Clawson, J. Hillman-Jackson, A. S. Zweig, K. Smith, A. Thakkapallayil, G. Barber, R. M. Kuhn, D. Karolchik, L. Armengol, C. P. Bird, P. I. W. de Bakker, A. D. Kern, N. Lopez-Bigas, J. D. Martin, B. E. Stranger, A. Woodroffe, E. Davydov, A. Dimas, E. Eyras, I. B. Hallgrímsdóttir, J. Huppert, M. C. Zody, G. R. Abecasis, X. Estivill, G. G. Bouffard, X. Guan, N. F. Hansen, J. R. Idol, V. V. B. Maduro, B. Maskeri, J. C. McDowell, M. Park, P. J. Thomas, A. C. Young, R. W. Blakesley, D. M. Muzny, E. Sodergren, D. A. Wheeler, K. C. Worley, H. Jiang, G. M. Weinstock, R. A. Gibbs, T. Graves, R. Fulton,

E. R. Mardis, R. K. Wilson, M. Clamp, J. Cuff, S. Gnerre, D. B. Jaffe, J. L. Chang, K. Lindblad-Toh, E. S. Lander, M. Koriabine, M. Nefedov, K. Osoegawa, Y. Yoshinaga, B. Zhu, and P. J. de Jong. Identification and analysis of functional elements in 1human genome by the ENCODE pilot project. *Nature*, 447(7146):799–816, Jun 2007. doi: 10.1038/nature05874.

J. M. Engreitz, A. Pandya-Jones, P. McDonel, A. Shishkin, K. Sirokman, C. Surka, S. Kadri, J. Xing, A. Goren, E. S. Lander, K. Plath, and M. Guttman. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*, 341(6147):1237973, Aug 2013. doi: 10.1126/science.1237973.

J. Ernst and M. Kellis. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol*, 28(8):817–825, Aug 2010. doi: 10.1038/nbt.1662.

C. Esnault, Y. Ghavi-Helm, S. Brun, J. Soutourina, N. Van Berkum, C. Boschiero, F. Holstege, and M. Werner. Mediator-dependent recruitment of TFIIH modules in preinitiation complex. *Mol Cell*, 31(3):337–346, Aug 2008. doi: 10.1016/j.molcel.2008.06.021.

R. Evans, J. A. Fairley, and S. G. Roberts. Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. *Genes Dev*, 15(22):2945–2949, Nov 2001. doi: 10.1101/gad.206901.

J. T. Finch, L. C. Lutter, D. Rhodes, R. S. Brown, B. Rushton, M. Levitt, and A. Klug. Structure of nucleosome core particles of chromatin. *Nature*, 269(5623):29–36, Sep 1977.

R. A. Flynn, A. E. Almada, J. R. Zamudio, and P. A. Sharp. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc Natl Acad Sci U S A*, 108(26):10460–10465, Jun 2011. doi: 10.1073/pnas.1106630108.

S. H. Friend, R. Bernards, S. Rogelj, R. A. Weinberg, J. M. Rapaport, D. M. Albert, and T. P. Dryja. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature*, 323(6089):643–646, 1986. doi: 10.1038/323643a0.

D. A. Fruman and C. Rommel. PI3K and cancer: lessons, challenges and opportunities. *Nat Rev Drug Discov*, 13(2):140–156, Feb 2014. doi: 10.1038/nrd4204.

R. M. Gill, P. A. Hamel, J. Zhe, E. Zacksenhaus, B. L. Gallie, and R. A. Phillips. Characterization of the human RB1 promoter and of elements involved in transcriptional regulation. *Cell Growth Differ*, 5(5):467–474, May 1994.

L. I. Gold, P. Eggleton, M. T. Sweetwyne, L. B. Van Duyn, M. R. Greives, S.-M. Naylor, M. Michalak, and J. E. Murphy-Ullrich. Calreticulin: non-endoplasmic reticulum functions in physiology and disease. *FASEB J*, 24(3):665–683, Mar 2010. doi: 10.1096/fj.09-145482.

V. Gotea, H. M. Petrykowska, and L. Elnitski. Bidirectional promoters as important drivers for the emergence of species-specific transcripts. *PLoS One*, 8(2):e57323, 2013. doi: 10.1371/journal.pone.0057323.

F. L. Graham, J. Smiley, W. C. Russell, and R. Nairn. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol*, 36(1):59–74, Jul 1977.

R. A. Green, R. Wollman, and K. B. Kaplan. APC and EB1 function together in mitosis to regulate spindle dynamics and chromosome alignment. *Mol Biol Cell*, 16(10):4609–4622, Oct 2005. doi: 10.1091/mbc.E05-03-0259.

P. Grote and B. G. Herrmann. The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. *RNA Biol*, 10 (10):1579–1585, Oct 2013. doi: 10.4161/rna.26165.

P. Grote, L. Wittler, D. Hendrix, F. Koch, S. Währisch, A. Beisaw, K. Macura, G. Bläss, M. Kellis, M. Werber, and B. G. Herrmann. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell*, 24(2): 206–214, Jan 2013. doi: 10.1016/j.devcel.2012.12.012.

S. Grünberg and S. Hahn. Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem Sci*, 38(12):603–611, Dec 2013. doi: 10.1016/j.tibs.2013.09.002.

M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, and R. A. Young. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, 130(1): 77–88, Jul 2007. doi: 10.1016/j.cell.2007.05.042.

J. Guo and D. H. Price. RNA polymerase II transcription elongation control. *Chem Rev*, 113(11):8583–8603, Nov 2013. doi: 10.1021/cr400105n.

M. Guttman, I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn, and E. S. Lander. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, 458(7235):223–227, Mar 2009. doi: 10.1038/nature07672.

M. Guttman, J. Donaghey, B. W. Carey, M. Garber, J. K. Grenier, G. Munson, G. Young, A. B. Lucas, R. Ach, L. Bruhn, X. Yang, I. Amit, A. Meissner, A. Regev, J. L. Rinn,

D. E. Root, and E. S. Lander. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, 477(7364):295–300, Sep 2011. doi: 10.1038/nature10398.

V. Haberle, N. Li, Y. Hadzhiev, C. Plessy, C. Previti, C. Nepal, J. Gehrig, X. Dong, A. Akalin, A. M. Suzuki, W. F. J. van IJcken, O. Armant, M. Ferg, U. Strähle, P. Carninci, F. Müller, and B. Lenhard. Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature*, 507(7492):381–385, Mar 2014. doi: 10.1038/nature12974.

P. A. Hamel, R. M. Gill, R. A. Phillips, and B. L. Gallie. Transcriptional repression of the E2-containing promoters EIIaE, c-myc, and RB1 by the product of the RB1 gene. *Mol Cell Biol*, 12(8):3431–3438, Aug 1992.

M. B. Harris, J. Mostecki, and P. B. Rothman. Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. *J Biol Chem*, 280(13):13114–13121, Apr 2005. doi: 10.1074/jbc.M412649200.

S. Haupt, M. Berger, Z. Goldberg, and Y. Haupt. Apoptosis - the p53 network. *J Cell Sci*, 116(Pt 20):4077–4085, Oct 2003. doi: 10.1242/jcs.00739.

N. D. Heintzman, G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis, and B. Ren. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459(7243):108–112, May 2009. doi: 10.1038/nature07829.

D. Helbling, B. U. Mueller, N. A. Timchenko, J. Schardt, M. Eyer, D. R. Betts, M. Jotterand, S. Meyer-Monard, M. F. Fey, and T. Pabst. CBFB-SMMHC is correlated with increased calreticulin expression and suppresses the granulocytic differentiation factor CEBPA in AML with inv(16). *Blood*, 106(4):1369–1375, Aug 2005. doi: 10.1182/blood-2004-11-4392.

K. Helin, E. Harlow, and A. Fattaey. Inhibition of E2F-1 transactivation by direct binding of the retinoblastoma protein. *Mol Cell Biol*, 13(10):6501–6508, Oct 1993.

B. Hoesel and J. A. Schmid. The complexity of NF-κB signaling in inflammation and cancer. *Mol Cancer*, 12:86, 2013. doi: 10.1186/1476-4598-12-86.

T. K. Howcroft, A. Raval, J. D. Weissman, A. Gegonne, and D. S. Singer. Distinct transcriptional pathways regulate basal and activated major histocompatibility complex class I expression. *Mol Cell Biol*, 23(10):3377–3391, May 2003.

D. W. Huang, B. T. Sherman, and R. A. Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1):44–57, 2009a. doi: 10.1038/nprot.2008.211.

D. W. Huang, B. T. Sherman, and R. A. Lempicki. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37(1):1–13, Jan 2009b. doi: 10.1093/nar/gkn923.

M. Huarte, M. Guttman, D. Feldser, M. Garber, M. J. Koziol, D. Kenzelmann-Broz, A. M. Khalil, O. Zuk, I. Amit, M. Rabani, L. D. Attardi, A. Regev, E. S. Lander, T. Jacks, and J. L. Rinn. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*, 142(3):409–419, Aug 2010. doi: 10.1016/j.cell.2010.06.040.

T. Hung, Y. Wang, M. F. Lin, A. K. Koegel, Y. Kotake, G. D. Grant, H. M. Horlings, N. Shah, C. Umbricht, P. Wang, Y. Wang, B. Kong, A. Langerød, A.-L. Børresen-Dale, S. K. Kim, M. van de Vijver, S. Sukumar, M. L. Whitfield, M. Kellis, Y. Xiong, D. J. Wong, and H. Y. Chang. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*, 43(7):621–629, Jul 2011. doi: 10.1038/ng.848.

M. K. Iyer, Y. S. Niknafs, R. Malik, U. Singhal, A. Sahu, Y. Hosono, T. R. Barrette, J. R. Prensner, J. R. Evans, S. Zhao, A. Poliakov, X. Cao, S. M. Dhanasekaran, Y.-M. Wu, D. R. Robinson, D. G. Beer, F. Y. Feng, H. K. Iyer, and A. M. Chinnaiyan. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet*, 47(3): 199–208, Mar 2015. doi: 10.1038/ng.3192.

F. JACOB and J. MONOD. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, 3:318–356, Jun 1961.

R. Javahery, A. Khachi, K. Lo, B. Zenzie-Gregory, and S. T. Smale. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol*, 14(1):116–127, Jan 1994.

T. H. Jensen, A. Jacquier, and D. Libri. Dealing with pervasive transcription. *Mol Cell*, 52(4):473–484, Nov 2013. doi: 10.1016/j.molcel.2013.10.032.

A. Johnson and J. B. Alberts. *Molecular Biology of the Cell*. Garland Science, 2002. ISBN 0849371619.

P. Johnsson, L. Lipovich, D. Grandér, and K. V. Morris. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta*, 1840(3): 1063–1071, Mar 2014. doi: 10.1016/j.bbagen.2013.10.035.

T. Juven-Gershon and J. T. Kadonaga. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol*, 339(2):225–229, Mar 2010. doi: 10.1016/j.ydbio.2009.08.009.

J. T. Kadonaga. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip Rev Dev Biol*, 1(1):40–51, 2012. doi: 10.1002/wdev.21.

M. H. Kagey, J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando, N. L. van Berkum, C. C. Ebmeier, J. Goossens, P. B. Rahl, S. S. Levine, D. J. Taatjes, J. Dekker, and R. A. Young. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314):430–435, Sep 2010. doi: 10.1038/nature09380.

S. Kageyama, T. Isono, H. Iwaki, Y. Wakabayashi, Y. Okada, K. Kontani, K. Yoshimura, A. Terai, Y. Arai, and T. Yoshiki. Identification by proteomic analysis of calreticulin as a marker for bladder cancer and evaluation of the diagnostic accuracy of its detection in urine. *Clin Chem*, 50(5):857–866, May 2004. doi: 10.1373/clinchem.2003.027425.

P. Kalitsis and R. Saffery. Inherent promoter bidirectionality facilitates maintenance of sequence integrity and transcription of parasitic DNA in mammalian genomes. *BMC Genomics*, 10:498, 2009. doi: 10.1186/1471-2164-10-498.

E. Kandiah, S. Trowitzsch, K. Gupta, M. Haffke, and I. Berger. More pieces to the puzzle: recent structural insights into class II transcription initiation. *Curr Opin Struct Biol*, 24:91–97, Feb 2014. doi: 10.1016/j.sbi.2013.12.005.

P. Kapranov, J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermüller, I. L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganesh, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammana, and T. R. Gingeras. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, 316(5830):1484–1488, Jun 2007. doi: 10.1126/science.1138341.

J. Kaufmann and S. T. Smale. Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev*, 8(7):821–829, Apr 1994.

D. Kelley and J. Rinn. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol*, 13(11):R107, 2012. doi: 10.1186/gb-2012-13-11-r107.

W. J. Kent. BLAT–the BLAST-like alignment tool. *Genome Res*, 12(4):656–664, Apr 2002. doi: 10.1101/gr.229202.ArticlepublishedonlinebeforeMarch2002.

W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and D. Haussler. The human genome browser at UCSC. *Genome Res*, 12(6):996–1006, Jun 2002. doi: 10.1101/gr.229102.ArticlepublishedonlinebeforeprintinMay2002.

D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, 14(4):R36, 2013. doi: 10.1186/gb-2013-14-4-r36.

T.-K. Kim, M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman, and M. E. Greenberg. Widespread transcription

at neuronal activity-regulated enhancers. *Nature*, 465(7295):182–187, May 2010. doi: 10.1038/nature09033.

T. Klampfl, H. Gisslinger, A. S. Harutyunyan, H. Nivarthi, E. Rumi, J. D. Milosevic, N. C. C. Them, T. Berg, B. Gisslinger, D. Pietra, D. Chen, G. I. Vladimer, K. Bagienski, C. Milanesi, I. C. Casetti, E. Sant'Antonio, V. Ferretti, C. Elena, F. Schischlik, C. Cleary, M. Six, M. Schalling, A. Schönegger, C. Bock, L. Malcovati, C. Pascutto, G. Superti-Furga, M. Cazzola, and R. Kralovics. Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med*, 369(25):2379–2390, Dec 2013. doi: 10.1056/NEJMoa1311347.

A. Klug, D. Rhodes, J. Smith, J. T. Finch, and J. O. Thomas. A low resolution structure for the histone core of the nucleosome. *Nature*, 287(5782):509–516, Oct 1980.

F. Koch, R. Fenouil, M. Gut, P. Cauchy, T. K. Albert, J. Zacarias-Cabeza, S. Spicuglia, A. L. de la Chapelle, M. Heidemann, C. Hintermair, D. Eick, I. Gut, P. Ferrier, and J.-C. Andrau. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol*, 18(8):956–963, Aug 2011. doi: 10.1038/nsmb.2085.

J. Krieg and T. Hunter. Identification of the two major epidermal growth factor-induced tyrosine phosphorylation sites in the microvillar core protein ezrin. *J Biol Chem*, 267 (27):19258–19265, Sep 1992.

H. Kwak and J. T. Lis. Control of transcriptional elongation. *Annu Rev Genet*, 47: 483–508, 2013. doi: 10.1146/annurev-genet-110711-155440.

M. Lacroix, S. El Messaoudi, G. Rodier, A. Le Cam, C. Sardet, and E. Fabbrizio. The histone-binding protein COPR5 is required for nuclear functions of the protein arginine methyltransferase PRMT5. *EMBO Rep*, 9(5):452–458, May 2008. doi: 10.1038/embor.2008.45.

T. Lagrange, A. N. Kapanidis, H. Tang, D. Reinberg, and R. H. Ebright. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev*, 12(1):34–44, Jan 1998.

F. Lai, U. A. Orom, M. Cesaroni, M. Beringer, D. J. Taatjes, G. A. Blobel, and R. Shiekhattar. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature*, 494(7438):497–501, Feb 2013. doi: 10.1038/nature11884.

M. T. Y. Lam, H. Cho, H. P. Lesch, D. Gosselin, S. Heinz, Y. Tanaka-Oishi, C. Benner, M. U. Kaikkonen, A. S. Kim, M. Kosaka, C. Y. Lee, A. Watt, T. R. Grossman, M. G. Rosenfeld, R. M. Evans, and C. K. Glass. Rev-Erbs repress macrophage gene expression

by inhibiting enhancer-directed transcription. *Nature*, 498(7455):511–515, Jun 2013. doi: 10.1038/nature12209.

E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh,

F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowki, and I. H. G. S. C. . Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, Feb 2001.

B. Langmead and S. L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4):357–359, Apr 2012. doi: 10.1038/nmeth.1923.

P. A. Latos, F. M. Pauler, M. V. Koerner, H. B. Şenergin, Q. J. Hudson, R. R. Stocsits, W. Allhoff, S. H. Stricker, R. M. Klement, K. E. Warczok, K. Aumayr, P. Pasierbek, and D. P. Barlow. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science*, 338(6113):1469–1472, Dec 2012. doi: 10.1126/science.1228110.

M. P. Lee, K. Howcroft, A. Kotekar, H. H. Yang, K. H. Buetow, and D. S. Singer. ATG deserts define a novel core promoter subclass. *Genome Res*, 15(9):1189–1197, Sep 2005. doi: 10.1101/gr.3873705.

T. I. Lee and R. A. Young. Transcriptional regulation and its misregulation in disease. *Cell*, 152(6):1237–1251, Mar 2013. doi: 10.1016/j.cell.2013.02.014.

W. H. Lee, J. Y. Shew, F. D. Hong, T. W. Sery, L. A. Donoso, L. J. Young, R. Bookstein, and E. Y. Lee. The retinoblastoma susceptibility gene encodes a nuclear phosphoprotein associated with DNA binding activity. *Nature*, 329(6140):642–645, 1987. doi: 10.1038/329642a0.

P. A. Leighton, R. S. Ingram, J. Eggenschwiler, A. Efstratiadis, and S. M. Tilghman. Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature*, 375(6526):34–39, May 1995. doi: 10.1038/375034a0.

B. Lenhard, A. Sandelin, and P. Carninci. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet*, 13(4):233–245, Apr 2012. doi: 10.1038/nrg3163.

L. A. Lettice, S. J. H. Heaney, L. A. Purdie, L. Li, P. de Beer, B. A. Oostra, D. Goode, G. Elgar, R. E. Hill, and E. de Graaff. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet*, 12(14):1725–1735, Jul 2003.

B. A. Lewis, T. K. Kim, and S. H. Orkin. A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. *Proc Natl Acad Sci U S A*, 97(13):7172–7177, Jun 2000. doi: 10.1073/pnas.120181197.

W. Li, D. Notani, Q. Ma, B. Tanasa, E. Nunez, A. Y. Chen, D. Merkurjev, J. Zhang, K. Ohgi, X. Song, S. Oh, H.-S. Kim, C. K. Glass, and M. G. Rosenfeld. Functional

roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, 498(7455):516–520, Jun 2013. doi: 10.1038/nature12210.

M. Lieber, B. Smith, A. Szakal, W. Nelson-Rees, and G. Todaro. A continuous tumor-cell line from a human lung carcinoma with properties of type II alveolar epithelial cells. *Int J Cancer*, 17(1):62–70, Jan 1976.

J. M. Lin, P. J. Collins, N. D. Trinklein, Y. Fu, H. Xi, R. M. Myers, and Z. Weng. Transcription factor binding and modified histones in human bidirectional promoters. *Genome Res*, 17(6):818–827, Jun 2007. doi: 10.1101/gr.5623407.

M. Lippman, G. Bolan, and K. Huff. Interactions of antiestrogens with human breast cancer in long-term tissue culture. *Cancer Treat Rep*, 60(10):1421–1429, Oct 1976.

J. Liu and E. S. Maxwell. Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene. *Nucleic Acids Res*, 18(22):6565–6571, Nov 1990.

G. Livide, M. C. Epistolato, M. Amenduni, V. Disciglio, A. Marozza, M. A. Mencarelli, P. Toti, S. Lazzi, T. Hadjistilianou, S. De Francesco, A. D'Ambrosio, A. Renieri, and F. Ariani. Epigenetic and copy number variation analysis in retinoblastoma by MS-MLPA. *Pathol Oncol Res*, 18(3):703–712, Jul 2012. doi: 10.1007/s12253-012-9498-8.

D. López-Terrada, S. W. Cheung, M. J. Finegold, and B. B. Knowles. Hep G2 is a hepatoblastoma-derived cell line. *Hum Pathol*, 40(10):1512–1515, Oct 2009. doi: 10.1016/j.humpath.2009.07.003.

N. Louis, C. Evelegh, and F. L. Graham. Cloning and sequencing of the cellular-viral junctions from the human adenovirus type 5 transformed 293 cell line. *Virology*, 233 (2):423–429, Jul 1997. doi: 10.1006/viro.1997.8597.

B. B. Lozzio and C. B. Lozzio. Properties and usefulness of the original K-562 human myelogenous leukemia cell line. *Leuk Res*, 3(6):363–370, 1979.

S. J. Lubbe, A. M. Pittman, B. Olver, A. Lloyd, J. Vijayakrishnan, S. Naranjo, S. Dobbins, P. Broderick, J. L. Gómez-Skarmeta, and R. S. Houlston. The 14q22.2 colorectal cancer variant rs4444235 shows cis-acting regulation of BMP4. *Oncogene*, 31(33):3777–3784, Aug 2012. doi: 10.1038/onc.2011.564.

J. Luo. Glycogen synthase kinase 3beta (GSK3beta) in tumorigenesis and cancer chemotherapy. *Cancer Lett*, 273(2):194–200, Jan 2009. doi: 10.1016/j.canlet.2008.05.045.

M. J. Luteijn and R. F. Ketting. PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nat Rev Genet*, 14(8):523–534, Aug 2013. doi: 10.1038/nrg3495.

Z.-M. Lwin, C. Guo, A. Salim, G. W.-C. Yip, F.-T. Chew, J. Nan, A. A. Thike, P.-H. Tan, and B.-H. Bay. Clinicopathological significance of calreticulin in breast invasive ductal carcinoma. *Mod Pathol*, 23(12):1559–1566, Dec 2010. doi: 10.1038/modpathol.2010.173.

M. Macaluso, M. Montanari, and A. Giordano. Rb family proteins as modulators of gene expression and new aspects regarding the interaction with chromatin remodeling enzymes. *Oncogene*, 25(38):5263–5267, Aug 2006. doi: 10.1038/sj.onc.1209680.

K. F. Macleod, Y. Hu, and T. Jacks. Loss of Rb activates both p53-dependent and independent cell death pathways in the developing mouse nervous system. *EMBO J*, 15(22):6178–6188, Nov 1996.

A. L. Manning and N. J. Dyson. pRB, a tumor suppressor with a stabilizing presence. *Trends Cell Biol*, 21(8):433–441, Aug 2011. doi: 10.1016/j.tcb.2011.05.003.

S. Marquardt, R. Escalante-Chong, N. Pho, J. Wang, L. S. Churchman, M. Springer, and S. Buratowski. A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell*, 157(7):1712–1723, Jun 2014. doi: 10.1016/j.cell.2014.04.036.

A. Marsico, M. R. Huska, J. Lasserre, H. Hu, D. Vucicevic, A. Musahl, U. Orom, and M. Vingron. PROmiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. *Genome Biol*, 14(8):R84, 2013. doi: 10.1186/gb-2013-14-8-r84.

Massachusetts Institute of Technology. CRISPR - The CRISPR design tool. 2014. `http://crispr.mit.edu`.

G. A. Maston, S. K. Evans, and M. R. Green. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet*, 7:29–59, 2006. doi: 10.1146/annurev.genom.7.080505.115623.

N. Mavaddat, S. Peock, D. Frost, S. Ellis, R. Platte, E. Fineberg, D. G. Evans, L. Izatt, R. A. Eeles, J. Adlard, R. Davidson, D. Eccles, T. Cole, J. Cook, C. Brewer, M. Tischkowitz, F. Douglas, S. Hodgson, L. Walker, M. E. Porteous, P. J. Morrison, L. E. Side, M. J. Kennedy, C. Houghton, A. Donaldson, M. T. Rogers, H. Dorkins, Z. Miedzybrodzka, H. Gregory, J. Eason, J. Barwell, E. McCann, A. Murray, A. C. Antoniou, D. F. Easton, and E. M. B. R. A. C. E. . Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. *J Natl Cancer Inst*, 105(11):812–822, Jun 2013. doi: 10.1093/jnci/djt095.

S. Memczak, M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S. D. Mackowiak, L. H. Gregersen, M. Munschauer, A. Loewer, U. Ziebold, M. Landthaler, C. Kocks, F. le Noble, and N. Rajewsky. Circular RNAs are a large class of animal

RNAs with regulatory potency. *Nature*, 495(7441):333–338, Mar 2013. doi: 10.1038/nature11928.

M. Michalak, J. Groenendyk, E. Szabo, L. I. Gold, and M. Opas. Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. *Biochem J*, 417(3): 651–666, Feb 2009. doi: 10.1042/BJ20081847.

K. V. Morris and J. S. Mattick. The rise of regulatory RNA. *Nat Rev Genet*, 15(6): 423–437, Jun 2014. doi: 10.1038/nrg3722.

A.-S. Musahl, X. Huang, S. Rusakiewicz, E. Ntini, A. Marsico, G. Kroemer, O. Kepp, and U. A. Ørom. A long non-coding RNA links calreticulin-mediated immunogenic cell removal to RB1 transcription. *Oncogene*, Jan 2015. doi: 10.1038/onc.2014.424.

G. W. Muse, D. A. Gilchrist, S. Nechaev, R. Shah, J. S. Parker, S. F. Grissom, J. Zeitlinger, and K. Adelman. RNA polymerase is poised for activation across the genome. *Nat Genet*, 39(12):1507–1511, Dec 2007. doi: 10.1038/ng.2007.21.

J. Nangalia, C. E. Massie, E. J. Baxter, F. L. Nice, G. Gundem, D. C. Wedge, E. Avezov, J. Li, K. Kollmann, D. G. Kent, A. Aziz, A. L. Godfrey, J. Hinton, I. Martincorena, P. Van Loo, A. V. Jones, P. Guglielmelli, P. Tarpey, H. P. Harding, J. D. Fitzpatrick, C. T. Goudie, C. A. Ortmann, S. J. Loughran, K. Raine, D. R. Jones, A. P. Butler, J. W. Teague, S. O'Meara, S. McLaren, M. Bianchi, Y. Silber, D. Dimitropoulou, D. Bloxham, L. Mudie, M. Maddison, B. Robinson, C. Keohane, C. Maclean, K. Hill, K. Orchard, S. Tauro, M.-Q. Du, M. Greaves, D. Bowen, B. J. P. Huntly, C. N. Harrison, N. C. P. Cross, D. Ron, A. M. Vannucchi, E. Papaemmanuil, P. J. Campbell, and A. R. Green. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med*, 369(25):2391–2405, Dec 2013. doi: 10.1056/NEJMoa1312542.

A. Necsulea, M. Soumillon, M. Warnefors, A. Liechti, T. Daish, U. Zeller, J. C. Baker, F. Grützner, and H. Kaessmann. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, 505(7485):635–640, Jan 2014. doi: 10.1038/nature12943.

H. Neil, C. Malabat, Y. d'Aubenton Carafa, Z. Xu, L. M. Steinmetz, and A. Jacquier. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, 457(7232):1038–1042, Feb 2009. doi: 10.1038/nature07747.

W. W. Nichols, D. G. Murphy, V. J. Cristofalo, L. H. Toji, A. E. Greene, and S. A. Dwight. Characterization of a new human diploid cell strain, IMR-90. *Science*, 196 (4285):60–63, Apr 1977.

D. B. Nikolov, H. Chen, E. D. Halay, A. A. Usheva, K. Hisatake, D. K. Lee, R. G. Roeder, and S. K. Burley. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*, 377(6545):119–128, Sep 1995. doi: 10.1038/377119a0.

E. Ntini, A. I. Järvelin, J. Bornholdt, Y. Chen, M. Boyd, M. Jørgensen, R. Andersson, I. Hoof, A. Schein, P. R. Andersen, P. K. Andersen, P. Preker, E. Valen, X. Zhao, V. Pelechano, L. M. Steinmetz, A. Sandelin, and T. H. Jensen. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol*, 20(8):923–928, Aug 2013. doi: 10.1038/nsmb.2640.

M. Obeid, A. Tesniere, F. Ghiringhelli, G. M. Fimia, L. Apetoh, J.-L. Perfettini, M. Castedo, G. Mignot, T. Panaretakis, N. Casares, D. Métivier, N. Larochette, P. van Endert, F. Ciccosanti, M. Piacentini, L. Zitvogel, and G. Kroemer. Calreticulin exposure dictates the immunogenicity of cancer cell death. *Nat Med*, 13(1):54–61, Jan 2007. doi: 10.1038/nm1523.

N. Ohtani-Fujita, T. Fujita, A. Aoike, N. E. Osifchin, P. D. Robbins, and T. Sakai. CpG methylation inactivates the promoter activity of the human retinoblastoma tumor-suppressor gene. *Oncogene*, 8(4):1063–1067, Apr 1993.

U. A. Ørom, T. Derrien, M. Beringer, K. Gumireddy, A. Gardini, G. Bussotti, F. Lai, M. Zytnicki, C. Notredame, Q. Huang, R. Guigo, and R. Shiekhattar. Long noncoding RNAs with enhancer-like function in human cells. *Cell*, 143(1):46–58, Oct 2010. doi: 10.1016/j.cell.2010.09.001.

T. Panaretakis, O. Kepp, U. Brockmeier, A. Tesniere, A.-C. Bjorklund, D. C. Chapman, M. Durchschlag, N. Joza, G. Pierron, P. van Endert, J. Yuan, L. Zitvogel, F. Madeo, D. B. Williams, and G. Kroemer. Mechanisms of pre-apoptotic calreticulin exposure in immunogenic cell death. *EMBO J*, 28(5):578–590, Mar 2009. doi: 10.1038/emboj.2009.1.

G. A. Patikoglou, J. L. Kim, L. Sun, S. H. Yang, T. Kodadek, and S. K. Burley. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev*, 13(24):3217–3230, Dec 1999.

G. D. Penny, G. F. Kay, S. A. Sheardown, S. Rastan, and N. Brockdorff. Requirement for Xist in X chromosome inactivation. *Nature*, 379(6561):131–137, Jan 1996. doi: 10.1038/379131a0.

V. Perissi, A. Aggarwal, C. K. Glass, D. W. Rose, and M. G. Rosenfeld. A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. *Cell*, 116(4):511–526, Feb 2004.

B. M. Peterlin and D. H. Price. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell*, 23(3):297–305, Aug 2006. doi: 10.1016/j.molcel.2006.06.014.

S. E. Pike, L. Yao, K. D. Jones, B. Cherney, E. Appella, K. Sakaguchi, H. Nakhasi, J. Teruya-Feldstein, P. Wirth, G. Gupta, and G. Tosato. Vasostatin, a calreticulin

fragment, inhibits angiogenesis and suppresses tumor growth. *J Exp Med*, 188(12): 2349–2356, Dec 1998.

S. E. Pike, L. Yao, J. Setsuda, K. D. Jones, B. Cherney, E. Appella, K. Sakaguchi, H. Nakhasi, C. D. Atreya, J. Teruya-Feldstein, P. Wirth, G. Gupta, and G. Tosato. Calreticulin and calreticulin fragments are endothelial cell inhibitors that suppress tumor growth. *Blood*, 94(7):2461–2468, Oct 1999.

K. Plath, J. Fang, S. K. Mlynarczyk-Evans, R. Cao, K. A. Worringer, H. Wang, C. C. de la Cruz, A. P. Otte, B. Panning, and Y. Zhang. Role of histone H3 lysine 27 methylation in X inactivation. *Science*, 300(5616):131–135, Apr 2003. doi: 10.1126/science.1084274.

J. Ponjavic, B. Lenhard, C. Kai, J. Kawai, P. Carninci, Y. Hayashizaki, and A. Sandelin. Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol*, 7(8):R78, 2006. doi: 10.1186/gb-2006-7-8-R78.

J. Pontén and E. Saksela. Two established in vitro cell lines from human mesenchymal tumours. *Int J Cancer*, 2(5):434–447, Sep 1967.

P. Preker, J. Nielsen, S. Kammler, S. Lykke-Andersen, M. S. Christensen, C. K. Mapendano, M. H. Schierup, and T. H. Jensen. RNA exosome depletion reveals transcription upstream of active human promoters. *Science*, 322(5909):1851–1854, Dec 2008. doi: 10.1126/science.1164096.

P. Preker, K. Almvig, M. S. Christensen, E. Valen, C. K. Mapendano, A. Sandelin, and T. H. Jensen. PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res*, 39(16):7179–7193, Sep 2011. doi: 10.1093/nar/gkr370.

N. J. Proudfoot. Ending the message: poly(A) signals then and now. *Genes Dev*, 25(17): 1770–1782, Sep 2011. doi: 10.1101/gad.17268411.

Y. Ramanathan, S. M. Rajpara, S. M. Reza, E. Lees, S. Shuman, M. B. Mathews, and T. Pe'ery. Three RNA polymerase II carboxyl-terminal domain kinases display distinct substrate preferences. *J Biol Chem*, 276(14):10913–10920, Apr 2001. doi: 10.1074/jbc.M010975200.

F. Recillas-Targa, M. J. Pikaart, B. Burgess-Beusse, A. C. Bell, M. D. Litt, A. G. West, M. Gaszner, and G. Felsenfeld. Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A*, 99(10):6883–6888, May 2002. doi: 10.1073/pnas.102179399.

J.-J. M. Riethoven. Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. *Methods Mol Biol*, 674:33–42, 2010. doi: 10.1007/978-1-60761-854-6_3.

J. L. Rinn and H. Y. Chang. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*, 81:145–166, 2012. doi: 10.1146/annurev-biochem-051410-092902.

J. L. Rinn, M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal, and H. Y. Chang. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, 129(7):1311–1323, Jun 2007. doi: 10.1016/j.cell.2007.05.022.

A. Roberts, H. Pimentel, C. Trapnell, and L. Pachter. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*, 27(17):2325–2329, Sep 2011. doi: 10.1093/bioinformatics/btr355.

M. G. Rosenfeld, V. V. Lunyak, and C. K. Glass. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes Dev*, 20(11):1405–1428, Jun 2006. doi: 10.1101/gad.1424806.

A. L. Roy and D. S. Singer. Core promoters in transcription: old problem, new insights. *Trends Biochem Sci*, 40(3):165–171, Mar 2015. doi: 10.1016/j.tibs.2015.01.007.

T. Sakai, N. Ohtani, T. L. McGee, P. D. Robbins, and T. P. Dryja. Oncogenic germ-line mutations in Sp1 and ATF sites in the human retinoblastoma gene. *Nature*, 353(6339): 83–86, Sep 1991. doi: 10.1038/353083a0.

A. Sandelin, P. Carninci, B. Lenhard, J. Ponjavic, Y. Hayashizaki, and D. A. Hume. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet*, 8(6):424–436, Jun 2007. doi: 10.1038/nrg2026.

E. Savoysky, T. Mizuno, Y. Sowa, H. Watanabe, J. Sawada, H. Nomura, Y. Ohsugi, H. Handa, and T. Sakai. The retinoblastoma binding factor 1 (RBF-1) site in RB gene promoter binds preferentially E4TF1, a member of the Ets transcription factors family. *Oncogene*, 9(7):1839–1846, Jul 1994.

K. Schaukowitch, J.-Y. Joo, X. Liu, J. K. Watts, C. Martinez, and T.-K. Kim. Enhancer RNA facilitates NELF release from immediate early genes. *Mol Cell*, 56(1):29–42, Oct 2014. doi: 10.1016/j.molcel.2014.08.023.

W. F. SCHERER, J. T. SYVERTON, and G. O. GEY. Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. *J Exp Med*, 97(5):695–710, May 1953.

M. Seifert and J. Reichrath. The role of the human DNA mismatch repair gene hMSH2 in DNA repair, cell cycle control and apoptosis: implications for pathogenesis, progression and therapy of cancer. *J Mol Histol*, 37(5-7):301–307, Sep 2006. doi: 10.1007/s10735-006-9062-5.

A. C. Seila, J. M. Calabrese, S. S. Levine, G. W. Yeo, P. B. Rahl, R. A. Flynn, R. A. Young, and P. A. Sharp. Divergent transcription from active promoters. *Science*, 322 (5909):1849–1851, Dec 2008. doi: 10.1126/science.1162253.

A. C. Seila, L. J. Core, J. T. Lis, and P. A. Sharp. Divergent transcription: a new feature of active promoters. *Cell Cycle*, 8(16):2557–2564, Aug 2009.

B. Shan, C. Y. Chang, D. Jones, and W. H. Lee. The transcription factor E2F-1 mediates the autoregulation of RB gene expression. *Mol Cell Biol*, 14(1):299–309, Jan 1994.

A. J. Shatkin and J. L. Manley. The ends of the affair: capping and polyadenylation. *Nat Struct Biol*, 7(10):838–842, Oct 2000. doi: 10.1038/79583.

C. J. Sherr. Cancer cell cycles. *Science*, 274(5293):1672–1677, Dec 1996.

Y. Shiio, T. Yamamoto, and N. Yamaguchi. Negative regulation of Rb expression by the p53 gene product. *Proc Natl Acad Sci U S A*, 89(12):5206–5210, Jun 1992.

A. Siepel, G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, and D. Haussler. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*, 15(8):1034–1050, Aug 2005. doi: 10.1101/gr.3715005.

A. A. Sigova, A. C. Mullen, B. Molinie, S. Gupta, D. A. Orlando, M. G. Guenther, A. E. Almada, C. Lin, P. A. Sharp, C. C. Giallourakis, and R. A. Young. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci U S A*, 110(8):2876–2881, Feb 2013. doi: 10.1073/pnas.1221904110.

M. D. Simon, C. I. Wang, P. V. Kharchenko, J. A. West, B. A. Chapman, A. A. Alekseyenko, M. L. Borowsky, M. I. Kuroda, and R. E. Kingston. The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci U S A*, 108(51):20497–20502, Dec 2011. doi: 10.1073/pnas.1113536108.

S. T. Smale and J. T. Kadonaga. The RNA polymerase II core promoter. *Annu Rev Biochem*, 72:449–479, 2003. doi: 10.1146/annurev.biochem.72.121801.161520.

T. M. M. Søgaard and J. Q. Svejstrup. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J Biol Chem*, 282(19):14113–14120, May 2007. doi: 10.1074/jbc.M701345200.

Y. Sowa, Y. Shiio, T. Fujita, T. Matsumoto, Y. Okuyama, D. Kato, J. Inoue, J. Sawada, M. Goto, H. Watanabe, H. Handa, and T. Sakai. Retinoblastoma binding factor 1 site in the core promoter region of the human RB gene is activated by hGABP/E4TF1. *Cancer Res*, 57(15):3145–3148, Aug 1997.

L. Srinivasan and M. L. Atchison. YY1 DNA binding and PcG recruitment requires CtBP. *Genes Dev*, 18(21):2596–2601, Nov 2004. doi: 10.1101/gad.1228204.

J. A. Stamatoyannopoulos, A. Goodwin, T. Joyce, and C. H. Lowrey. NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region. *EMBO J*, 14(1):106–116, Jan 1995.

W. Sun and J. Yang. Functional mechanisms for human tumor suppressors. *J Cancer*, 1:136–140, 2010.

Y. Suzuki, T. Tsunoda, J. Sese, H. Taira, J. Mizushima-Sugano, H. Hata, T. Ota, T. Isogai, T. Tanaka, Y. Nakamura, A. Suyama, Y. Sakaki, S. Morishita, K. Okubo, and S. Sugano. Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res*, 11(5):677–684, May 2001. doi: 10.1101/gr.164001.

R. J. Taft, E. A. Glazov, N. Cloonan, C. Simons, S. Stephen, G. J. Faulkner, T. Lassmann, A. R. R. Forrest, S. M. Grimmond, K. Schroder, K. Irvine, T. Arakawa, M. Nakamura, A. Kubosaki, K. Hayashida, C. Kawazu, M. Murata, H. Nishiyori, S. Fukuda, J. Kawai, C. O. Daub, D. A. Hume, H. Suzuki, V. Orlando, P. Carninci, Y. Hayashizaki, and J. S. Mattick. Tiny RNAs associated with transcription start sites in animals. *Nat Genet*, 41(5):572–578, May 2009. doi: 10.1038/ng.312.

T. Takaki, K. Fukasawa, I. Suzuki-Takahashi, and H. Hirai. Cdk-mediated phosphorylation of pRB regulates HDAC binding in vitro. *Biochem Biophys Res Commun*, 316 (1):252–255, Mar 2004. doi: 10.1016/j.bbrc.2004.02.044.

S. M. Tan-Wong, J. B. Zaugg, J. Camblong, Z. Xu, D. W. Zhang, H. E. Mischo, A. Z. Ansari, N. M. Luscombe, L. M. Steinmetz, and N. J. Proudfoot. Gene loops enhance transcriptional directionality. *Science*, 338(6107):671–675, Nov 2012. doi: 10.1126/science.1224350.

The GENCODE Consortium. Version 22 (October 2014 freeze, GRCh38) - Ensembl 79. 2014. `http://www.gencodegenes.org/releases/22.html`.

N. D. Trinklein, S. F. Aldred, S. J. Hartman, D. I. Schroeder, R. P. Otillar, and R. M. Myers. An abundance of bidirectional promoters in the human genome. *Genome Res*, 14(1):62–66, Jan 2004. doi: 10.1101/gr.1982804.

K. Y. Tsai, Y. Hu, K. F. Macleod, D. Crowley, L. Yamasaki, and T. Jacks. Mutation of E2f-1 suppresses apoptosis and inappropriate S phase entry and extends survival of Rb-deficient mouse embryos. *Mol Cell*, 2(3):293–304, Sep 1998.

M.-C. Tsai, O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang, F. Lan, Y. Shi, E. Segal, and H. Y. Chang. Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, 329(5992):689–693, Aug 2010. doi: 10.1126/science.1192002.

R. K. Tsai and D. E. Discher. Inhibition of "self" engulfment through deactivation of myosin-II at the phagocytic synapse between human cells. *J Cell Biol*, 180(5):989–1003, Mar 2008. doi: 10.1083/jcb.200708043.

S. Tsuchiya, M. Yamabe, Y. Yamaguchi, Y. Kobayashi, T. Konno, and K. Tada. Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *Int J Cancer*, 26(2):171–176, Aug 1980.

M. Uesaka, O. Nishimura, Y. Go, K. Nakashima, K. Agata, and T. Imamura. Bidirectional promoters are the major source of gene activation-associated non-coding RNAs in mammals. *BMC Genomics*, 15:35, 2014. doi: 10.1186/1471-2164-15-35.

I. Ulitsky and D. P. Bartel. lincRNAs: genomics, evolution, and mechanisms. *Cell*, 154 (1):26–46, Jul 2013. doi: 10.1016/j.cell.2013.06.020.

A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. Primer3–new capabilities and interfaces. *Nucleic Acids Res*, 40(15):e115, Aug 2012. doi: 10.1093/nar/gks596.

S. Valadkhan. snRNAs as the catalysts of pre-mRNA splicing. *Curr Opin Chem Biol*, 9 (6):603–608, Dec 2005. doi: 10.1016/j.cbpa.2005.10.008.

J. Vandamme, P. Völkel, C. Rosnoblet, P. Le Faou, and P.-O. Angrand. Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. *Mol Cell Proteomics*, 10(4):M110.002642, Apr 2011. doi: 10.1074/mcp.M110.002642.

B. J. Venters and B. F. Pugh. Genomic organization of human transcription initiation complexes. *Nature*, 502(7469):53–58, Oct 2013. doi: 10.1038/nature12535.

P. Viatour and J. Sage. Newly identified aspects of tumor suppression by RB. *Dis Model Mech*, 4(5):581–585, Sep 2011. doi: 10.1242/dmm.008060.

C. Wakano, J. S. Byun, L.-J. Di, and K. Gardner. The dual lives of bidirectional promoters. *Biochim Biophys Acta*, 1819(7):688–693, Jul 2012. doi: 10.1016/j.bbagrm.2012.02.006.

K. C. Wang, Y. W. Yang, B. Liu, A. Sanyal, R. Corces-Zimmerman, Y. Chen, B. R. Lajoie, A. Protacio, R. A. Flynn, R. A. Gupta, J. Wysocka, M. Lei, J. Dekker, J. A. Helms, and H. Y. Chang. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, 472(7341):120–124, Apr 2011. doi: 10.1038/nature09819.

W.-A. Wang, J. Groenendyk, and M. Michalak. Calreticulin signaling in health and disease. *Int J Biochem Cell Biol*, 44(6):842–846, Jun 2012. doi: 10.1016/j.biocel.2012.02.009.

X. Wang, S. Arai, X. Song, D. Reichart, K. Du, G. Pascual, P. Tempst, M. G. Rosenfeld, C. K. Glass, and R. Kurokawa. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature*, 454(7200):126–130, Jul 2008. doi: 10.1038/nature06992.

M. Waser, N. Mesaeli, C. Spencer, and M. Michalak. Regulation of calreticulin gene expression by calcium. *J Cell Biol*, 138(3):547–557, Aug 1997.

I. Whitehouse, O. J. Rando, J. Delrow, and T. Tsukiyama. Chromatin remodelling at promoters suppresses antisense transcription. *Nature*, 450(7172):1031–1035, Dec 2007. doi: 10.1038/nature06391.

X. Wu and P. A. Sharp. Divergent transcription: a driving force for new gene origination? *Cell*, 155(5):990–996, Nov 2013. doi: 10.1016/j.cell.2013.10.048.

X. Xie, J. Lu, E. J. Kulbokas, T. R. Golub, V. Mootha, K. Lindblad-Toh, E. S. Lander, and M. Kellis. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, 434(7031):338–345, Mar 2005. doi: 10.1038/nature03441.

Y. Yamaguchi, T. Takagi, T. Wada, K. Yano, A. Furuya, S. Sugimoto, J. Hasegawa, and H. Handa. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*, 97(1):41–51, Apr 1999.

M. Q. Yang and L. L. Elnitski. Diversity of core promoter elements comprising human bidirectional promoters. *BMC Genomics*, 9 Suppl 2:S3, 2008. doi: 10.1186/1471-2164-9-S2-S3.

M. Q. Yang, L. M. Koehly, and L. L. Elnitski. Comprehensive annotation of bidirectional promoters identifies co-regulation among breast and ovarian cancer genes. *PLoS Comput Biol*, 3(4):e72, Apr 2007. doi: 10.1371/journal.pcbi.0030072.

W. Yu, D. Gius, P. Onyango, K. Muldoon-Jacobs, J. Karp, A. P. Feinberg, and H. Cui. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*, 451(7175):202–206, Jan 2008. doi: 10.1038/nature06468.

N. Yudkovsky, J. A. Ranish, and S. Hahn. A transcription reinitiation intermediate that is stabilized by activator. *Nature*, 408(6809):225–229, Nov 2000. doi: 10.1038/35041603.

J. Zeitlinger, A. Stark, M. Kellis, J.-W. Hong, S. Nechaev, K. Adelman, M. Levine, and R. A. Young. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat Genet*, 39(12):1512–1516, Dec 2007. doi: 10.1038/ng.2007.26.

D. Zheng and B. Tian. RNA-binding proteins in regulation of alternative cleavage and polyadenylation. *Adv Exp Med Biol*, 825:97–127, 2014. doi: 10.1007/978-1-4939-1221-6_3.

Y. Zhou, X. Zhang, and A. Klibanski. MEG3 noncoding RNA: a tumor suppressor. *J Mol Endocrinol*, 48(3):R45–R53, Jun 2012. doi: 10.1530/JME-12-0008.

List of Figures

List of Tables

Appendix A

Table A1: List of 1,107 bidirectionally expressed ncRNA/PCG pairs

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
1	1822910	1822911	ENST00000412228.1	+	1822494	1822495	ENSG00000078369.13	-	415
1	3816936	3816937	ENST00000413332.1	+	3816856	3816857	ENSG00000198912.6	-	79
1	3816968	3816969	ENST00000442673.1	+	3816856	3816857	ENSG00000198912.6	-	111
1	3816980	3816981	ENST00000439488.1	+	3816856	3816857	ENSG00000198912.6	-	123
1	6296300	6296301	ENST00000441724.1	+	6296031	6296032	ENSG00000116237.11	-	268
1	8086798	8086799	ENST00000445300.1	+	8086367	8086368	ENSG00000116285.8	-	430
1	21059373	21059374	ENST00000436642.1	+	21059329	21059330	ENSG00000189410.7	-	43
1	28969740	28969741	ENST00000420776.1	+	28969596	28969597	ENSG00000120656.7	-	143
1	40254648	40254649	ENST00000566366.1	+	40254532	40254533	ENSG00000116985.6	-	115
1	41708045	41708046	ENST00000425554.1	+	41707825	41707826	ENSG00000010803.12	-	219
1	41708247	41708248	ENST00000445073.1	+	41707825	41707826	ENSG00000010803.12	-	421
1	43424720	43424721	ENST00000431759.1	+	43424529	43424530	ENSG00000117394.15	-	190
1	43424775	43424776	ENST00000416689.1	+	43424529	43424530	ENSG00000117394.15	-	245
1	53686335	53686336	ENST00000569869.1	+	53686288	53686289	ENSG00000162384.9	-	46
1	53704282	53704283	ENST00000458151.1	+	53704281	53704282	ENSG00000162385.6	-	0
1	53793905	53793906	ENST00000445039.2	+	53793741	53793742	ENSG00000157193.10	-	163
1	55353236	55353237	ENST00000443284.1	+	55352890	55352891	ENSG00000116133.7	-	345
1	63154153	63154154	ENST00000453229.1	+	63153968	63153969	ENSG00000116641.11	-	184
1	71547036	71547037	ENST00000596952.1	+	71546979	71546980	ENSG00000132485.8	-	56
1	71547044	71547045	ENST00000413421.1	+	71546979	71546980	ENSG00000132485.8	-	64
1	94312730	94312731	ENST00000565336.1	+	94312705	94312706	ENSG00000137936.12	-	24
1	95393122	95393123	ENST00000452846.1	+	95392833	95392834	ENSG00000117519.11	-	288
1	110950999	110951000	ENST00000608253.1	+	110950563	110950564	ENSG00000134248.9	-	435
1	113258294	113258295	ENST00000566195.1	+	113258098	113258099	ENSG00000155367.11	-	195
1	119683331	119683332	ENST00000457043.1	+	119683293	119683294	ENSG00000116874.7	-	37
1	119683353	119683354	ENST00000418015.1	+	119683293	119683294	ENSG00000116874.7	-	59
1	145827205	145827206	ENST00000437377.1	+	145827102	145827103	ENSG00000117262.14	-	102
1	146644350	146644351	ENST00000440377.2	+	146644128	146644129	ENSG00000131791.6	-	221
1	147634989	147634990	ENST00000432038.1	+	147634885	147634886	ENSG00000203836.7	-	103
1	147635092	147635093	ENST00000411978.1	+	147634885	147634886	ENSG00000203836.7	-	206
1	147635156	147635157	ENST00000598757.1	+	147634885	147634886	ENSG00000203836.7	-	270
1	147635200	147635201	ENST00000608244.1	+	147634885	147634886	ENSG00000203836.7	-	314
1	151300425	151300426	ENST00000609583.1	+	151300190	151300191	ENSG00000143393.12	-	234
1	153950219	153950220	ENST00000608236.1	+	153950163	153950164	ENSG00000143543.10	-	55
1	154909846	154909847	ENST00000604546.1	+	154909466	154909467	ENSG00000163344.5	-	379
1	161337812	161337813	ENST00000437833.2	+	161337663	161337664	ENSG00000188931.3	-	148
1	173991647	173991648	ENST00000424181.1	+	173991434	173991435	ENSG00000135870.7	-	212
1	176176784	176176785	ENST00000456125.1	+	176176628	176176629	ENSG00000143207.15	-	155
1	185286911	185286912	ENST00000609881.1	+	185286460	185286461	ENSG00000116679.11	-	450
1	186649754	186649755	ENST00000608917.1	+	186649558	186649559	ENSG00000073756.7	-	195
1	200993077	200993078	ENST00000446333.1	+	200992827	200992828	ENSG00000116852.10	-	249
1	200993089	200993090	ENST00000458003.1	+	200992827	200992828	ENSG00000116852.10	-	261
1	211849104	211849105	ENST00000415202.1	+	211848959	211848960	ENSG00000117650.8	-	144
1	212004303	212004304	ENST00000430623.1	+	212004113	212004114	ENSG00000123684.8	-	189
1	222763304	222763305	ENST00000413074.1	+	222763274	222763275	ENSG00000143498.13	-	29
1	229644248	229644249	ENST00000417605.1	+	229644102	229644103	ENSG00000069248.9	-	145
1	245028040	245028041	ENST00000610145.1	+	245027843	245027844	ENSG00000153187.12	-	196
1	249153363	249153364	ENST00000417047.1	+	249153342	249153343	ENSG00000171163.11	-	20
10	6622381	6622382	ENST00000445427.1	+	6622262	6622263	ENSG00000065675.10	-	118
10	6622387	6622388	ENST00000455810.1	+	6622262	6622263	ENSG00000065675.10	-	124
10	21463283	21463284	ENST00000417845.1	+	21463115	21463116	ENSG00000078114.14	-	167
10	35104695	35104696	ENST00000446211.1	+	35104252	35104253	ENSG00000148498.11	-	442
10	38265753	38265754	ENST00000412789.1	+	38265560	38265561	ENSG00000175395.11	-	192
10	75385754	75385755	ENST00000595595.1	+	75385710	75385711	ENSG00000166348.13	-	43
10	88281702	88281703	ENST00000428940.2	+	88281571	88281572	ENSG00000062650.13	-	130
10	101380812	101380813	ENST00000566847.1	+	101380365	101380366	ENSG00000155287.6	-	446

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
10	104211359	104211360	ENST00000597488.1	+	104211299	104211300	ENSG00000120055.5	-	59
10	119806335	119806336	ENST00000435944.1	+	119806113	119806114	ENSG00000107560.6	-	221
10	119806505	119806506	ENST00000426021.1	+	119806113	119806114	ENSG00000107560.6	-	391
10	119806515	119806516	ENST00000454781.1	+	119806113	119806114	ENSG00000107560.6	-	401
10	119806535	119806536	ENST00000454857.1	+	119806113	119806114	ENSG00000107560.6	-	421
10	127371798	127371799	ENST00000596068.1	+	127371712	127371713	ENSG00000175018.8	-	85
10	127371808	127371809	ENST00000607914.1	+	127371712	127371713	ENSG00000175018.8	-	95
10	127371812	127371813	ENST00000415305.2	+	127371712	127371713	ENSG00000175018.8	-	99
10	127371862	127371863	ENST00000449693.1	+	127371712	127371713	ENSG00000175018.8	-	149
11	1330999	1331000	ENST00000530897.1	+	1330883	1330884	ENSG00000078902.11	-	115
11	5959981	5959982	ENST00000528915.1	+	5959848	5959849	ENSG00000132256.14	-	132
11	8190714	8190715	ENST00000499752.2	+	8190601	8190602	ENSG00000166405.10	-	112
11	8986595	8986596	ENST00000532599.1	+	8986557	8986558	ENSG00000175348.6	-	37
11	32457322	32457323	ENST00000459866.1	+	32457175	32457176	ENSG00000184937.8	-	146
11	32457347	32457348	ENST00000525436.1	+	32457175	32457176	ENSG00000184937.8	-	171
11	33796245	33796246	ENST00000533046.1	+	33796088	33796089	ENSG00000110429.9	-	156
11	33796251	33796252	ENST00000530352.1	+	33796088	33796089	ENSG00000110429.9	-	162
11	59436785	59436786	ENST00000534120.1	+	59436452	59436453	ENSG00000166889.13	-	332
11	60674261	60674262	ENST00000544421.1	+	60674059	60674060	ENSG00000110107.4	-	201
11	64014526	64014527	ENST00000538355.1	+	64014412	64014413	ENSG00000173457.6	-	113
11	64546426	64546427	ENST00000594089.1	+	64546257	64546258	ENSG00000168066.16	-	168
11	77850817	77850818	ENST00000532831.1	+	77850705	77850706	ENSG00000159063.8	-	111
11	77850849	77850850	ENST00000500113.1	+	77850705	77850706	ENSG00000159063.8	-	143
11	77850861	77850862	ENST00000527321.1	+	77850705	77850706	ENSG00000159063.8	-	155
11	82783108	82783109	ENST00000527627.1	+	82782964	82782965	ENSG00000137502.5	-	143
11	82783124	82783125	ENST00000526795.1	+	82782964	82782965	ENSG00000137502.5	-	159
11	82783146	82783147	ENST00000533528.1	+	82782964	82782965	ENSG00000137502.5	-	181
11	82783160	82783161	ENST00000533708.1	+	82782964	82782965	ENSG00000137502.5	-	195
11	82783164	82783165	ENST00000534499.1	+	82782964	82782965	ENSG00000137502.5	-	199
11	82783385	82783386	ENST00000528156.1	+	82782964	82782965	ENSG00000137502.5	-	420
11	82783397	82783398	ENST00000530270.2	+	82782964	82782965	ENSG00000137502.5	-	432
11	86666661	86666662	ENST00000499504.3	+	86666432	86666433	ENSG00000174804.3	-	228
11	94965735	94965736	ENST00000543573.1	+	94965704	94965705	ENSG00000149212.6	-	30
11	116644105	116644106	ENST00000439104.1	+	116643703	116643704	ENSG00000137656.7	-	401
11	119252488	119252489	ENST00000577297.1	+	119252435	119252436	ENSG00000036672.11	-	52
11	119600293	119600294	ENST00000533253.1	+	119599793	119599794	ENSG00000110400.6	-	499
11	124632327	124632328	ENST00000532579.1	+	124632185	124632186	ENSG00000149564.7	-	141
11	124670806	124670807	ENST00000529392.1	+	124670568	124670569	ENSG00000120458.5	-	237
11	129872833	129872834	ENST00000530583.1	+	129872729	129872730	ENSG00000170325.10	-	103
11	130184888	130184889	ENST00000602376.1	+	130184580	130184581	ENSG00000196323.7	-	307
11	130184923	130184924	ENST00000532116.3	+	130184580	130184581	ENSG00000196323.7	-	342
11	130184934	130184935	ENST00000602310.1	+	130184580	130184581	ENSG00000196323.7	-	353
12	13153385	13153386	ENST00000543321.1	+	13153206	13153207	ENSG00000013583.4	-	178
12	30908008	30908009	ENST00000500076.2	+	30907884	30907885	ENSG00000110888.13	-	123
12	31744246	31744247	ENST00000537346.1	+	31744030	31744031	ENSG00000170456.10	-	215
12	49182930	49182931	ENST00000547774.1	+	49182819	49182820	ENSG00000174233.7	-	110
12	49525493	49525494	ENST00000551496.1	+	49525179	49525180	ENSG00000123416.11	-	313
12	54813569	54813570	ENST00000552053.1	+	54813243	54813244	ENSG00000161638.6	-	325
12	57824899	57824900	ENST00000547552.1	+	57824787	57824788	ENSG00000179912.15	-	111
12	59314420	59314421	ENST00000547590.1	+	59314302	59314303	ENSG00000139263.7	-	117
12	65153301	65153302	ENST00000434563.3	+	65153226	65153227	ENSG00000135677.6	-	74
12	92539957	92539958	ENST00000499685.2	+	92539672	92539673	ENSG00000133639.3	-	284
12	100536836	100536837	ENST00000550886.1	+	100536625	100536626	ENSG00000111647.8	-	210
12	110318481	110318482	ENST00000446473.2	+	110318292	110318293	ENSG00000139433.5	-	188
12	111807086	111807087	ENST00000552663.1	+	111806924	111806925	ENSG00000198324.10	-	161
12	114846559	114846560	ENST00000528549.1	+	114846246	114846247	ENSG00000089225.15	-	312
12	117537286	117537287	ENST00000547006.1	+	117537283	117537284	ENSG00000088992.13	-	2
12	123849500	123849501	ENST00000543072.1	+	123849389	123849390	ENSG00000139697.7	-	110
13	27746396	27746397	ENST00000452222.1	+	27746032	27746033	ENSG00000152484.9	-	363
13	45915554	45915555	ENST00000412946.2	+	45915504	45915505	ENSG00000133112.12	-	49
13	45915555	45915556	ENST00000520622.1	+	45915504	45915505	ENSG00000133112.12	-	50
13	45915564	45915565	ENST00000520590.1	+	45915504	45915505	ENSG00000133112.12	-	59
13	45915626	45915627	ENST00000523506.1	+	45915504	45915505	ENSG00000133112.12	-	121
13	45915646	45915647	ENST00000521336.1	+	45915504	45915505	ENSG00000133112.12	-	141
13	45915648	45915649	ENST00000524062.1	+	45915504	45915505	ENSG00000133112.12	-	143
13	45915673	45915674	ENST00000520310.1	+	45915504	45915505	ENSG00000133112.12	-	168
13	45915727	45915728	ENST00000520924.1	+	45915504	45915505	ENSG00000133112.12	-	222
13	52378433	52378434	ENST00000456688.1	+	52378292	52378293	ENSG00000102796.6	-	140
13	74993310	74993311	ENST00000423629.1	+	74993251	74993252	ENSG00000177596.1	-	58
13	114567141	114567142	ENST00000608651.1	+	114567045	114567046	ENSG00000183087.10	-	95
14	21852451	21852452	ENST00000565098.1	+	21852424	21852425	ENSG00000092201.5	-	26
14	44976610	44976611	ENST00000557465.1	+	44976481	44976482	ENSG00000189139.5	-	128
14	44976612	44976613	ENST00000555433.1	+	44976481	44976482	ENSG00000189139.5	-	130
14	53620072	53620073	ENST00000554235.1	+	53619999	53620000	ENSG00000100523.10	-	72

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
14	64010287	64010288	ENST00000561909.1	+	64010091	64010092	ENSG00000154001.9	-	195
14	69446399	69446400	ENST00000553961.1	+	69446156	69446157	ENSG00000072110.9	-	242
14	91884610	91884611	ENST00000557524.1	+	91884187	91884188	ENSG00000015133.14	-	422
14	101295638	101295639	ENST00000523671.2	+	101295536	101295537	ENSG00000267918.1	-	101
14	101295948	101295949	ENST00000452514.2	+	101295536	101295537	ENSG00000267918.1	-	411
14	104314058	104314059	ENST00000556586.1	+	104313926	104313927	ENSG00000088808.12	-	131
15	26110442	26110443	ENST00000557558.1	+	26110316	26110317	ENSG00000206190.7	-	125
15	35838396	35838397	ENST00000501169.2	+	35838393	35838394	ENSG00000134146.7	-	2
15	35838584	35838585	ENST00000559210.1	+	35838393	35838394	ENSG00000134146.7	-	190
15	40213243	40213244	ENST00000499797.2	+	40213092	40213093	ENSG00000166073.4	-	150
15	40213271	40213272	ENST00000558675.1	+	40213092	40213093	ENSG00000166073.4	-	178
15	40331512	40331513	ENST00000504245.1	+	40331388	40331389	ENSG00000140319.6	-	123
15	40331550	40331551	ENST00000560341.1	+	40331388	40331389	ENSG00000140319.6	-	161
15	48938148	48938149	ENST00000558061.1	+	48938045	48938046	ENSG00000166147.9	-	102
15	50647664	50647665	ENST00000499624.2	+	50647604	50647605	ENSG00000104064.12	-	59
15	50647742	50647743	ENST00000561289.1	+	50647604	50647605	ENSG00000104064.12	-	137
15	50647750	50647751	ENST00000499326.1	+	50647604	50647605	ENSG00000104064.12	-	145
15	62352702	62352703	ENST00000560813.2	+	62352671	62352672	ENSG00000129003.11	-	30
15	62352713	62352714	ENST00000558368.2	+	62352671	62352672	ENSG00000129003.11	-	41
15	74753606	74753607	ENST00000499217.2	+	74753522	74753523	ENSG00000138629.11	-	83
15	74753648	74753649	ENST00000567286.1	+	74753522	74753523	ENSG00000138629.11	-	125
15	74753677	74753678	ENST00000564621.1	+	74753522	74753523	ENSG00000138629.11	-	154
15	74753685	74753686	ENST00000568853.1	+	74753522	74753523	ENSG00000138629.11	-	162
15	89878284	89878285	ENST00000569473.1	+	89878091	89878092	ENSG00000140521.7	-	192
15	89878502	89878503	ENST00000562356.1	+	89878091	89878092	ENSG00000140521.7	-	410
15	90645881	90645882	ENST00000561101.1	+	90645735	90645736	ENSG00000182054.5	-	145
15	91565849	91565850	ENST00000556904.1	+	91565832	91565833	ENSG00000184056.10	-	16
15	91565851	91565852	ENST00000557804.1	+	91565832	91565833	ENSG00000184056.10	-	18
15	91565852	91565853	ENST00000501381.3	+	91565832	91565833	ENSG00000184056.10	-	19
16	2014960	2014961	ENST00000531523.1	+	2014860	2014861	ENSG00000140988.11	-	99
16	2015035	2015036	ENST00000564014.1	+	2014860	2014861	ENSG00000140988.11	-	174
16	2015185	2015186	ENST00000459373.1	+	2014860	2014861	ENSG00000140988.11	-	324
16	2318624	2318625	ENST00000567888.1	+	2318412	2318413	ENSG00000205937.7	-	211
16	2318664	2318665	ENST00000562838.1	+	2318412	2318413	ENSG00000205937.7	-	251
16	8963108	8963109	ENST00000570290.1	+	8962865	8962866	ENSG00000153048.6	-	242
16	18938185	18938186	ENST00000565782.1	+	18937775	18937776	ENSG00000157106.12	-	409
16	21314568	21314569	ENST00000444326.1	+	21314403	21314404	ENSG00000103316.6	-	164
16	29875155	29875156	ENST00000398859.3	+	29875056	29875057	ENSG00000103502.9	-	98
16	30366762	30366763	ENST00000563252.1	+	30366681	30366682	ENSG00000169217.4	-	80
16	47007900	47007901	ENST00000562536.1	+	47007698	47007699	ENSG00000069345.7	-	201
16	48654230	48654231	ENST00000565055.1	+	48654058	48654059	ENSG00000102921.3	-	171
16	52112534	52112535	ENST00000568711.1	+	52112385	52112386	ENSG00000261190.1	-	148
16	67515217	67515218	ENST00000602592.1	+	67515139	67515140	ENSG00000159720.7	-	77
16	67518145	67518146	ENST00000602476.1	+	67517715	67517716	ENSG00000159723.4	-	429
16	81110924	81110925	ENST00000501068.2	+	81110871	81110872	ENSG00000166455.9	-	52
16	82203901	82203902	ENST00000563841.1	+	82203830	82203831	ENSG00000135698.5	-	70
16	84150656	84150657	ENST00000565382.1	+	84150510	84150511	ENSG00000140943.12	-	145
16	88729743	88729744	ENST00000565633.1	+	88729568	88729569	ENSG00000167508.6	-	174
16	88729777	88729778	ENST00000569786.1	+	88729568	88729569	ENSG00000167508.6	-	208
16	88729788	88729789	ENST00000563475.1	+	88729568	88729569	ENSG00000167508.6	-	219
17	5372412	5372413	ENST00000571506.1	+	5372379	5372380	ENSG00000005100.8	-	32
17	8080130	8080131	ENST00000581248.1	+	8079716	8079717	ENSG00000179029.10	-	413
17	8869213	8869214	ENST00000585297.1	+	8869028	8869029	ENSG00000141506.9	-	184
17	16557218	16557219	ENST00000577569.1	+	16557169	16557170	ENSG00000197566.5	-	48
17	19622372	19622373	ENST00000577087.2	+	19622291	19622292	ENSG00000180638.13	-	80
17	38083995	38083996	ENST00000578802.1	+	38083853	38083854	ENSG00000172057.5	-	141
17	38084097	38084098	ENST00000578478.1	+	38083853	38083854	ENSG00000172057.5	-	243
17	38084120	38084121	ENST00000582263.1	+	38083853	38083854	ENSG00000172057.5	-	266
17	40086888	40086889	ENST00000593239.1	+	40086794	40086795	ENSG00000131473.12	-	93
17	40086898	40086899	ENST00000377540.1	+	40086794	40086795	ENSG00000131473.12	-	103
17	40086909	40086910	ENST00000591658.1	+	40086794	40086795	ENSG00000131473.12	-	114
17	42299283	42299284	ENST00000563394.1	+	42298993	42298994	ENSG00000108312.10	-	289
17	43025280	43025281	ENST00000591013.1	+	43025081	43025082	ENSG00000186185.9	-	198
17	43339558	43339559	ENST00000585351.1	+	43339478	43339479	ENSG00000184361.8	-	79
17	58156670	58156671	ENST00000589740.1	+	58156291	58156292	ENSG00000068097.10	-	378
17	58603654	58603655	ENST00000559739.1	+	58603579	58603580	ENSG00000062725.5	-	74
17	58603660	58603661	ENST00000558027.1	+	58603579	58603580	ENSG00000062725.5	-	80
17	60885861	60885862	ENST00000584542.1	+	60885704	60885705	ENSG00000173838.7	-	156
17	76356534	76356535	ENST00000592569.1	+	76356157	76356158	ENSG00000184557.3	-	376
17	79885705	79885706	ENST00000582106.1	+	79885589	79885590	ENSG00000197063.6	-	115
17	79995781	79995782	ENST00000584705.1	+	79995607	79995608	ENSG00000169738.3	-	173
17	79995907	79995908	ENST00000582558.1	+	79995607	79995608	ENSG00000169738.3	-	299
18	268148	268149	ENST00000581677.1	+	268049	268050	ENSG00000079134.7	-	98
18	9615262	9615263	ENST00000582435.1	+	9615237	9615238	ENSG00000154845.11	-	24

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
18	70535623	70535624	ENST00000580564.1	+	70535380	70535381	ENSG00000166342.14	-	242
19	663482	663483	ENST00000591866.1	+	663276	663277	ENSG00000070423.13	-	205
19	1238178	1238179	ENST00000592843.1	+	1238025	1238026	ENSG00000099625.8	-	152
19	2328629	2328630	ENST00000452401.2	+	2328618	2328619	ENSG00000130332.10	-	10
19	2328658	2328659	ENST00000586332.1	+	2328618	2328619	ENSG00000130332.10	-	39
19	2328672	2328673	ENST00000590982.1	+	2328618	2328619	ENSG00000130332.10	-	53
19	2328676	2328677	ENST00000586377.2	+	2328618	2328619	ENSG00000130332.10	-	57
19	2328683	2328684	ENST00000593198.1	+	2328618	2328619	ENSG00000130332.10	-	64
19	2328692	2328693	ENST00000590295.2	+	2328618	2328619	ENSG00000130332.10	-	73
19	2328695	2328696	ENST00000592738.2	+	2328618	2328619	ENSG00000130332.10	-	76
19	4903092	4903093	ENST00000592666.1	+	4902878	4902879	ENSG00000205784.2	-	213
19	9609354	9609355	ENST00000589751.1	+	9609282	9609283	ENSG00000198028.3	-	71
19	9904297	9904298	ENST00000590046.1	+	9903855	9903856	ENSG00000196605.3	-	441
19	22715428	22715429	ENST00000598832.1	+	22715286	22715287	ENSG00000197360.5	-	141
19	22715473	22715474	ENST00000601708.1	+	22715286	22715287	ENSG00000197360.5	-	186
19	22715579	22715580	ENST00000594200.1	+	22715286	22715287	ENSG00000197360.5	-	292
19	33793763	33793764	ENST00000592982.2	+	33793469	33793470	ENSG00000245848.2	-	293
19	38307999	38308000	ENST00000589653.1	+	38307939	38307940	ENSG00000189144.9	-	59
19	38308051	38308052	ENST00000590433.1	+	38307939	38307940	ENSG00000189144.9	-	111
19	38308125	38308126	ENST00000592103.1	+	38307939	38307940	ENSG00000189144.9	-	185
19	47164735	47164736	ENST00000500689.1	+	47164394	47164395	ENSG00000197380.6	-	340
19	56905180	56905181	ENST00000593109.1	+	56904913	56904914	ENSG00000018869.12	-	266
19	56989500	56989501	ENST00000585445.1	+	56989433	56989434	ENSG00000198046.7	-	66
19	56989524	56989525	ENST00000586091.1	+	56989433	56989434	ENSG00000198046.7	-	90
19	56989526	56989527	ENST00000594783.1	+	56989433	56989434	ENSG00000198046.7	-	92
19	56989552	56989553	ENST00000588158.1	+	56989433	56989434	ENSG00000198046.7	-	118
19	56989559	56989560	ENST00000591797.1	+	56989433	56989434	ENSG00000198046.7	-	125
19	56989609	56989610	ENST00000601875.1	+	56989433	56989434	ENSG00000198046.7	-	175
19	57183636	57183637	ENST00000599726.1	+	57183150	57183151	ENSG00000127903.12	-	485
19	57352270	57352271	ENST00000599641.1	+	57352096	57352097	ENSG00000269699.1	-	173
19	57989017	57989018	ENST00000595422.1	+	57988937	57988938	ENSG00000197128.7	-	79
19	57989067	57989068	ENST00000594562.1	+	57988937	57988938	ENSG00000197128.7	-	129
19	58951815	58951816	ENST00000595059.1	+	58951588	58951589	ENSG00000131849.10	-	226
2	9696028	9696029	ENST00000607241.1	+	9695920	9695921	ENSG00000151694.8	-	107
2	10588820	10588821	ENST00000553181.1	+	10588629	10588630	ENSG00000115758.8	-	190
2	10830470	10830471	ENST00000607781.1	+	10830100	10830101	ENSG00000115761.11	-	369
2	20101786	20101787	ENST00000607190.1	+	20101746	20101747	ENSG00000183891.5	-	39
2	20251895	20251896	ENST00000452342.2	+	20251788	20251789	ENSG00000068697.6	-	106
2	25194995	25194996	ENST00000434897.1	+	25194962	25194963	ENSG00000115137.7	-	32
2	25195013	25195014	ENST00000428614.1	+	25194962	25194963	ENSG00000115137.7	-	50
2	25195034	25195035	ENST00000422449.1	+	25194962	25194963	ENSG00000115137.7	-	71
2	25195052	25195053	ENST00000421842.1	+	25194962	25194963	ENSG00000115137.7	-	89
2	27580007	27580008	ENST00000453289.1	+	27579867	27579868	ENSG00000115207.9	-	139
2	39664510	39664511	ENST00000443038.1	+	39664452	39664453	ENSG00000011566.10	-	57
2	39664543	39664544	ENST00000422128.1	+	39664452	39664453	ENSG00000011566.10	-	90
2	39664557	39664558	ENST00000449569.1	+	39664452	39664453	ENSG00000011566.10	-	104
2	39664572	39664573	ENST00000445520.1	+	39664452	39664453	ENSG00000011566.10	-	119
2	48133221	48133222	ENST00000439870.1	+	48132931	48132932	ENSG00000138081.15	-	289
2	51259739	51259740	ENST00000440698.1	+	51259673	51259674	ENSG00000179915.16	-	65
2	74375136	74375137	ENST00000529783.1	+	74375120	74375121	ENSG00000163170.7	-	15
2	74375166	74375167	ENST00000423477.2	+	74375120	74375121	ENSG00000163170.7	-	45
2	74375172	74375173	ENST00000533563.1	+	74375120	74375121	ENSG00000163170.7	-	51
2	86116403	86116404	ENST00000455121.3	+	86116136	86116137	ENSG00000115525.12	-	266
2	98280680	98280681	ENST00000450072.1	+	98280569	98280570	ENSG00000115073.6	-	110
2	98280702	98280703	ENST00000603172.1	+	98280569	98280570	ENSG00000115073.6	-	132
2	98280710	98280711	ENST00000605866.1	+	98280569	98280570	ENSG00000115073.6	-	140
2	98280724	98280725	ENST00000609604.1	+	98280569	98280570	ENSG00000115073.6	-	154
2	98280735	98280736	ENST00000609703.1	+	98280569	98280570	ENSG00000115073.6	-	165
2	122407226	122407227	ENST00000414554.2	+	122407162	122407163	ENSG00000074054.13	-	63
2	122407374	122407375	ENST00000413904.2	+	122407162	122407163	ENSG00000074054.13	-	211
2	122407559	122407560	ENST00000439321.1	+	122407162	122407163	ENSG00000074054.13	-	396
2	122407658	122407659	ENST00000447668.2	+	122407162	122407163	ENSG00000074054.13	-	495
2	166651361	166651362	ENST00000425688.1	+	166651191	166651192	ENSG00000115339.9	-	169
2	172967734	172967735	ENST00000448117.1	+	172967627	172967628	ENSG00000115844.6	-	106
2	175352117	175352118	ENST00000444196.1	+	175351821	175351822	ENSG00000163328.9	-	295
2	175352131	175352132	ENST00000417038.1	+	175351821	175351822	ENSG00000163328.9	-	309
2	175352240	175352241	ENST00000606406.1	+	175351821	175351822	ENSG00000163328.9	-	418
2	191399581	191399582	ENST00000457407.1	+	191399447	191399448	ENSG00000189362.7	-	133
2	198176117	198176118	ENST00000442984.1	+	198175896	198175897	ENSG00000065413.12	-	220
2	216300976	216300977	ENST00000412951.1	+	216300894	216300895	ENSG00000115414.14	-	81
2	227664862	227664863	ENST00000607970.1	+	227664474	227664475	ENSG00000169047.5	-	387
20	2644998	2644999	ENST00000418739.1	+	2644864	2644865	ENSG00000101365.16	-	133
20	8000549	8000550	ENST00000457707.1	+	8000475	8000476	ENSG00000125827.4	-	73
20	18040137	18040138	ENST00000429853.1	+	18039831	18039832	ENSG00000125850.6	-	305

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
20	19738792	19738793	ENST00000412571.1	+	19738678	19738679	ENSG00000268628.1	-	113
20	25604869	25604870	ENST00000420803.1	+	25604810	25604811	ENSG00000170191.4	-	58
20	32262323	32262324	ENST00000606866.1	+	32262268	32262269	ENSG00000125967.12	-	54
20	41818862	41818863	ENST00000430025.1	+	41818609	41818610	ENSG00000196090.8	-	252
20	42839600	42839601	ENST00000439943.1	+	42839430	42839431	ENSG00000132823.6	-	169
20	42839632	42839633	ENST00000437730.1	+	42839430	42839431	ENSG00000132823.6	-	201
20	42839722	42839723	ENST00000442383.1	+	42839430	42839431	ENSG00000132823.6	-	291
20	42839892	42839893	ENST00000435163.1	+	42839430	42839431	ENSG00000132823.6	-	461
20	47895179	47895180	ENST00000428008.1	+	47894962	47894963	ENSG00000124201.10	-	216
20	55841853	55841854	ENST00000412321.1	+	55841684	55841685	ENSG00000101144.8	-	168
20	55841858	55841859	ENST00000426580.1	+	55841684	55841685	ENSG00000101144.8	-	173
20	57090435	57090436	ENST00000427140.1	+	57090186	57090187	ENSG00000198768.6	-	248
20	57090555	57090556	ENST00000420279.1	+	57090186	57090187	ENSG00000198768.6	-	368
20	57090599	57090600	ENST00000447767.1	+	57090186	57090187	ENSG00000198768.6	-	412
20	57090624	57090625	ENST00000427794.1	+	57090186	57090187	ENSG00000198768.6	-	437
20	62258580	62258581	ENST00000449500.1	+	62258393	62258394	ENSG00000101216.6	-	186
20	62258603	62258604	ENST00000411579.1	+	62258393	62258394	ENSG00000101216.6	-	209
21	34100426	34100427	ENST00000458479.1	+	34100358	34100359	ENSG00000159082.13	-	67
21	46707967	46707968	ENST00000454115.2	+	46707812	46707813	ENSG00000186866.12	-	154
21	46707977	46707978	ENST00000400362.2	+	46707812	46707813	ENSG00000186866.12	-	164
22	26908503	26908504	ENST00000566814.1	+	26908470	26908471	ENSG00000100109.12	-	32
22	26908521	26908522	ENST00000565764.1	+	26908470	26908471	ENSG00000100109.12	-	50
22	29196671	29196672	ENST00000458080.1	+	29196584	29196585	ENSG00000100219.12	-	86
22	29196692	29196693	ENST00000418292.1	+	29196584	29196585	ENSG00000100219.12	-	107
22	29196697	29196698	ENST00000585003.1	+	29196584	29196585	ENSG00000100219.12	-	112
22	37099963	37099964	ENST00000430281.1	+	37099602	37099603	ENSG00000166862.6	-	360
22	42486971	42486972	ENST00000536447.2	+	42486958	42486959	ENSG00000184983.5	-	12
22	42487208	42487209	ENST00000595777.1	+	42486958	42486959	ENSG00000184983.5	-	249
22	42487356	42487357	ENST00000600968.1	+	42486958	42486959	ENSG00000184983.5	-	397
22	42487406	42487407	ENST00000434834.1	+	42486958	42486959	ENSG00000184983.5	-	447
22	43011250	43011251	ENST00000602478.1	+	43010967	43010968	ENSG00000100227.13	-	282
22	44208336	44208337	ENST00000563715.1	+	44208216	44208217	ENSG00000186976.10	-	119
22	44208373	44208374	ENST00000564696.1	+	44208216	44208217	ENSG00000186976.10	-	156
3	48885370	48885371	ENST00000412171.2	+	48885278	48885279	ENSG00000114302.11	-	91
3	48885390	48885391	ENST00000416209.2	+	48885278	48885279	ENSG00000114302.11	-	111
3	52273316	52273317	ENST00000464958.1	+	52273275	52273276	ENSG00000247596.4	-	40
3	57678932	57678933	ENST00000465933.1	+	57678815	57678816	ENSG00000174839.8	-	116
3	67705121	67705122	ENST00000464420.1	+	67705037	67705038	ENSG00000172340.10	-	83
3	67705182	67705183	ENST00000482677.1	+	67705037	67705038	ENSG00000172340.10	-	144
3	111852270	111852271	ENST00000563632.1	+	111852151	111852152	ENSG00000174500.8	-	118
3	119813742	119813743	ENST00000484076.1	+	119813263	119813264	ENSG00000082701.10	-	478
3	120068357	120068358	ENST00000494869.1	+	120068185	120068186	ENSG00000163428.3	-	171
3	129612714	129612715	ENST00000605830.1	+	129612418	129612419	ENSG00000172765.12	-	295
3	129612725	129612726	ENST00000605698.1	+	129612418	129612419	ENSG00000172765.12	-	306
3	136471472	136471473	ENST00000564748.1	+	136471219	136471220	ENSG00000118007.8	-	252
3	139108657	139108658	ENST00000504670.1	+	139108573	139108574	ENSG00000184432.5	-	83
3	139108660	139108661	ENST00000507362.1	+	139108573	139108574	ENSG00000184432.5	-	86
3	139108672	139108673	ENST00000512622.1	+	139108573	139108574	ENSG00000184432.5	-	98
3	139108680	139108681	ENST00000510068.1	+	139108573	139108574	ENSG00000184432.5	-	106
3	139108718	139108719	ENST00000514729.1	+	139108573	139108574	ENSG00000184432.5	-	144
3	149096006	149096007	ENST00000484046.1	+	149095651	149095652	ENSG00000169908.6	-	354
3	150421830	150421831	ENST00000475393.1	+	150421757	150421758	ENSG00000163645.10	-	72
3	180707569	180707570	ENST00000461063.2	+	180707561	180707562	ENSG00000205981.2	-	7
3	195638920	195638921	ENST00000448113.1	+	195638815	195638816	ENSG00000061938.12	-	104
3	195638965	195638966	ENST00000424819.1	+	195638815	195638816	ENSG00000061938.12	-	149
3	196669494	196669495	ENST00000602845.1	+	196669467	196669468	ENSG00000114503.6	-	26
4	1107427	1107428	ENST00000504969.1	+	1107349	1107350	ENSG00000178222.8	-	77
4	1243898	1243899	ENST00000514984.1	+	1243740	1243741	ENSG00000159692.11	-	157
4	1244047	1244048	ENST00000581398.1	+	1243740	1243741	ENSG00000159692.11	-	306
4	1714548	1714549	ENST00000605571.1	+	1714281	1714282	ENSG00000163950.8	-	266
4	2420701	2420702	ENST00000382849.2	+	2420389	2420390	ENSG00000159733.9	-	311
4	4544145	4544146	ENST00000514763.1	+	4544072	4544073	ENSG00000168818.5	-	72
4	6202460	6202461	ENST00000508601.1	+	6202317	6202318	ENSG00000152969.12	-	142
4	10686627	10686628	ENST00000505494.1	+	10686488	10686489	ENSG00000109684.10	-	138
4	21950638	21950639	ENST00000510705.3	+	21950421	21950422	ENSG00000185774.10	-	216
4	39640760	39640761	ENST00000533736.1	+	39640709	39640710	ENSG00000163683.7	-	50
4	39640801	39640802	ENST00000532680.1	+	39640709	39640710	ENSG00000163683.7	-	91
4	42659513	42659514	ENST00000562054.1	+	42659121	42659122	ENSG00000124406.12	-	391
4	53525573	53525574	ENST00000503051.1	+	53525501	53525502	ENSG00000109189.8	-	71
4	57253791	57253792	ENST00000602927.1	+	57253665	57253666	ENSG00000157426.9	-	125
4	66536248	66536249	ENST00000514260.1	+	66536212	66536213	ENSG00000145242.9	-	35
4	66536327	66536328	ENST00000507117.1	+	66536212	66536213	ENSG00000145242.9	-	114
4	68566998	68566999	ENST00000498917.2	+	68566896	68566897	ENSG00000033178.8	-	101
4	68567051	68567052	ENST00000506606.1	+	68566896	68566897	ENSG00000033178.8	-	154

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
4	68567113	68567114	ENST00000514109.1	+	68566896	68566897	ENSG00000033178.8	-	216
4	74124926	74124927	ENST00000502790.1	+	74124514	74124515	ENSG00000132466.13	-	411
4	85887946	85887947	ENST00000318186.3	+	85887543	85887544	ENSG00000163625.11	-	402
4	89206094	89206095	ENST00000500009.2	+	89205920	89205921	ENSG00000163644.10	-	173
4	90032651	90032652	ENST00000603357.1	+	90032548	90032549	ENSG00000138640.10	-	102
4	99580055	99580056	ENST00000569927.1	+	99579779	99579780	ENSG00000168785.3	-	275
4	100010102	100010103	ENST00000499178.2	+	100009951	100009952	ENSG00000197894.6	-	150
4	100871647	100871648	ENST00000507494.1	+	100871544	100871545	ENSG00000164032.7	-	102
4	100871680	100871681	ENST00000501976.2	+	100871544	100871545	ENSG00000164032.7	-	135
4	110224191	110224192	ENST00000500526.1	+	110223812	110223813	ENSG00000188517.10	-	378
4	120988666	120988667	ENST00000508362.1	+	120988228	120988229	ENSG00000164109.9	-	437
4	122791738	122791739	ENST00000567769.1	+	122791651	122791652	ENSG00000138686.5	-	86
4	141677682	141677683	ENST00000609937.1	+	141677273	141677274	ENSG00000109436.7	-	408
4	153457416	153457417	ENST00000604157.1	+	153457252	153457253	ENSG00000109670.9	-	163
4	153457580	153457581	ENST00000584019.1	+	153457252	153457253	ENSG00000109670.9	-	327
4	157563732	157563733	ENST00000507972.1	+	157563605	157563606	ENSG00000251283.1	-	126
4	169931581	169931582	ENST00000506933.1	+	169931425	169931426	ENSG00000145439.7	-	155
4	174451522	174451523	ENST00000507062.1	+	174451379	174451380	ENSG00000164107.7	-	142
4	174451592	174451593	ENST00000512929.1	+	174451379	174451380	ENSG00000164107.7	-	212
4	174451609	174451610	ENST00000515350.1	+	174451379	174451380	ENSG00000164107.7	-	229
4	174451610	174451611	ENST00000503198.1	+	174451379	174451380	ENSG00000164107.7	-	230
4	174451611	174451612	ENST00000515376.1	+	174451379	174451380	ENSG00000164107.7	-	231
4	174451613	174451614	ENST00000509866.1	+	174451379	174451380	ENSG00000164107.7	-	233
4	174451614	174451615	ENST00000503474.1	+	174451379	174451380	ENSG00000164107.7	-	234
4	174451625	174451626	ENST00000515345.1	+	174451379	174451380	ENSG00000164107.7	-	245
4	174451626	174451627	ENST00000512209.2	+	174451379	174451380	ENSG00000164107.7	-	246
4	174451630	174451631	ENST00000504740.1	+	174451379	174451380	ENSG00000164107.7	-	250
4	174451635	174451636	ENST00000508887.1	+	174451379	174451380	ENSG00000164107.7	-	255
4	174451658	174451659	ENST00000505817.1	+	174451379	174451380	ENSG00000164107.7	-	278
4	174451666	174451667	ENST00000514431.1	+	174451379	174451380	ENSG00000164107.7	-	286
4	185395956	185395957	ENST00000605834.1	+	185395733	185395734	ENSG00000168310.6	-	222
5	524820	524821	ENST00000515085.1	+	524446	524447	ENSG00000066230.6	-	373
5	1887446	1887447	ENST00000514569.1	+	1887349	1887350	ENSG00000113430.5	-	96
5	9546312	9546313	ENST00000508179.1	+	9546186	9546187	ENSG00000112902.7	-	125
5	9546402	9546403	ENST00000509788.1	+	9546186	9546187	ENSG00000112902.7	-	215
5	16617334	16617335	ENST00000504935.1	+	16617166	16617167	ENSG00000154153.9	-	167
5	32174577	32174578	ENST00000606994.1	+	32174455	32174456	ENSG00000113384.9	-	121
5	35938924	35938925	ENST00000503269.1	+	35938880	35938881	ENSG00000152611.7	-	43
5	43515376	43515377	ENST00000504277.1	+	43515246	43515247	ENSG00000172244.4	-	129
5	55290995	55290996	ENST00000500093.2	+	55290820	55290821	ENSG00000134352.15	-	174
5	76383262	76383263	ENST00000514640.1	+	76383147	76383148	ENSG00000132846.5	-	114
5	76383288	76383289	ENST00000514114.1	+	76383147	76383148	ENSG00000132846.5	-	140
5	83680665	83680666	ENST00000507060.1	+	83680610	83680611	ENSG00000164176.8	-	54
5	132299462	132299463	ENST00000607688.1	+	132299325	132299326	ENSG00000072364.8	-	136
5	133340838	133340839	ENST00000606089.1	+	133340823	133340824	ENSG00000213585.6	-	14
5	148442880	148442881	ENST00000515519.1	+	148442725	148442726	ENSG00000169247.7	-	154
5	148443018	148443019	ENST00000509139.1	+	148442725	148442726	ENSG00000169247.7	-	292
5	148443049	148443050	ENST00000515304.1	+	148442725	148442726	ENSG00000169247.7	-	323
5	148443066	148443067	ENST00000507318.1	+	148442725	148442726	ENSG00000169247.7	-	340
5	148443172	148443173	ENST00000507373.1	+	148442725	148442726	ENSG00000169247.7	-	446
5	176981825	176981826	ENST00000606358.1	+	176981541	176981542	ENSG00000146067.11	-	283
5	180688213	180688214	ENST00000514146.1	+	180688118	180688119	ENSG00000183718.4	-	94
5	180688223	180688224	ENST00000507434.1	+	180688118	180688119	ENSG00000183718.4	-	104
5	180688225	180688226	ENST00000509252.1	+	180688118	180688119	ENSG00000183718.4	-	106
6	2245982	2245983	ENST00000530346.1	+	2245925	2245926	ENSG00000112699.6	-	56
6	2246002	2246003	ENST00000532124.1	+	2245925	2245926	ENSG00000112699.6	-	76
6	2246005	2246006	ENST00000531092.1	+	2245925	2245926	ENSG00000112699.6	-	79
6	2246011	2246012	ENST00000456943.2	+	2245925	2245926	ENSG00000112699.6	-	85
6	2246016	2246017	ENST00000529893.1	+	2245925	2245926	ENSG00000112699.6	-	90
6	2246059	2246060	ENST00000530833.1	+	2245925	2245926	ENSG00000112699.6	-	133
6	2246075	2246076	ENST00000525811.1	+	2245925	2245926	ENSG00000112699.6	-	149
6	2246076	2246077	ENST00000534441.1	+	2245925	2245926	ENSG00000112699.6	-	150
6	2246087	2246088	ENST00000533653.1	+	2245925	2245926	ENSG00000112699.6	-	161
6	2246089	2246090	ENST00000534468.1	+	2245925	2245926	ENSG00000112699.6	-	163
6	4136306	4136307	ENST00000427049.2	+	4135830	4135831	ENSG00000198721.8	-	475
6	11044746	11044747	ENST00000607275.1	+	11044546	11044547	ENSG00000197977.3	-	199
6	11044926	11044927	ENST00000456616.1	+	11044546	11044547	ENSG00000197977.3	-	379
6	16762143	16762144	ENST00000450930.1	+	16761721	16761722	ENSG00000124788.13	-	421
6	34664722	34664723	ENST00000606496.1	+	34664635	34664636	ENSG00000196821.5	-	86
6	46459789	46459790	ENST00000415787.1	+	46459708	46459709	ENSG00000172348.10	-	80
6	52442105	52442106	ENST00000606714.1	+	52441712	52441713	ENSG00000065308.4	-	392
6	74233564	74233565	ENST00000429386.1	+	74233519	74233520	ENSG00000156508.13	-	44
6	75994749	75994750	ENST00000607221.1	+	75994683	75994684	ENSG00000112697.11	-	65
6	97731179	97731180	ENST00000457513.1	+	97731092	97731093	ENSG00000146263.7	-	86

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
6	101329347	101329348	ENST00000565695.2	+	101329247	101329248	ENSG00000112249.9	-	99
6	109703998	109703999	ENST00000563105.1	+	109703761	109703762	ENSG00000135535.10	-	236
6	135818939	135818940	ENST00000579339.1	+	135818913	135818914	ENSG00000135541.16	-	25
6	135818944	135818945	ENST00000580741.1	+	135818913	135818914	ENSG00000135541.16	-	30
6	135818992	135818993	ENST00000421378.2	+	135818913	135818914	ENSG00000135541.16	-	78
6	135819075	135819076	ENST00000579057.1	+	135818913	135818914	ENSG00000135541.16	-	161
6	135819094	135819095	ENST00000436554.1	+	135818913	135818914	ENSG00000135541.16	-	180
6	135819107	135819108	ENST00000438618.2	+	135818913	135818914	ENSG00000135541.16	-	193
6	157744996	157744997	ENST00000603032.1	+	157744632	157744633	ENSG00000215712.6	-	363
6	166796677	166796678	ENST00000568025.1	+	166796485	166796486	ENSG00000060762.14	-	191
7	560028	560029	ENST00000452622.1	+	559932	559933	ENSG00000197461.9	-	95
7	1499573	1499574	ENST00000445345.1	+	1499137	1499138	ENSG00000164877.14	-	435
7	5465401	5465402	ENST00000609130.1	+	5465044	5465045	ENSG00000182095.10	-	356
7	10980050	10980051	ENST00000604183.1	+	10979882	10979883	ENSG00000189043.5	-	167
7	20257200	20257201	ENST00000439058.1	+	20257026	20257027	ENSG00000183742.8	-	173
7	20257209	20257210	ENST00000430859.1	+	20257026	20257027	ENSG00000183742.8	-	182
7	27135743	27135744	ENST00000429611.3	+	27135614	27135615	ENSG00000105991.7	-	128
7	27135852	27135853	ENST00000425358.2	+	27135614	27135615	ENSG00000105991.7	-	237
7	27197963	27197964	ENST00000602610.1	+	27197554	27197555	ENSG00000122592.6	-	408
7	27225027	27225028	ENST00000522674.1	+	27224841	27224842	ENSG00000005073.5	-	185
7	27225153	27225154	ENST00000520395.1	+	27224841	27224842	ENSG00000005073.5	-	311
7	27240040	27240041	ENST00000521028.2	+	27239724	27239725	ENSG00000106031.6	-	315
7	27240056	27240057	ENST00000472494.1	+	27239724	27239725	ENSG00000106031.6	-	331
7	44888015	44888016	ENST00000443162.1	+	44887681	44887682	ENSG00000105968.14	-	333
7	55640864	55640865	ENST00000454777.1	+	55640680	55640681	ENSG00000154978.8	-	183
7	77045990	77045991	ENST00000608884.1	+	77045716	77045717	ENSG00000186088.11	-	273
7	79083276	79083277	ENST00000414797.1	+	79082889	79082890	ENSG00000187391.13	-	386
7	79083308	79083309	ENST00000422093.1	+	79082889	79082890	ENSG00000187391.13	-	418
7	79083326	79083327	ENST00000451809.1	+	79082889	79082890	ENSG00000187391.13	-	436
7	79083328	79083329	ENST00000448195.1	+	79082889	79082890	ENSG00000187391.13	-	438
7	79083331	79083332	ENST00000424477.1	+	79082889	79082890	ENSG00000187391.13	-	441
7	95225994	95225995	ENST00000432265.1	+	95225802	95225803	ENSG00000004799.7	-	191
7	123389122	123389123	ENST00000607957.1	+	123389120	123389121	ENSG00000106299.7	-	1
7	139877061	139877062	ENST00000566699.1	+	139876834	139876835	ENSG00000006459.6	-	226
7	150038860	150038861	ENST00000563946.1	+	150038762	150038763	ENSG00000106538.5	-	97
7	151574550	151574551	ENST00000467458.1	+	151574209	151574210	ENSG00000106617.9	-	340
7	154795158	154795159	ENST00000608317.1	+	154794793	154794794	ENSG00000157212.14	-	364
7	156803499	156803500	ENST00000480284.1	+	156803344	156803345	ENSG00000130675.10	-	154
8	12051976	12051977	ENST00000528514.1	+	12051641	12051642	ENSG00000186523.10	-	334
8	17658854	17658855	ENST00000522768.1	+	17658425	17658426	ENSG00000129422.9	-	428
8	17942536	17942537	ENST00000521775.1	+	17942493	17942494	ENSG00000104763.13	-	42
8	17942558	17942559	ENST00000517798.1	+	17942493	17942494	ENSG00000104763.13	-	64
8	23082734	23082735	ENST00000500853.1	+	23082638	23082639	ENSG00000104689.5	-	95
8	38239882	38239883	ENST00000607047.1	+	38239789	38239790	ENSG00000147548.12	-	92
8	52811885	52811886	ENST00000518942.1	+	52811734	52811735	ENSG00000168300.9	-	150
8	52812209	52812210	ENST00000423716.1	+	52811734	52811735	ENSG00000168300.9	-	474
8	56987151	56987152	ENST00000521403.1	+	56987068	56987069	ENSG00000008988.5	-	82
8	60031777	60031778	ENST00000518993.1	+	60031766	60031767	ENSG00000198846.5	-	10
8	66754987	66754988	ENST00000607622.1	+	66754556	66754557	ENSG00000205268.6	-	430
8	68256210	68256211	ENST00000607397.1	+	68255911	68255912	ENSG00000066777.4	-	298
8	80680377	80680378	ENST00000607172.1	+	80680097	80680098	ENSG00000164683.12	-	279
8	94753421	94753422	ENST00000523945.1	+	94753244	94753245	ENSG00000183808.7	-	176
8	95565950	95565951	ENST00000523011.1	+	95565756	95565757	ENSG00000164944.7	-	193
8	103251622	103251623	ENST00000520820.1	+	103251345	103251346	ENSG00000048392.7	-	276
8	122653676	122653677	ENST00000520043.1	+	122653629	122653630	ENSG00000170961.6	-	46
8	143485013	143485014	ENST00000569285.1	+	143484600	143484601	ENSG00000171045.10	-	412
8	144816310	144816311	ENST00000533004.1	+	144815970	144815971	ENSG00000180921.6	-	339
9	6645956	6645957	ENST00000413145.1	+	6645649	6645650	ENSG00000178445.8	-	306
9	10613202	10613203	ENST00000429581.2	+	10612722	10612723	ENSG00000153707.11	-	479
9	24545950	24545951	ENST00000602851.1	+	24545943	24545944	ENSG00000205442.8	-	6
9	24545996	24545997	ENST00000602614.1	+	24545943	24545944	ENSG00000205442.8	-	52
9	33402855	33402856	ENST00000450864.1	+	33402642	33402643	ENSG00000165269.8	-	212
9	38620731	38620732	ENST00000484285.2	+	38620656	38620657	ENSG00000180071.14	-	74
9	38621085	38621086	ENST00000377680.3	+	38620656	38620657	ENSG00000180071.14	-	428
9	44402460	44402461	ENST00000425309.1	+	44402426	44402427	ENSG00000212952.5	-	33
9	71155952	71155953	ENST00000413269.3	+	71155782	71155783	ENSG00000181778.4	-	169
9	72287665	72287666	ENST00000567129.1	+	72287221	72287222	ENSG00000107282.5	-	443
9	84304628	84304629	ENST00000437181.1	+	84304219	84304220	ENSG00000196781.9	-	408
9	86323233	86323234	ENST00000531661.1	+	86323117	86323118	ENSG00000135018.9	-	115
9	96717879	96717880	ENST00000454594.1	+	96717653	96717654	ENSG00000131668.9	-	225
9	96718116	96718117	ENST00000453045.1	+	96717653	96717654	ENSG00000131668.9	-	462
9	123605274	123605275	ENST00000586907.1	+	123605261	123605262	ENSG00000095261.9	-	12
9	123605378	123605379	ENST00000442982.1	+	123605261	123605262	ENSG00000095261.9	-	116
9	128003942	128003943	ENST00000468244.1	+	128003608	128003609	ENSG00000044574.7	-	333

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
9	139440664	139440665	ENST00000429224.1	+	139440313	139440314	ENSG00000148400.9	-	350
X	46404928	46404929	ENST00000421685.2	+	46404891	46404892	ENSG00000251192.3	-	36
X	46404946	46404947	ENST00000609887.1	+	46404891	46404892	ENSG00000251192.3	-	54
X	57148095	57148096	ENST00000439622.1	+	57147979	57147980	ENSG00000186787.7	-	115
X	107979770	107979771	ENST00000436013.1	+	107979650	107979651	ENSG00000133124.10	-	119
X	151922445	151922446	ENST00000370292.3	+	151922363	151922364	ENSG00000184750.11	-	81
1	948572	948573	ENST00000458555.1	-	948803	948804	ENSG00000187608.5	+	230
1	1369952	1369953	ENST00000430109.1	-	1370241	1370242	ENSG00000179403.10	+	288
1	1369867	1369868	ENST00000454562.1	-	1370241	1370242	ENSG00000179403.10	+	373
1	1369899	1369900	ENST00000417917.1	-	1370241	1370242	ENSG00000179403.10	+	341
1	1981508	1981509	ENST00000449154.1	-	1981909	1981910	ENSG00000067606.11	+	400
1	3541306	3541307	ENST00000435049.1	-	3541566	3541567	ENSG00000158109.10	+	259
1	6844902	6844903	ENST00000442889.1	-	6845384	6845385	ENSG00000171735.14	+	481
1	15735928	15735929	ENST00000427824.1	-	15736391	15736392	ENSG00000142634.8	+	462
1	19923325	19923326	ENST00000416470.1	-	19923473	19923474	ENSG00000270136.1	+	147
1	27560842	27560843	ENST00000425205.1	-	27561007	27561008	ENSG00000142784.11	+	164
1	37940011	37940012	ENST00000424989.1	-	37940153	37940154	ENSG00000163874.8	+	141
1	40723638	40723639	ENST00000567508.1	-	40723779	40723780	ENSG00000084073.4	+	140
1	43824328	43824329	ENST00000424948.1	-	43824626	43824627	ENSG00000117399.9	+	297
1	44412222	44412223	ENST00000412378.1	-	44412611	44412612	ENSG00000117408.6	+	388
1	62208095	62208096	ENST00000605725.1	-	62208149	62208150	ENSG00000132849.14	+	53
1	76189695	76189696	ENST00000433521.2	-	76190036	76190037	ENSG00000117054.9	+	340
1	84543613	84543614	ENST00000605506.1	-	84543745	84543746	ENSG00000142875.15	+	131
1	87170144	87170145	ENST00000565575.1	-	87170259	87170260	ENSG00000097033.10	+	114
1	90098452	90098453	ENST00000415584.2	-	90098631	90098632	ENSG00000171488.10	+	178
1	90098305	90098306	ENST00000526694.1	-	90098631	90098632	ENSG00000171488.10	+	325
1	90098319	90098320	ENST00000528692.1	-	90098631	90098632	ENSG00000171488.10	+	311
1	93811367	93811368	ENST00000421202.1	-	93811445	93811446	ENSG00000117505.8	+	77
1	93811359	93811360	ENST00000411670.1	-	93811445	93811446	ENSG00000117505.8	+	85
1	93811404	93811405	ENST00000452347.1	-	93811445	93811446	ENSG00000117505.8	+	40
1	93811356	93811357	ENST00000438777.1	-	93811445	93811446	ENSG00000117505.8	+	88
1	93811339	93811340	ENST00000449305.1	-	93811445	93811446	ENSG00000117505.8	+	105
1	93811402	93811403	ENST00000457387.1	-	93811445	93811446	ENSG00000117505.8	+	42
1	93811383	93811384	ENST00000445076.1	-	93811445	93811446	ENSG00000117505.8	+	61
1	95285651	95285652	ENST00000442418.1	-	95285898	95285899	ENSG00000143036.12	+	246
1	95285836	95285837	ENST00000452922.1	-	95285898	95285899	ENSG00000143036.12	+	61
1	95285774	95285775	ENST00000421997.1	-	95285898	95285899	ENSG00000143036.12	+	123
1	95285765	95285766	ENST00000414374.1	-	95285898	95285899	ENSG00000143036.12	+	132
1	95285774	95285775	ENST00000418366.2	-	95285898	95285899	ENSG00000143036.12	+	123
1	100731675	100731676	ENST00000421185.1	-	100731763	100731764	ENSG00000137996.8	+	87
1	101702083	101702084	ENST00000432195.1	-	101702444	101702445	ENSG00000170989.8	+	360
1	104068104	104068105	ENST00000444810.1	-	104068313	104068314	ENSG00000185946.11	+	208
1	104068082	104068083	ENST00000447322.2	-	104068313	104068314	ENSG00000185946.11	+	230
1	113615726	113615727	ENST00000421157.1	-	113615831	113615832	ENSG00000198799.7	+	104
1	118148332	118148333	ENST00000440801.1	-	118148556	118148557	ENSG00000183508.4	+	223
1	118148391	118148392	ENST00000425010.1	-	118148556	118148557	ENSG00000183508.4	+	164
1	118148306	118148307	ENST00000456126.1	-	118148556	118148557	ENSG00000183508.4	+	249
1	149804108	149804109	ENST00000577853.1	-	149804221	149804222	ENSG00000183941.8	+	112
1	156610795	156610796	ENST00000606343.1	-	156611182	156611183	ENSG00000132692.14	+	386
1	161736016	161736017	ENST00000431097.2	-	161736084	161736085	ENSG00000118217.5	+	67
1	162531097	162531098	ENST00000563991.1	-	162531323	162531324	ENSG00000117143.9	+	225
1	167189748	167189749	ENST00000606967.1	-	167190066	167190067	ENSG00000143190.17	+	317
1	170501140	170501141	ENST00000421020.1	-	170501270	170501271	ENSG00000120370.8	+	129
1	173837126	173837127	ENST00000449289.1	-	173837220	173837221	ENSG00000185278.10	+	93
1	173837120	173837121	ENST00000452197.1	-	173837220	173837221	ENSG00000185278.10	+	99
1	173837124	173837125	ENST00000412059.1	-	173837220	173837221	ENSG00000185278.10	+	95
1	173837126	173837127	ENST00000431268.1	-	173837220	173837221	ENSG00000185278.10	+	93
1	173836952	173836953	ENST00000454068.1	-	173837220	173837221	ENSG00000185278.10	+	267
1	173837126	173837127	ENST00000449589.1	-	173837220	173837221	ENSG00000185278.10	+	93
1	173837126	173837127	ENST00000455838.1	-	173837220	173837221	ENSG00000185278.10	+	93
1	173837125	173837126	ENST00000416952.1	-	173837220	173837221	ENSG00000185278.10	+	94
1	173837128	173837129	ENST00000456293.1	-	173837220	173837221	ENSG00000185278.10	+	91
1	173837126	173837127	ENST00000443799.1	-	173837220	173837221	ENSG00000185278.10	+	93
1	173836866	173836867	ENST00000451607.1	-	173837220	173837221	ENSG00000185278.10	+	353
1	173837124	173837125	ENST00000421068.1	-	173837220	173837221	ENSG00000185278.10	+	95
1	173836826	173836827	ENST00000432536.1	-	173837220	173837221	ENSG00000185278.10	+	393
1	173837124	173837125	ENST00000458220.1	-	173837220	173837221	ENSG00000185278.10	+	95
1	173836887	173836888	ENST00000364084.1	-	173837220	173837221	ENSG00000185278.10	+	332
1	174128290	174128291	ENST00000426899.1	-	174128548	174128549	ENSG00000152061.17	+	257
1	174128270	174128271	ENST00000454467.1	-	174128548	174128549	ENSG00000152061.17	+	277
1	174128424	174128425	ENST00000430592.1	-	174128548	174128549	ENSG00000152061.17	+	123
1	178062734	178062735	ENST00000419458.1	-	178062864	178062865	ENSG00000075391.12	+	129
1	178062706	178062707	ENST00000452867.1	-	178062864	178062865	ENSG00000075391.12	+	157
1	179923574	179923575	ENST00000567904.1	-	179923873	179923874	ENSG00000135837.11	+	298

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
1	183440880	183440881	ENST00000421703.1	-	183441351	183441352	ENSG00000116698.16	+	470
1	183440863	183440864	ENST00000432837.1	-	183441351	183441352	ENSG00000116698.16	+	487
1	184355837	184355838	ENST00000605589.1	-	184356192	184356193	ENSG00000116667.8	+	354
1	203274436	203274437	ENST00000457348.1	-	203274619	203274620	ENSG00000159388.5	+	182
1	203274386	203274387	ENST00000432511.1	-	203274619	203274620	ENSG00000159388.5	+	232
1	206223547	206223548	ENST00000425896.1	-	206223976	206223977	ENSG00000198049.5	+	428
1	209848591	209848592	ENST00000445272.1	-	209848765	209848766	ENSG00000123689.5	+	173
1	213031429	213031430	ENST00000356684.3	-	213031597	213031598	ENSG00000162769.8	+	167
1	229406749	229406750	ENST00000429227.1	-	229406822	229406823	ENSG00000168118.7	+	72
1	229406774	229406775	ENST00000436334.1	-	229406822	229406823	ENSG00000168118.7	+	47
1	231664301	231664302	ENST00000416221.1	-	231664399	231664400	ENSG00000270106.1	+	97
1	231664301	231664302	ENST00000416221.1	-	231664399	231664400	ENSG00000116918.9	+	97
1	231664000	231664001	ENST00000440665.1	-	231664399	231664400	ENSG00000270106.1	+	398
1	231664000	231664001	ENST00000440665.1	-	231664399	231664400	ENSG00000116918.9	+	398
1	231664041	231664042	ENST00000450783.1	-	231664399	231664400	ENSG00000270106.1	+	357
1	231664041	231664042	ENST00000450783.1	-	231664399	231664400	ENSG00000116918.9	+	357
1	231664301	231664302	ENST00000454631.1	-	231664399	231664400	ENSG00000270106.1	+	97
1	231664301	231664302	ENST00000454631.1	-	231664399	231664400	ENSG00000116918.9	+	97
1	231664246	231664247	ENST00000425412.1	-	231664399	231664400	ENSG00000270106.1	+	152
1	231664246	231664247	ENST00000425412.1	-	231664399	231664400	ENSG00000116918.9	+	152
10	3108192	3108193	ENST00000607898.1	-	3108525	3108526	ENSG00000067057.12	+	332
10	8095446	8095447	ENST00000355358.1	-	8095567	8095568	ENSG00000107485.11	+	120
10	8095218	8095219	ENST00000417359.1	-	8095567	8095568	ENSG00000107485.11	+	348
10	8095411	8095412	ENST00000458727.1	-	8095567	8095568	ENSG00000107485.11	+	155
10	14920684	14920685	ENST00000609399.1	-	14920819	14920820	ENSG00000152455.11	+	134
10	17685876	17685877	ENST00000563601.1	-	17686124	17686125	ENSG00000136738.10	+	247
10	28821282	28821283	ENST00000528337.1	-	28821422	28821423	ENSG00000095787.17	+	139
10	35415584	35415585	ENST00000450742.1	-	35415719	35415720	ENSG00000095794.15	+	134
10	35415587	35415588	ENST00000450106.1	-	35415719	35415720	ENSG00000095794.15	+	131
10	43633781	43633782	ENST00000609407.1	-	43633934	43633935	ENSG00000169826.6	+	152
10	50506824	50506825	ENST00000437677.1	-	50507187	50507188	ENSG00000177354.7	+	362
10	50507062	50507063	ENST00000442700.1	-	50507187	50507188	ENSG00000177354.7	+	124
10	54073887	54073888	ENST00000420193.1	-	54074056	54074057	ENSG00000107984.5	+	168
10	88516054	88516055	ENST00000608826.1	-	88516407	88516408	ENSG00000107779.7	+	352
10	93558047	93558048	ENST00000432938.1	-	93558069	93558070	ENSG00000107854.5	+	21
10	93557952	93557953	ENST00000432246.1	-	93558069	93558070	ENSG00000107854.5	+	116
10	99609554	99609555	ENST00000427379.2	-	99609996	99609997	ENSG00000155265.6	+	441
10	104403896	104403897	ENST00000607967.1	-	104404253	104404254	ENSG00000171206.9	+	356
10	106113332	106113333	ENST00000435434.1	-	106113522	106113523	ENSG00000120051.10	+	189
10	119301829	119301830	ENST00000450314.2	-	119301955	119301956	ENSG00000170370.10	+	125
10	119301844	119301845	ENST00000440007.1	-	119301955	119301956	ENSG00000170370.10	+	110
10	127408028	127408029	ENST00000527483.1	-	127408084	127408085	ENSG00000107938.13	+	55
10	127408016	127408017	ENST00000531977.1	-	127408084	127408085	ENSG00000107938.13	+	67
10	127407933	127407934	ENST00000430970.1	-	127408084	127408085	ENSG00000107938.13	+	150
11	62623217	62623218	ENST00000540725.1	-	62623518	62623519	ENSG00000168003.12	+	300
11	62623338	62623339	ENST00000537925.1	-	62623518	62623519	ENSG00000168003.12	+	179
11	62623352	62623353	ENST00000537068.1	-	62623518	62623519	ENSG00000168003.12	+	165
11	62623356	62623357	ENST00000538654.1	-	62623518	62623519	ENSG00000168003.12	+	161
11	62623385	62623386	ENST00000542112.1	-	62623518	62623519	ENSG00000168003.12	+	132
11	62623283	62623284	ENST00000541416.1	-	62623518	62623519	ENSG00000168003.12	+	234
11	62623229	62623230	ENST00000544983.1	-	62623518	62623519	ENSG00000168003.12	+	288
11	62623101	62623102	ENST00000365607.1	-	62623518	62623519	ENSG00000168003.12	+	416
11	65337743	65337744	ENST00000567594.1	-	65337901	65337902	ENSG00000173465.3	+	157
11	75525841	75525842	ENST00000531263.1	-	75526212	75526213	ENSG00000198382.4	+	370
11	76155617	76155618	ENST00000530759.1	-	76155967	76155968	ENSG00000158636.12	+	349
11	76155699	76155700	ENST00000572035.1	-	76155967	76155968	ENSG00000158636.12	+	267
11	82904641	82904642	ENST00000529031.1	-	82904781	82904782	ENSG00000137494.9	+	139
11	82904661	82904662	ENST00000529811.1	-	82904781	82904782	ENSG00000137494.9	+	119
11	82904610	82904611	ENST00000529607.1	-	82904781	82904782	ENSG00000137494.9	+	170
11	82904703	82904704	ENST00000500634.2	-	82904781	82904782	ENSG00000137494.9	+	77
11	82904704	82904705	ENST00000528083.1	-	82904781	82904782	ENSG00000137494.9	+	76
11	113185145	113185146	ENST00000526487.1	-	113185251	113185252	ENSG00000149292.12	+	105
11	113185158	113185159	ENST00000533504.1	-	113185251	113185252	ENSG00000149292.12	+	92
11	118868713	118868714	ENST00000526453.1	-	118868852	118868853	ENSG00000186166.4	+	138
11	125034504	125034505	ENST00000532316.1	-	125034583	125034584	ENSG00000165495.11	+	78
11	128556322	128556323	ENST00000572256.1	-	128556430	128556431	ENSG00000151702.12	+	107
12	6642663	6642664	ENST00000537921.1	-	6643093	6643094	ENSG00000111640.10	+	429
12	22777948	22777949	ENST00000542076.1	-	22778009	22778010	ENSG00000139163.11	+	60
12	26111510	26111511	ENST00000500276.2	-	26111962	26111963	ENSG00000123094.11	+	451
12	29301607	29301608	ENST00000553075.1	-	29302036	29302037	ENSG00000064763.6	+	428
12	54378621	54378622	ENST00000509870.1	-	54378849	54378850	ENSG00000180818.4	+	227
12	54378541	54378542	ENST00000513165.1	-	54378849	54378850	ENSG00000180818.4	+	307
12	70132347	70132348	ENST00000501387.1	-	70132461	70132462	ENSG00000127328.17	+	113
12	70132341	70132342	ENST00000501300.1	-	70132461	70132462	ENSG00000127328.17	+	119

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
12	93771491	93771492	ENST00000552835.1	-	93771659	93771660	ENSG00000173598.9	+	167
12	93771483	93771484	ENST00000549806.1	-	93771659	93771660	ENSG00000173598.9	+	175
12	93771511	93771512	ENST00000548890.1	-	93771659	93771660	ENSG00000173598.9	+	147
12	96252470	96252471	ENST00000553163.1	-	96252706	96252707	ENSG00000139343.6	+	235
12	96252575	96252576	ENST00000551893.1	-	96252706	96252707	ENSG00000139343.6	+	130
12	105500956	105500957	ENST00000550088.1	-	105501102	105501103	ENSG00000136051.9	+	145
12	107349274	107349275	ENST00000570282.1	-	107349497	107349498	ENSG00000151135.5	+	222
12	112279705	112279706	ENST00000443596.1	-	112279782	112279783	ENSG00000089022.9	+	76
12	112279709	112279710	ENST00000442119.1	-	112279782	112279783	ENSG00000089022.9	+	72
12	112279513	112279514	ENST00000590479.1	-	112279782	112279783	ENSG00000089022.9	+	268
12	112279509	112279510	ENST00000609228.1	-	112279782	112279783	ENSG00000089022.9	+	272
12	112279513	112279514	ENST00000609983.1	-	112279782	112279783	ENSG00000089022.9	+	268
12	118814079	118814080	ENST00000605329.1	-	118814185	118814186	ENSG00000111707.7	+	105
12	121077955	121077956	ENST00000544339.1	-	121078355	121078356	ENSG00000157782.5	+	399
12	130646767	130646768	ENST00000505807.2	-	130647004	130647005	ENSG00000111432.4	+	236
12	130646688	130646689	ENST00000542000.1	-	130647004	130647005	ENSG00000111432.4	+	315
12	130646800	130646801	ENST00000537095.1	-	130647004	130647005	ENSG00000111432.4	+	203
12	133613496	133613497	ENST00000592296.1	-	133613878	133613879	ENSG00000198040.6	+	381
12	133613807	133613808	ENST00000443154.3	-	133613878	133613879	ENSG00000198040.6	+	70
13	25875575	25875576	ENST00000568856.2	-	25875662	25875663	ENSG00000139496.11	+	86
13	28712329	28712330	ENST00000563843.1	-	28712643	28712644	ENSG00000152520.9	+	313
13	31506720	31506721	ENST00000411835.2	-	31506840	31506841	ENSG00000175664.5	+	119
13	31506389	31506390	ENST00000589840.1	-	31506840	31506841	ENSG00000175664.5	+	450
13	31506619	31506620	ENST00000429200.2	-	31506840	31506841	ENSG00000175664.5	+	220
13	48877796	48877797	ENST00000433480.2	-	48877887	48877888	ENSG00000139687.9	+	90
13	48877794	48877795	ENST00000436963.1	-	48877887	48877888	ENSG00000139687.9	+	92
13	50656126	50656127	ENST00000235290.3	-	50656307	50656308	ENSG00000176124.7	+	180
13	50656113	50656114	ENST00000458725.1	-	50656307	50656308	ENSG00000176124.7	+	193
13	50656127	50656128	ENST00000433070.2	-	50656307	50656308	ENSG00000176124.7	+	179
13	50656107	50656108	ENST00000443587.1	-	50656307	50656308	ENSG00000176124.7	+	199
13	50656108	50656109	ENST00000421758.1	-	50656307	50656308	ENSG00000176124.7	+	198
13	50656113	50656114	ENST00000449579.1	-	50656307	50656308	ENSG00000176124.7	+	193
13	96329047	96329048	ENST00000606011.1	-	96329393	96329394	ENSG00000102580.10	+	345
13	96329178	96329179	ENST00000499499.2	-	96329393	96329394	ENSG00000102580.10	+	214
13	99852963	99852964	ENST00000426037.2	-	99853028	99853029	ENSG00000134882.11	+	64
13	100153305	100153306	ENST00000366259.2	-	100153671	100153672	ENSG00000125304.8	+	365
14	20811565	20811566	ENST00000516869.1	-	20811741	20811742	ENSG00000129484.9	+	175
14	24422577	24422578	ENST00000399886.2	-	24422795	24422796	ENSG00000157326.14	+	217
14	45366279	45366280	ENST00000554389.1	-	45366498	45366499	ENSG00000179476.3	+	218
14	45553179	45553180	ENST00000556389.1	-	45553302	45553303	ENSG00000185246.13	+	122
14	60712370	60712371	ENST00000532515.1	-	60712470	60712471	ENSG00000100614.13	+	99
14	60712053	60712054	ENST00000553269.1	-	60712470	60712471	ENSG00000100614.13	+	416
14	60712285	60712286	ENST00000553775.1	-	60712470	60712471	ENSG00000100614.13	+	184
14	69658136	69658137	ENST00000556182.1	-	69658228	69658229	ENSG00000081177.14	+	91
14	71108014	71108015	ENST00000500016.1	-	71108504	71108505	ENSG00000133985.2	+	489
14	75894392	75894393	ENST00000558267.1	-	75894419	75894420	ENSG00000140044.8	+	26
14	103589343	103589344	ENST00000560742.1	-	103589779	103589780	ENSG00000185215.4	+	435
14	103995408	103995409	ENST00000568177.1	-	103995521	103995522	ENSG00000166166.8	+	112
15	33602859	33602860	ENST00000559457.1	-	33603163	33603164	ENSG00000198838.7	+	303
15	44829120	44829121	ENST00000313807.4	-	44829255	44829256	ENSG00000104131.8	+	134
15	44829091	44829092	ENST00000559356.1	-	44829255	44829256	ENSG00000104131.8	+	163
15	44829097	44829098	ENST00000560049.1	-	44829255	44829256	ENSG00000104131.8	+	157
15	59063172	59063173	ENST00000500929.2	-	59063391	59063392	ENSG00000128923.6	+	218
15	67813399	67813400	ENST00000559702.1	-	67813406	67813407	ENSG00000189227.4	+	6
15	67834941	67834942	ENST00000604760.1	-	67835047	67835048	ENSG00000137764.15	+	105
15	69591096	69591097	ENST00000563004.1	-	69591286	69591287	ENSG00000137819.9	+	189
15	74165706	74165707	ENST00000569137.1	-	74165949	74165950	ENSG00000167139.4	+	242
15	78556328	78556329	ENST00000559954.1	-	78556428	78556429	ENSG00000140403.8	+	99
15	91260279	91260280	ENST00000558105.1	-	91260558	91260559	ENSG00000197299.6	+	278
15	93014775	93014776	ENST00000554440.1	-	93014884	93014885	ENSG00000183643.2	+	108
15	97326541	97326542	ENST00000558722.1	-	97326619	97326620	ENSG00000185594.4	+	77
16	729776	729777	ENST00000567091.1	-	730224	730225	ENSG00000103266.6	+	447
16	729736	729737	ENST00000571933.1	-	730224	730225	ENSG00000103266.6	+	487
16	2205358	2205359	ENST00000563192.1	-	2205699	2205700	ENSG00000131653.8	+	340
16	3162489	3162490	ENST00000576943.1	-	3162561	3162562	ENSG00000122386.6	+	71
16	3179299	3179300	ENST00000570901.1	-	3179778	3179779	ENSG00000085644.9	+	478
16	3179554	3179555	ENST00000573414.1	-	3179778	3179779	ENSG00000085644.9	+	223
16	25122799	25122800	ENST00000563962.1	-	25123050	25123051	ENSG00000205629.7	+	250
16	25122734	25122735	ENST00000563176.1	-	25123050	25123051	ENSG00000205629.7	+	315
16	25122875	25122876	ENST00000569920.1	-	25123050	25123051	ENSG00000205629.7	+	174
16	28303384	28303385	ENST00000501520.1	-	28303840	28303841	ENSG00000188322.4	+	455
16	56225005	56225006	ENST00000501259.1	-	56225302	56225303	ENSG00000087258.9	+	296
16	56763879	56763880	ENST00000561663.1	-	56764017	56764018	ENSG00000102900.8	+	137
16	57126214	57126215	ENST00000565829.1	-	57126449	57126450	ENSG00000140848.12	+	234

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
16	67596211	67596212	ENST00000562846.1	-	67596310	67596311	ENSG00000102974.10	+	98
16	74330648	74330649	ENST00000569389.1	-	74330673	74330674	ENSG00000103035.6	+	24
16	74330659	74330660	ENST00000562888.1	-	74330673	74330674	ENSG00000103035.6	+	13
16	83841438	83841439	ENST00000561599.1	-	83841448	83841449	ENSG00000230989.2	+	9
16	89284060	89284061	ENST00000570267.1	-	89284118	89284119	ENSG00000170100.9	+	57
17	259821	259822	ENST00000599026.1	-	260118	260119	ENSG00000187624.7	+	296
17	1619503	1619504	ENST00000334146.3	-	1619817	1619818	ENSG00000167716.14	+	313
17	1619544	1619545	ENST00000574306.1	-	1619817	1619818	ENSG00000167716.14	+	272
17	1619505	1619506	ENST00000576749.1	-	1619817	1619818	ENSG00000167716.14	+	311
17	1619490	1619491	ENST00000570416.1	-	1619817	1619818	ENSG00000167716.14	+	326
17	1619634	1619635	ENST00000571595.1	-	1619817	1619818	ENSG00000167716.14	+	182
17	1619504	1619505	ENST00000571091.1	-	1619817	1619818	ENSG00000167716.14	+	312
17	1619503	1619504	ENST00000576489.1	-	1619817	1619818	ENSG00000167716.14	+	313
17	1619501	1619502	ENST00000575626.1	-	1619817	1619818	ENSG00000167716.14	+	315
17	1619503	1619504	ENST00000608245.1	-	1619817	1619818	ENSG00000167716.14	+	313
17	1619501	1619502	ENST00000610106.1	-	1619817	1619818	ENSG00000167716.14	+	315
17	1619504	1619505	ENST00000609990.1	-	1619817	1619818	ENSG00000167716.14	+	312
17	1619501	1619502	ENST00000608198.1	-	1619817	1619818	ENSG00000167716.14	+	315
17	1619502	1619503	ENST00000609442.1	-	1619817	1619818	ENSG00000167716.14	+	314
17	1619503	1619504	ENST00000608913.1	-	1619817	1619818	ENSG00000167716.14	+	313
17	1619540	1619541	ENST00000573075.1	-	1619817	1619818	ENSG00000167716.14	+	276
17	1619503	1619504	ENST00000574016.1	-	1619817	1619818	ENSG00000167716.14	+	313
17	4710307	4710308	ENST00000571067.1	-	4710391	4710392	ENSG00000129219.9	+	83
17	4981407	4981408	ENST00000574352.1	-	4981543	4981544	ENSG00000180787.5	+	135
17	7486835	7486836	ENST00000573187.1	-	7486847	7486848	ENSG00000129255.10	+	11
17	13972811	13972812	ENST00000602743.1	-	13972813	13972814	ENSG00000006695.6	+	1
17	13972774	13972775	ENST00000449363.1	-	13972813	13972814	ENSG00000006695.6	+	38
17	13972795	13972796	ENST00000602539.1	-	13972813	13972814	ENSG00000006695.6	+	17
17	19912546	19912547	ENST00000564549.1	-	19912657	19912658	ENSG00000128487.12	+	110
17	33569981	33569982	ENST00000590478.1	-	33570055	33570056	ENSG00000166750.5	+	73
17	35293959	35293960	ENST00000528383.1	-	35294084	35294085	ENSG00000132130.7	+	124
17	35293956	35293957	ENST00000532387.2	-	35294084	35294085	ENSG00000132130.7	+	127
17	35293950	35293951	ENST00000529264.1	-	35294084	35294085	ENSG00000132130.7	+	133
17	35293920	35293921	ENST00000525111.1	-	35294084	35294085	ENSG00000132130.7	+	163
17	41322419	41322420	ENST00000590740.1	-	41322498	41322499	ENSG00000188554.9	+	78
17	42385760	42385761	ENST00000586388.1	-	42385781	42385782	ENSG00000108309.8	+	20
17	43922122	43922123	ENST00000581125.1	-	43922256	43922257	ENSG00000185294.5	+	133
17	45000399	45000400	ENST00000572349.1	-	45000483	45000484	ENSG00000108433.11	+	83
17	45726781	45726782	ENST00000580045.1	-	45726842	45726843	ENSG00000108424.5	+	60
17	45973177	45973178	ENST00000582787.1	-	45973516	45973517	ENSG00000167182.11	+	338
17	45973134	45973135	ENST00000577279.1	-	45973516	45973517	ENSG00000167182.11	+	381
17	45973157	45973158	ENST00000580459.1	-	45973516	45973517	ENSG00000167182.11	+	358
17	46125411	46125412	ENST00000578660.1	-	46125691	46125692	ENSG00000082641.11	+	279
17	46125434	46125435	ENST00000584428.1	-	46125691	46125692	ENSG00000082641.11	+	256
17	48133102	48133103	ENST00000499842.1	-	48133332	48133333	ENSG00000005884.13	+	229
17	48585688	48585689	ENST00000502300.1	-	48585745	48585746	ENSG00000136449.9	+	56
17	55162382	55162383	ENST00000576871.1	-	55162453	55162454	ENSG00000121057.8	+	70
17	55162384	55162385	ENST00000576313.1	-	55162453	55162454	ENSG00000121057.8	+	68
17	56160564	56160565	ENST00000584805.1	-	56160776	56160777	ENSG00000264364.2	+	211
17	59476966	59476967	ENST00000590421.1	-	59477257	59477258	ENSG00000121068.9	+	290
17	59476948	59476949	ENST00000591313.1	-	59477257	59477258	ENSG00000121068.9	+	308
17	70116933	70116934	ENST00000533232.1	-	70117161	70117162	ENSG00000125398.5	+	227
17	72209480	72209481	ENST00000532794.1	-	72209653	72209654	ENSG00000141540.6	+	172
17	72209446	72209447	ENST00000531617.1	-	72209653	72209654	ENSG00000141540.6	+	206
17	79008500	79008501	ENST00000573167.1	-	79008948	79008949	ENSG00000175866.11	+	447
17	80415574	80415575	ENST00000578344.1	-	80416056	80416057	ENSG00000141562.13	+	481
17	80674130	80674131	ENST00000574471.1	-	80674559	80674560	ENSG00000141560.10	+	428
18	904481	904482	ENST00000582921.1	-	904944	904945	ENSG00000141433.8	+	462
18	3247083	3247084	ENST00000609924.1	-	3247479	3247480	ENSG00000101608.8	+	395
18	9334438	9334439	ENST00000584509.1	-	9334765	9334766	ENSG00000128791.7	+	326
18	19748928	19748929	ENST00000583490.1	-	19749404	19749405	ENSG00000141448.4	+	475
18	31158150	31158151	ENST00000591558.1	-	31158579	31158580	ENSG00000141431.5	+	428
18	33767410	33767411	ENST00000568654.1	-	33767482	33767483	ENSG00000075643.5	+	71
18	42259681	42259682	ENST00000592638.1	-	42260138	42260139	ENSG00000152217.12	+	456
18	54814228	54814229	ENST00000590942.1	-	54814293	54814294	ENSG00000228075.4	+	64
18	72265059	72265060	ENST00000580048.1	-	72265106	72265107	ENSG00000215421.5	+	46
18	72265035	72265036	ENST00000585279.1	-	72265106	72265107	ENSG00000215421.5	+	70
18	72264804	72264805	ENST00000577806.1	-	72265106	72265107	ENSG00000215421.5	+	301
18	77439744	77439745	ENST00000317008.4	-	77439801	77439802	ENSG00000060069.12	+	56
19	9945688	9945689	ENST00000591174.1	-	9945933	9945934	ENSG00000127445.9	+	244
19	10764519	10764520	ENST00000591501.1	-	10764937	10764938	ENSG00000129351.13	+	417
19	16435324	16435325	ENST00000588799.1	-	16435628	16435629	ENSG00000127528.5	+	303
19	33182655	33182656	ENST00000592431.1	-	33182867	33182868	ENSG00000213965.3	+	211
19	33210500	33210501	ENST00000587554.1	-	33210659	33210660	ENSG00000173809.11	+	158

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
19	35417715	35417716	ENST00000604333.1	-	35417807	35417808	ENSG00000168661.10	+	91
19	36103567	36103568	ENST00000589603.1	-	36103646	36103647	ENSG00000249115.4	+	78
19	37178350	37178351	ENST00000433232.1	-	37178514	37178515	ENSG00000189042.9	+	163
19	37178311	37178312	ENST00000425254.2	-	37178514	37178515	ENSG00000189042.9	+	202
19	37178338	37178339	ENST00000590952.1	-	37178514	37178515	ENSG00000189042.9	+	175
19	37997525	37997526	ENST00000588845.1	-	37997841	37997842	ENSG00000188227.8	+	315
19	44529413	44529414	ENST00000592583.1	-	44529494	44529495	ENSG00000159885.9	+	80
19	52901009	52901010	ENST00000601562.1	-	52901102	52901103	ENSG00000167555.9	+	92
19	52901018	52901019	ENST00000596746.1	-	52901102	52901103	ENSG00000167555.9	+	83
19	52900921	52900922	ENST00000598892.1	-	52901102	52901103	ENSG00000167555.9	+	180
19	54959779	54959780	ENST00000416022.1	-	54960065	54960066	ENSG00000167615.12	+	285
19	57049943	57049944	ENST00000590613.1	-	57050317	57050318	ENSG00000196867.3	+	373
2	8818940	8818941	ENST00000433340.1	-	8818975	8818976	ENSG00000115738.5	+	34
2	8818773	8818774	ENST00000433592.1	-	8818975	8818976	ENSG00000115738.5	+	201
2	11272895	11272896	ENST00000447433.1	-	11273179	11273180	ENSG00000150873.7	+	283
2	11272946	11272947	ENST00000590373.1	-	11273179	11273180	ENSG00000150873.7	+	232
2	29320390	29320391	ENST00000446073.1	-	29320571	29320572	ENSG00000115295.15	+	180
2	32390825	32390826	ENST00000608489.1	-	32390933	32390934	ENSG00000152683.10	+	107
2	45168632	45168633	ENST00000456467.1	-	45168902	45168903	ENSG00000138083.3	+	269
2	47572127	47572128	ENST00000448713.1	-	47572297	47572298	ENSG00000119888.6	+	169
2	47572212	47572213	ENST00000441997.1	-	47572297	47572298	ENSG00000119888.6	+	84
2	47572104	47572105	ENST00000419035.1	-	47572297	47572298	ENSG00000119888.6	+	192
2	47572127	47572128	ENST00000450550.1	-	47572297	47572298	ENSG00000119888.6	+	169
2	48667735	48667736	ENST00000609028.1	-	48667737	48667738	ENSG00000162869.11	+	1
2	61108448	61108449	ENST00000439412.1	-	61108656	61108657	ENSG00000162924.9	+	207
2	64680931	64680932	ENST00000441630.1	-	64681103	64681104	ENSG00000119862.8	+	171
2	64751226	64751227	ENST00000561559.1	-	64751465	64751466	ENSG00000119844.10	+	238
2	70314451	70314452	ENST00000457076.1	-	70314585	70314586	ENSG00000169564.5	+	133
2	70314474	70314475	ENST00000415222.1	-	70314585	70314586	ENSG00000169564.5	+	110
2	70314137	70314138	ENST00000599673.1	-	70314585	70314586	ENSG00000169564.5	+	447
2	70314446	70314447	ENST00000434781.1	-	70314585	70314586	ENSG00000169564.5	+	138
2	70314131	70314132	ENST00000366234.3	-	70314585	70314586	ENSG00000169564.5	+	453
2	70314474	70314475	ENST00000425333.1	-	70314585	70314586	ENSG00000169564.5	+	110
2	70314146	70314147	ENST00000444410.1	-	70314585	70314586	ENSG00000169564.5	+	438
2	70314435	70314436	ENST00000458698.2	-	70314585	70314586	ENSG00000169564.5	+	149
2	70314552	70314553	ENST00000413791.1	-	70314585	70314586	ENSG00000169564.5	+	32
2	70314137	70314138	ENST00000596573.1	-	70314585	70314586	ENSG00000169564.5	+	447
2	70314566	70314567	ENST00000594548.1	-	70314585	70314586	ENSG00000169564.5	+	18
2	108443346	108443347	ENST00000457647.2	-	108443388	108443389	ENSG00000196862.8	+	41
2	108442910	108442911	ENST00000609354.1	-	108443388	108443389	ENSG00000196862.8	+	477
2	108443291	108443292	ENST00000594764.1	-	108443388	108443389	ENSG00000196862.8	+	96
2	108443325	108443326	ENST00000593452.1	-	108443388	108443389	ENSG00000196862.8	+	62
2	108442978	108442979	ENST00000609972.1	-	108443388	108443389	ENSG00000196862.8	+	409
2	109150637	109150638	ENST00000440975.1	-	109150857	109150858	ENSG00000169756.12	+	219
2	109745385	109745386	ENST00000567491.1	-	109745804	109745805	ENSG00000172985.8	+	418
2	113403266	113403267	ENST00000457336.1	-	113403434	113403435	ENSG00000144136.6	+	167
2	114647326	114647327	ENST00000602760.1	-	114647537	114647538	ENSG00000115091.7	+	210
2	124782749	124782750	ENST00000438816.1	-	124782864	124782865	ENSG00000155052.14	+	114
2	139259267	139259268	ENST00000431985.1	-	139259371	139259372	ENSG00000144228.4	+	103
2	139259243	139259244	ENST00000414911.1	-	139259371	139259372	ENSG00000144228.4	+	127
2	155554328	155554329	ENST00000443901.1	-	155554811	155554812	ENSG00000162989.3	+	482
2	160568940	160568941	ENST00000453016.1	-	160569000	160569001	ENSG00000136536.10	+	59
2	160568528	160568529	ENST00000418770.1	-	160569000	160569001	ENSG00000136536.10	+	471
2	160568945	160568946	ENST00000607836.1	-	160569000	160569001	ENSG00000136536.10	+	54
2	177053267	177053268	ENST00000417086.1	-	177053307	177053308	ENSG00000128645.11	+	39
2	177052996	177052997	ENST00000436126.1	-	177053307	177053308	ENSG00000128645.11	+	310
2	177053206	177053207	ENST00000425005.1	-	177053307	177053308	ENSG00000128645.11	+	100
2	177053258	177053259	ENST00000452365.1	-	177053307	177053308	ENSG00000128645.11	+	48
2	182756389	182756390	ENST00000567327.1	-	182756560	182756561	ENSG00000138434.12	+	170
2	200775804	200775805	ENST00000457577.3	-	200775979	200775980	ENSG00000178074.5	+	174
2	200775677	200775678	ENST00000417006.1	-	200775979	200775980	ENSG00000178074.5	+	301
2	200775777	200775778	ENST00000598349.1	-	200775979	200775980	ENSG00000178074.5	+	201
2	200775860	200775861	ENST00000599977.1	-	200775979	200775980	ENSG00000178074.5	+	118
2	200775881	200775882	ENST00000596619.1	-	200775979	200775980	ENSG00000178074.5	+	97
2	200775722	200775723	ENST00000608040.1	-	200775979	200775980	ENSG00000178074.5	+	256
2	200775812	200775813	ENST00000608498.1	-	200775979	200775980	ENSG00000178074.5	+	166
2	200775813	200775814	ENST00000608419.1	-	200775979	200775980	ENSG00000178074.5	+	165
2	214148928	214148929	ENST00000360083.3	-	214149113	214149114	ENSG00000144451.14	+	184
2	231860745	231860746	ENST00000414876.1	-	231860836	231860837	ENSG00000173699.11	+	90
2	231860746	231860747	ENST00000446741.1	-	231860836	231860837	ENSG00000173699.11	+	89
2	231860722	231860723	ENST00000434094.1	-	231860836	231860837	ENSG00000173699.11	+	113
2	231860726	231860727	ENST00000418330.1	-	231860836	231860837	ENSG00000173699.11	+	109
2	231860746	231860747	ENST00000426904.1	-	231860836	231860837	ENSG00000173699.11	+	89
2	231860595	231860596	ENST00000441063.1	-	231860836	231860837	ENSG00000173699.11	+	240

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
2	231860743	231860744	ENST00000457803.1	-	231860836	231860837	ENSG00000173699.11	+	92
2	241526115	241526116	ENST00000567819.1	-	241526133	241526134	ENSG00000142330.15	+	17
20	305877	305878	ENST00000414676.1	-	306207	306208	ENSG00000177732.6	+	329
20	31804872	31804873	ENST00000419613.1	-	31805116	31805117	ENSG00000131059.7	+	243
20	35201559	35201560	ENST00000559455.1	-	35201891	35201892	ENSG00000118707.5	+	331
20	35201793	35201794	ENST00000559804.1	-	35201891	35201892	ENSG00000118707.5	+	97
20	37590753	37590754	ENST00000570096.1	-	37590942	37590943	ENSG00000101452.10	+	188
20	43595042	43595043	ENST00000434401.1	-	43595115	43595116	ENSG00000101109.7	+	72
20	47662580	47662581	ENST00000417781.1	-	47662849	47662850	ENSG00000124207.12	+	268
21	33031812	33031813	ENST00000449339.1	-	33031935	33031936	ENSG00000142168.10	+	122
21	45875166	45875167	ENST00000426578.1	-	45875369	45875370	ENSG00000160233.6	+	202
22	18560562	18560563	ENST00000426483.1	-	18560689	18560690	ENSG00000215193.8	+	126
22	25960428	25960429	ENST00000412773.1	-	25960816	25960817	ENSG00000100077.10	+	387
22	30115847	30115848	ENST00000416352.1	-	30116073	30116074	ENSG00000100314.3	+	225
22	30115736	30115737	ENST00000451180.1	-	30116073	30116074	ENSG00000100314.3	+	336
22	30115702	30115703	ENST00000420180.1	-	30116073	30116074	ENSG00000100314.3	+	370
22	39077791	39077792	ENST00000412067.1	-	39077953	39077954	ENSG00000100216.4	+	161
22	45559661	45559662	ENST00000426282.2	-	45559722	45559723	ENSG00000093000.14	+	60
22	45559539	45559540	ENST00000432502.1	-	45559722	45559723	ENSG00000093000.14	+	182
22	47158459	47158460	ENST00000564152.1	-	47158518	47158519	ENSG00000054611.9	+	58
22	51176566	51176567	ENST00000449652.1	-	51176624	51176625	ENSG00000100312.6	+	57
3	4534846	4534847	ENST00000412804.1	-	4535032	4535033	ENSG00000150995.13	+	185
3	8543283	8543284	ENST00000446281.1	-	8543393	8543394	ENSG00000071282.7	+	109
3	8543331	8543332	ENST00000452802.1	-	8543393	8543394	ENSG00000071282.7	+	61
3	8543340	8543341	ENST00000420095.1	-	8543393	8543394	ENSG00000071282.7	+	52
3	8543314	8543315	ENST00000455811.2	-	8543393	8543394	ENSG00000071282.7	+	78
3	9439176	9439177	ENST00000522525.1	-	9439299	9439300	ENSG00000168137.11	+	122
3	9439178	9439179	ENST00000520447.1	-	9439299	9439300	ENSG00000168137.11	+	120
3	9439112	9439113	ENST00000467069.2	-	9439299	9439300	ENSG00000168137.11	+	186
3	9439122	9439123	ENST00000494680.2	-	9439299	9439300	ENSG00000168137.11	+	176
3	9439157	9439158	ENST00000481221.2	-	9439299	9439300	ENSG00000168137.11	+	141
3	9439187	9439188	ENST00000469846.2	-	9439299	9439300	ENSG00000168137.11	+	111
3	9439179	9439180	ENST00000518437.1	-	9439299	9439300	ENSG00000168137.11	+	119
3	14989013	14989014	ENST00000424349.1	-	14989091	14989092	ENSG00000177463.11	+	77
3	14989011	14989012	ENST00000440079.1	-	14989091	14989092	ENSG00000177463.11	+	79
3	23244023	23244024	ENST00000452251.1	-	23244511	23244512	ENSG00000182247.5	+	487
3	23244033	23244034	ENST00000421375.1	-	23244511	23244512	ENSG00000182247.5	+	477
3	23244068	23244069	ENST00000430018.1	-	23244511	23244512	ENSG00000182247.5	+	442
3	26664180	26664181	ENST00000435884.1	-	26664297	26664298	ENSG00000179796.7	+	116
3	32280069	32280070	ENST00000565519.1	-	32280171	32280172	ENSG00000170293.4	+	101
3	47422488	47422489	ENST00000568593.1	-	47422501	47422502	ENSG00000076201.10	+	12
3	49591798	49591799	ENST00000421598.1	-	49591922	49591923	ENSG00000164061.4	+	123
3	49591772	49591773	ENST00000433882.1	-	49591922	49591923	ENSG00000164061.4	+	149
3	110788805	110788806	ENST00000467426.1	-	110788918	110788919	ENSG00000177707.6	+	112
3	119217027	119217028	ENST00000609598.1	-	119217379	119217380	ENSG00000113845.5	+	351
3	126113693	126113694	ENST00000506660.1	-	126113782	126113783	ENSG00000163885.7	+	88
3	126113430	126113431	ENST00000505467.1	-	126113782	126113783	ENSG00000163885.7	+	351
3	152879563	152879564	ENST00000487827.1	-	152880029	152880030	ENSG00000181467.2	+	465
3	156390658	156390659	ENST00000463449.1	-	156391024	156391025	ENSG00000163659.8	+	365
3	158288695	158288696	ENST00000479233.1	-	158288952	158288953	ENSG00000178053.13	+	256
3	159733647	159733648	ENST00000462431.1	-	159733811	159733812	ENSG00000242107.1	+	163
3	159943085	159943086	ENST00000486168.1	-	159943423	159943424	ENSG00000180044.3	+	337
3	160472929	160472930	ENST00000566372.1	-	160473390	160473391	ENSG00000163590.9	+	460
3	169684027	169684028	ENST00000487580.1	-	169684423	169684424	ENSG00000008952.12	+	395
3	169684013	169684014	ENST00000483289.2	-	169684423	169684424	ENSG00000008952.12	+	409
3	178865760	178865761	ENST00000435560.1	-	178865902	178865903	ENSG00000121879.3	+	141
3	180319722	180319723	ENST00000472596.1	-	180319918	180319919	ENSG00000163728.6	+	195
3	182511181	182511182	ENST00000488882.1	-	182511288	182511289	ENSG00000058063.11	+	106
4	3076240	3076241	ENST00000503893.1	-	3076408	3076409	ENSG00000197386.6	+	167
4	56262008	56262009	ENST00000592823.1	-	56262124	56262125	ENSG00000134851.8	+	115
4	56262008	56262009	ENST00000599135.1	-	56262124	56262125	ENSG00000134851.8	+	115
4	56262003	56262004	ENST00000601433.1	-	56262124	56262125	ENSG00000134851.8	+	120
4	56262013	56262014	ENST00000608136.1	-	56262124	56262125	ENSG00000134851.8	+	110
4	56261996	56261997	ENST00000598819.1	-	56262124	56262125	ENSG00000134851.8	+	127
4	79697126	79697127	ENST00000564925.1	-	79697496	79697497	ENSG00000138756.13	+	369
4	95128706	95128707	ENST00000501965.2	-	95128762	95128763	ENSG00000163104.13	+	55
4	95678683	95678684	ENST00000510795.1	-	95679119	95679120	ENSG00000138696.6	+	435
4	103422475	103422476	ENST00000563833.1	-	103422486	103422487	ENSG00000109320.7	+	10
4	106473511	106473512	ENST00000514879.1	-	106473777	106473778	ENSG00000236699.4	+	265
4	109541552	109541553	ENST00000507248.1	-	109541722	109541723	ENSG00000109475.12	+	169
4	109541584	109541585	ENST00000506795.1	-	109541722	109541723	ENSG00000109475.12	+	137
4	109541550	109541551	ENST00000509984.1	-	109541722	109541723	ENSG00000109475.12	+	171
4	109541615	109541616	ENST00000510212.1	-	109541722	109541723	ENSG00000109475.12	+	106
4	110354919	110354920	ENST00000510971.1	-	110354928	110354929	ENSG00000138802.7	+	8

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
4	110736565	110736566	ENST00000609440.1	-	110736666	110736667	ENSG00000109534.12	+	100
4	113152751	113152752	ENST00000562919.1	-	113152893	113152894	ENSG00000138660.7	+	141
4	134070267	134070268	ENST00000505289.1	-	134070470	134070471	ENSG00000138650.7	+	202
4	134070270	134070271	ENST00000509715.1	-	134070470	134070471	ENSG00000138650.7	+	199
4	144106013	144106014	ENST00000507826.1	-	144106070	144106071	ENSG00000170185.5	+	56
4	144105981	144105982	ENST00000507486.1	-	144106070	144106071	ENSG00000170185.5	+	88
4	145567138	145567139	ENST00000508269.1	-	145567173	145567174	ENSG00000164161.5	+	34
4	145567063	145567064	ENST00000503066.1	-	145567173	145567174	ENSG00000164161.5	+	109
4	148538395	148538396	ENST00000508072.1	-	148538534	148538535	ENSG00000164168.3	+	138
4	152329986	152329987	ENST00000508847.1	-	152330368	152330369	ENSG00000164142.11	+	381
4	156129582	156129583	ENST00000511017.1	-	156129781	156129782	ENSG00000185149.5	+	198
4	174290965	174290966	ENST00000608794.1	-	174291120	174291121	ENSG00000164105.3	+	154
4	183064928	183064929	ENST00000505873.1	-	183065140	183065141	ENSG00000218336.3	+	211
4	183065040	183065041	ENST00000511052.1	-	183065140	183065141	ENSG00000218336.3	+	99
4	184020351	184020352	ENST00000578387.1	-	184020446	184020447	ENSG00000151718.11	+	94
4	184425646	184425647	ENST00000457303.3	-	184426147	184426148	ENSG00000168556.5	+	500
5	271630	271631	ENST00000512642.1	-	271736	271737	ENSG00000249915.3	+	105
5	612324	612325	ENST00000506629.1	-	612387	612388	ENSG00000112877.6	+	62
5	5140166	5140167	ENST00000512155.1	-	5140443	5140444	ENSG00000145536.11	+	276
5	10353712	10353713	ENST00000561606.1	-	10353815	10353816	ENSG00000145495.10	+	102
5	14664712	14664713	ENST00000563101.1	-	14664773	14664774	ENSG00000154124.4	+	60
5	14664681	14664682	ENST00000567048.1	-	14664773	14664774	ENSG00000154124.4	+	91
5	14664383	14664384	ENST00000564167.1	-	14664773	14664774	ENSG00000154124.4	+	389
5	33440724	33440725	ENST00000507251.1	-	33440802	33440803	ENSG00000113407.9	+	77
5	38845869	38845870	ENST00000513480.1	-	38845960	38845961	ENSG00000145623.8	+	90
5	38845930	38845931	ENST00000512519.1	-	38845960	38845961	ENSG00000145623.8	+	29
5	44808878	44808879	ENST00000503452.1	-	44809027	44809028	ENSG00000112996.5	+	148
5	44808843	44808844	ENST00000514597.1	-	44809027	44809028	ENSG00000112996.5	+	183
5	44808842	44808843	ENST00000505302.1	-	44809027	44809028	ENSG00000112996.5	+	184
5	44808860	44808861	ENST00000508945.1	-	44809027	44809028	ENSG00000112996.5	+	166
5	44808767	44808768	ENST00000508123.1	-	44809027	44809028	ENSG00000112996.5	+	259
5	44808860	44808861	ENST00000503179.1	-	44809027	44809028	ENSG00000112996.5	+	166
5	44808827	44808828	ENST00000505401.1	-	44809027	44809028	ENSG00000112996.5	+	199
5	44808837	44808838	ENST00000505637.1	-	44809027	44809028	ENSG00000112996.5	+	189
5	72251525	72251526	ENST00000606587.1	-	72251808	72251809	ENSG00000157107.9	+	282
5	72794117	72794118	ENST00000607001.1	-	72794233	72794234	ENSG00000145741.11	+	115
5	77656216	77656217	ENST00000513755.1	-	77656339	77656340	ENSG00000085365.13	+	122
5	77656330	77656331	ENST00000421004.3	-	77656339	77656340	ENSG00000085365.13	+	8
5	79783771	79783772	ENST00000508000.1	-	79783788	79783789	ENSG00000152380.5	+	16
5	99870889	99870890	ENST00000499025.1	-	99871009	99871010	ENSG00000174132.8	+	119
5	99870965	99870966	ENST00000504833.1	-	99871009	99871010	ENSG00000174132.8	+	43
5	118373362	118373363	ENST00000506486.1	-	118373467	118373468	ENSG00000172869.10	+	104
5	126112177	126112178	ENST00000509185.2	-	126112315	126112316	ENSG00000113368.7	+	137
5	131705339	131705340	ENST00000457998.2	-	131705444	131705445	ENSG00000197375.8	+	104
5	139487227	139487228	ENST00000499203.2	-	139487362	139487363	ENSG00000185129.4	+	134
5	139487050	139487051	ENST00000522747.1	-	139487362	139487363	ENSG00000185129.4	+	311
5	139487233	139487234	ENST00000521563.1	-	139487362	139487363	ENSG00000185129.4	+	128
5	146614421	146614422	ENST00000504297.1	-	146614526	146614527	ENSG00000169302.10	+	104
5	153825409	153825410	ENST00000501280.3	-	153825517	153825518	ENSG00000164576.7	+	107
5	153825381	153825382	ENST00000522312.1	-	153825517	153825518	ENSG00000164576.7	+	135
5	162864298	162864299	ENST00000458002.2	-	162864575	162864576	ENSG00000113328.14	+	276
5	162864331	162864332	ENST00000503504.1	-	162864575	162864576	ENSG00000113328.14	+	243
6	4021448	4021449	ENST00000415144.1	-	4021501	4021502	ENSG00000112739.12	+	52
6	7541570	7541571	ENST00000561592.1	-	7541808	7541809	ENSG00000096696.9	+	237
6	10747801	10747802	ENST00000606522.1	-	10747992	10747993	ENSG00000137210.9	+	190
6	10747895	10747896	ENST00000606652.1	-	10747992	10747993	ENSG00000137210.9	+	96
6	30026523	30026524	ENST00000422224.1	-	30026676	30026677	ENSG00000066379.10	+	152
6	30293910	30293911	ENST00000453558.1	-	30294256	30294257	ENSG00000204599.10	+	345
6	30294180	30294181	ENST00000602550.1	-	30294256	30294257	ENSG00000204599.10	+	75
6	30294163	30294164	ENST00000438412.1	-	30294256	30294257	ENSG00000204599.10	+	92
6	30294110	30294111	ENST00000444126.1	-	30294256	30294257	ENSG00000204599.10	+	145
6	30294123	30294124	ENST00000449544.1	-	30294256	30294257	ENSG00000204599.10	+	132
6	30294152	30294153	ENST00000454129.1	-	30294256	30294257	ENSG00000204599.10	+	103
6	30294162	30294163	ENST00000454269.1	-	30294256	30294257	ENSG00000204599.10	+	93
6	30294139	30294140	ENST00000602290.1	-	30294256	30294257	ENSG00000204599.10	+	116
6	30293913	30293914	ENST00000602498.1	-	30294256	30294257	ENSG00000204599.10	+	342
6	35704723	35704724	ENST00000452048.1	-	35704809	35704810	ENSG00000157343.4	+	85
6	37786993	37786994	ENST00000415890.1	-	37787275	37787276	ENSG00000156639.7	+	281
6	38682992	38682993	ENST00000439844.2	-	38683117	38683118	ENSG00000124721.13	+	124
6	41513980	41513981	ENST00000440194.1	-	41514164	41514165	ENSG00000137166.10	+	183
6	41513913	41513914	ENST00000414386.1	-	41514164	41514165	ENSG00000137166.10	+	250
6	41513811	41513812	ENST00000439386.1	-	41514164	41514165	ENSG00000137166.10	+	352
6	46097327	46097328	ENST00000444038.2	-	46097730	46097731	ENSG00000001561.6	+	402
6	46097442	46097443	ENST00000437249.2	-	46097730	46097731	ENSG00000001561.6	+	287

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
6	47445307	47445308	ENST00000604014.1	-	47445525	47445526	ENSG00000198087.7	+	217
6	53659071	53659072	ENST00000429053.1	-	53659295	53659296	ENSG00000137269.10	+	223
6	69344853	69344854	ENST00000603261.1	-	69345259	69345260	ENSG00000135298.9	+	405
6	69344918	69344919	ENST00000604392.1	-	69345259	69345260	ENSG00000135298.9	+	340
6	89790385	89790386	ENST00000606729.1	-	89790470	89790471	ENSG00000146278.10	+	84
6	96025325	96025326	ENST00000564541.1	-	96025419	96025420	ENSG00000172469.10	+	93
6	134210119	134210120	ENST00000606544.1	-	134210276	134210277	ENSG00000118526.6	+	156
6	134210143	134210144	ENST00000607641.1	-	134210276	134210277	ENSG00000118526.6	+	132
6	146920066	146920067	ENST00000419168.2	-	146920101	146920102	ENSG00000118492.12	+	34
6	163834991	163834992	ENST00000604200.1	-	163835032	163835033	ENSG00000112531.12	+	40
6	167412552	167412553	ENST00000444102.1	-	167412670	167412671	ENSG00000213066.7	+	117
6	168227388	168227389	ENST00000359760.5	-	168227602	168227603	ENSG00000130396.16	+	213
6	168227134	168227135	ENST00000414943.1	-	168227602	168227603	ENSG00000130396.16	+	467
7	17338069	17338070	ENST00000419382.1	-	17338246	17338247	ENSG00000106546.8	+	176
7	17338069	17338070	ENST00000452249.1	-	17338246	17338247	ENSG00000106546.8	+	176
7	17338082	17338083	ENST00000415246.1	-	17338246	17338247	ENSG00000106546.8	+	163
7	23145321	23145322	ENST00000419813.1	-	23145353	23145354	ENSG00000122550.13	+	31
7	29603278	29603279	ENST00000447171.1	-	29603427	29603428	ENSG00000176532.3	+	148
7	35840215	35840216	ENST00000437235.3	-	35840542	35840543	ENSG00000122545.13	+	326
7	35840197	35840198	ENST00000412856.1	-	35840542	35840543	ENSG00000122545.13	+	344
7	35840225	35840226	ENST00000424194.1	-	35840542	35840543	ENSG00000122545.13	+	316
7	39605837	39605838	ENST00000439751.2	-	39605975	39605976	ENSG00000241127.3	+	137
7	39989353	39989354	ENST00000569710.1	-	39989636	39989637	ENSG00000065883.10	+	282
7	77325581	77325582	ENST00000440088.1	-	77325760	77325761	ENSG00000187257.10	+	178
7	77325578	77325579	ENST00000416650.1	-	77325760	77325761	ENSG00000187257.10	+	181
7	77325581	77325582	ENST00000398043.2	-	77325760	77325761	ENSG00000187257.10	+	178
7	77325569	77325570	ENST00000447009.1	-	77325760	77325761	ENSG00000187257.10	+	190
7	86781597	86781598	ENST00000433446.1	-	86781677	86781678	ENSG00000135164.14	+	79
7	86974830	86974831	ENST00000359941.5	-	86974997	86974998	ENSG00000005469.7	+	166
7	86974778	86974779	ENST00000610086.1	-	86974997	86974998	ENSG00000005469.7	+	218
7	86974801	86974802	ENST00000421293.1	-	86974997	86974998	ENSG00000005469.7	+	195
7	127292065	127292066	ENST00000490314.1	-	127292234	127292235	ENSG00000197157.6	+	168
7	129251470	129251471	ENST00000608694.1	-	129251555	129251556	ENSG00000106459.10	+	84
7	144052311	144052312	ENST00000470435.1	-	144052381	144052382	ENSG00000050327.10	+	69
7	155089250	155089251	ENST00000609974.1	-	155089486	155089487	ENSG00000186480.8	+	235
8	1921775	1921776	ENST00000517676.1	-	1922044	1922045	ENSG00000176595.3	+	268
8	6264068	6264069	ENST00000500118.2	-	6264113	6264114	ENSG00000147316.8	+	44
8	6264062	6264063	ENST00000606853.1	-	6264113	6264114	ENSG00000147316.8	+	50
8	9911763	9911764	ENST00000562143.1	-	9911778	9911779	ENSG00000175806.10	+	14
8	42010280	42010281	ENST00000564481.1	-	42010464	42010465	ENSG00000070718.7	+	183
8	42128428	42128429	ENST00000523459.1	-	42128820	42128821	ENSG00000104365.9	+	391
8	42128437	42128438	ENST00000518994.1	-	42128820	42128821	ENSG00000104365.9	+	382
8	42128714	42128715	ENST00000518213.1	-	42128820	42128821	ENSG00000104365.9	+	105
8	61429338	61429339	ENST00000530725.1	-	61429416	61429417	ENSG00000104388.10	+	77
8	61429353	61429354	ENST00000532232.1	-	61429416	61429417	ENSG00000104388.10	+	62
8	64081000	64081001	ENST00000603538.1	-	64081112	64081113	ENSG00000185728.12	+	111
8	67341211	67341212	ENST00000499642.1	-	67341263	67341264	ENSG00000179041.2	+	51
8	81397853	81397854	ENST00000605948.1	-	81397854	81397855	ENSG00000205189.7	+	0
8	86089275	86089276	ENST00000562577.1	-	86089460	86089461	ENSG00000133740.6	+	184
8	86089296	86089297	ENST00000566000.1	-	86089460	86089461	ENSG00000133740.6	+	163
8	90769954	90769955	ENST00000519655.2	-	90769975	90769976	ENSG00000104312.6	+	20
8	90769938	90769939	ENST00000504145.1	-	90769975	90769976	ENSG00000104312.6	+	36
8	90769591	90769592	ENST00000523859.1	-	90769975	90769976	ENSG00000104312.6	+	383
8	92082350	92082351	ENST00000522817.1	-	92082424	92082425	ENSG00000155100.6	+	73
8	92082416	92082417	ENST00000524003.1	-	92082424	92082425	ENSG00000155100.6	+	7
8	100025271	100025272	ENST00000521696.1	-	100025494	100025495	ENSG00000132549.14	+	222
8	125486804	125486805	ENST00000499418.2	-	125486979	125486980	ENSG00000170881.4	+	174
8	125486594	125486595	ENST00000519861.1	-	125486979	125486980	ENSG00000170881.4	+	384
8	125486816	125486817	ENST00000530778.1	-	125486979	125486980	ENSG00000170881.4	+	162
8	126010439	126010440	ENST00000523030.1	-	126010739	126010740	ENSG00000104549.7	+	299
8	143751387	143751388	ENST00000422119.2	-	143751726	143751727	ENSG00000167653.4	+	338
8	143751411	143751412	ENST00000512113.2	-	143751726	143751727	ENSG00000167653.4	+	314
8	143751405	143751406	ENST00000503272.1	-	143751726	143751727	ENSG00000167653.4	+	320
8	144450717	144450718	ENST00000518049.1	-	144451057	144451058	ENSG00000158106.8	+	339
9	2621412	2621413	ENST00000416826.2	-	2621834	2621835	ENSG00000147852.11	+	421
9	4679470	4679471	ENST00000609131.1	-	4679559	4679560	ENSG00000106993.7	+	88
9	4679501	4679502	ENST00000607997.1	-	4679559	4679560	ENSG00000106993.7	+	57
9	35658014	35658015	ENST00000602361.1	-	35658301	35658302	ENSG00000159884.7	+	286
9	35658013	35658014	ENST00000363046.1	-	35658301	35658302	ENSG00000159884.7	+	287
9	72435654	72435655	ENST00000439418.1	-	72435709	72435710	ENSG00000204711.4	+	54
9	72435582	72435583	ENST00000453410.1	-	72435709	72435710	ENSG00000204711.4	+	126
9	72435598	72435599	ENST00000526458.1	-	72435709	72435710	ENSG00000204711.4	+	110
9	98637868	98637869	ENST00000429781.1	-	98637983	98637984	ENSG00000182150.11	+	114
9	98637551	98637552	ENST00000427259.1	-	98637983	98637984	ENSG00000182150.11	+	431

Continued on next page

Table A1: *Continued from previous page*

Chr	Start	End	Transcript ID	Str	Start	End	Gene ID	Str	Dist
9	102668882	102668883	ENST00000529965.1	-	102668915	102668916	ENSG00000136874.6	+	32
9	108320309	108320310	ENST00000421614.1	-	108320411	108320412	ENSG00000106692.9	+	101
9	115512735	115512736	ENST00000440009.1	-	115513118	115513119	ENSG00000148158.12	+	382
X	11129228	11129229	ENST00000608176.1	-	11129421	11129422	ENSG00000004961.10	+	192
X	11129257	11129258	ENST00000608576.1	-	11129421	11129422	ENSG00000004961.10	+	163
X	11129233	11129234	ENST00000433747.2	-	11129421	11129422	ENSG00000004961.10	+	187
X	11129260	11129261	ENST00000608916.1	-	11129421	11129422	ENSG00000004961.10	+	160
X	23801072	23801073	ENST00000366134.2	-	23801290	23801291	ENSG00000130066.12	+	217
X	48367225	48367226	ENST00000445586.1	-	48367350	48367351	ENSG00000102312.16	+	124
X	118602224	118602225	ENST00000446986.1	-	118602363	118602364	ENSG00000005022.5	+	138
X	130192119	130192120	ENST00000412420.1	-	130192216	130192217	ENSG00000147256.6	+	96
X	146993334	146993335	ENST00000594922.1	-	146993469	146993470	ENSG00000102081.9	+	134
X	147582134	147582135	ENST00000456981.1	-	147582139	147582140	ENSG00000155966.9	+	4
X	149009869	149009870	ENST00000427671.1	-	149009941	149009942	ENSG00000156009.5	+	71
X	151883037	151883038	ENST00000415810.1	-	151883082	151883083	ENSG00000183305.9	+	44

Appendix B

Table B1: List of genes regulated by ncRNA-RB1

Gene ID	Fold-Change $\frac{kd\_ncRNA-RB1}{control}$
SLC45A1	2.200856122
CLSTN1	2.052514424
SPEN	2.069549204
DBT	2.046796354
AMY2A	2.120613752
AMY1B	0.482050047
AL592284.1	2.597208218
POLR3GL	0.475976256
BX842679.1	0.200342672
FLG	2.009668955
SYT11	2.523909518
KCNH1	2.046999604
CSGALNACT2	2.099423549
MARCH8	0.484836161
ARHGAP19-SLIT1	2.039554513
CTSW	2.016604618
DGAT2	0.433594527
RAB39A	2.964628456
UPK2	2.284171349
CBL	2.122902015
CD163	2.109484004
PTPRO	2.075954368
SLCO1B7	2.045546942
RP11-125O5.2	2.571298443
C12orf68	2.118379153
KRT84	0.151470779
HOXC8	0.287645099
AL359736.1	0.377747099
TTC7B	0.416658877
OTUB2	2.100004373
AL117190.2	2.197194635
MAP1A	2.381616911
FBN1	2.497451912
SHC4	2.083192263
RSL24D1	2.487629084
RBPMS2	0.316453401
SLC51B	0.478123123
RP11-210M15.2	2.017645666
RP11-89K11.1	2.020746078
APOBR	0.344020477
CMTM4	0.432268128
DEF8	0.286416002
CENPBD1	0.388158174
DBNDD1	0.357193652
GAS8	0.19740295
C16orf3	0.332046425
URAHP	0.225547274
PRDM7	0.112208436
ALOXE3	2.116305715
OMG	2.160162744
COPRS	0.441495917
RP11-1055B8.6	0.495849799
RP11-595B24.2	2.010587803
CRB3	0.467426508
C19orf80	2.621786745
CALR	0.492081732

Continued on next page

Table B1: *Continued from previous page*

Gene ID	Fold-Change $\frac{kd\_ncRNA-RB1}{control}$
RAB8A	0.406311533
SPINT2	0.445735337
LMTK3	21.74051663
SULT2B1	19.92038851
SPHK2	6.091379755
DBP	8.273742235
CA11	8.719269838
NTN5	20.57105759
FUT2	9.572867666
MAMSTR	85.97012861
FUT1	7.796864869
HSD17B14	4.733837457
DHDH	97.94071323
GP6	2.257961265
KRTCAP3	0.472891765
MRPL33	0.485832661
AC110084.1	0.389201934
AC007401.2	2.213433197
EPCAM	0.402526456
SPOPL	2.268759887
ARL6IP6	0.466299071
CHN1	0.466660949
ABCA12	2.099322046
MARCH4	2.07049195
SEC14L6	2.16698168
TTLL1	0.471709007
CDPF1	0.392831547
TSEN2	2.010354222
ADAMTS9	2.444829279
ZMAT3	2.350760587
AL590235.1	0.353380957
SMIM14	0.448806802
RBM47	0.422038329
IL8	2.035527583
CXCL5	0.235821495
SCARB2	0.468738074
CCNG2	0.496192408
AGPAT9	0.375176131
CDS1	0.494183429
HHIP	2.50113646
GLRB	0.478348281
RP11-404L6.2	2.974914831
POLR3G	2.039879453
FSTL4	2.163318407
PPP2R2B	2.413621617
SPINK9	2.221763016
DOCK2	2.070621931
FAM196B	2.084408787
CPLX2	2.03929938
RNF182	2.97167249
HIST1H4J	0.430188197
SPDEF	2.003949031
AL035588.1	0.412556322
COL12A1	2.352489754
TSPAN13	0.417739443
SNX10	0.486812049
CACNA2D1	2.358725145
CDK6	2.035218173
TMEM209	2.012902741
HIPK2	2.318693281
REXO1L1	2.123007569
MMP16	2.213718459
C8orf47	0.433759717
DNAJC25-GNG10	0.460545163
GNG10	0.454720497
COL5A1	2.309645421
NACC2	0.468154288
PNRC2	2.143791548
TXLNG	2.086003683
CDKL5	2.115138397
ARX	2.289863401
XAGE2	0.467589514
PGAM4	2.15137716
CT45A4	2.583966826

Continued on next page

Table B1: *Continued from previous page*

Gene ID	Fold-Change $\frac{kd\_ncRNA-RB1}{control}$
SPANXB2	2.011163436
RAB39B	2.784522673

Table B2: List of genes regulated by RB1

Gene ID	Fold-Change $\frac{kd\_RB1}{control}$
AL645608.1	0.496348767
ISG15	6.011936026
TNFRSF9	2.356636447
C1orf195	0.385091733
IFI6	8.765113707
AL929472.1	2.199898166
RNF11	0.429031613
IFI44	7.056040518
GBP1	3.334995929
ADAMTSL4	2.028343956
IFI16	2.566274123
RP11-565P22.6	0.488806935
NEK7	2.105128092
ZBED6	2.234199938
CDC42BPA	2.066592983
TET1	2.178094494
IFIT2	13.86814038
IFIT3	7.63782133
IFIT1	22.46841766
IFIT5	2.015337238
TRUB1	0.412436981
IRF7	4.612488379
TRIM21	2.366134984
ARHGAP1	0.495274562
RTN4RL2	0.451009799
BATF2	5.643334226
SC5D	2.002218921
ETS1	0.398759806
RBP5	2.159461508
SLC2A14	2.028005284
SLC2A3	2.099080959
ARHGDIB	0.411161608
KRT76	5.105891517
STAT2	2.57440064
CTD-2021H9.3	2.361680203
OAS1	4.430859636
OASL	14.56966859
CDK2AP1	0.42396939
RB1	0.230052032
LPAR6	0.382430448
RP11-468E2.4	2.343458152
IRF9	3.560310501
REC8	2.497430144
RP11-463J10.2	2.042905848
RP11-463C8.4	2.049107955
DICER1	2.103249894
GOLGA8F	0.397372179
DMXL2	2.146800707
GOLGA6L10	0.471862856
MMP25	2.421685609
ATP2A1	0.321290785
LAT	0.418663927
MT1F	0.316066589
MT1G	0.456719007
MT1X	0.418306129
NLRC5	2.521587392
CDH1	0.409909829
XAF1	6.334215702
SMCR8	2.13780816
DHX40	0.495999473
RNF213	2.142623537
RAB12	0.309236482
ANGPTL6	2.193284385
ZNF66	2.202291343
PRODH2	0.446919063

Continued on next page

Table B2: *Continued from previous page*

Gene ID	Fold-Change $\frac{kd\_RB1}{control}$
ZNF155	2.705246996
ZNF223	2.410867606
CTC-512J12.6	2.141150446
FOXA3	0.411965521
GYS1	28.57961281
RUVBL2	7.921797905
LHB	387.322803
CGB	68.35917932
CTB-60B18.6	164.6034365
CGB1	166.5939343
CGB2	186.2109748
CGB5	79.907749
NTF4	683.1759702
CGB8	70.71345588
CGB7	399.0746372
KCNA7	24.12471339
SNRNP70	3.955501268
LIN7B	157.2697791
C19orf73	56.52321157
PPFIA3	126.8576515
PRRG2	2.046114502
AC003006.7	2.097964729
AC004017.1	2.59196841
CTD-2583A14.10	2.014372286
UCN	0.403009393
EIF2AK2	2.09184868
TET3	2.113257406
INHBB	0.426525455
FMNL2	2.05165731
IFIH1	7.190730798
SP110	3.347794622
GBX2	0.403674585
ZNFX1	2.163459951
HELZ2	5.009543596
MX1	20.53270757
C2CD2	2.008169149
USP18	2.337882576
USP41	2.180782614
SERPIND1	2.248774077
TRANK1	6.121010385
RAD54L2	2.152380357
ZBTB20	2.25005579
PARP9	3.182996679
DTX3L	2.961074245
PARP14	2.804421962
PRR23C	2.06954897
PLSCR1	2.006500705
HES1	0.494739855
PPM1K	2.258972743
HERC6	2.72827973
HERC5	2.266687923
GPRIN3	2.094768223
KIAA1109	2.022597876
DDX60	3.072625407
DDX60L	2.711165839
C5orf42	2.126252783
MAP1B	2.048509157
PPP2CA	0.433223368
CDKL3	0.490103952
CD83	0.380728174
HIST1H2BK	0.386798675
HIST1H2AK	2.073774474
HIST1H2BO	0.430632403
COL21A1	0.43103872
DST	2.004345678
LIN28B	2.149980153
GJA1	0.474988475
SAMD9	4.720507484
SAMD9L	6.893337582
OCM2	34.74198481
LMTK2	4.677842253
BHLHA15	34.85102075
TECPR1	18.78460001
BRI3	3.155724838

Continued on next page

Table B2: *Continued from previous page*

Gene ID	Fold-Change $\frac{kd\_RB1}{control}$
BAIAP2L1	3.424074807
RP11-514P8.7	0.488551893
SYPL1	0.485884106
AKR1B15	0.389139568
PARP12	2.707541259
ADRA1A	2.3200835
PARP10	2.829161647
CD274	2.318929348
DDX58	4.913801125
TGFBR1	0.417639969
ABCA1	2.107517179
RP11-101E3.5	2.098806621
C9orf69	0.49422568
CU459201.1	2.200566852
SCML2	2.126180295
RP2	2.025834543
CHIC1	2.030217957
ZDHHC9	2.04303871
F8	0.494078195

Table B3: List of genes regulated by nc-RNA-RB1 and RB1

Gene ID	Fold-Change $\frac{kd\_ncRNA-RB1}{control}$	Fold-Change $\frac{kd\_RB1}{control}$
KIAA0754	2.193886781	2.326282358
ZBTB37	2.525633656	2.414134319
IGFN1	2.279235496	2.062475111
NCR3LG1	2.00177361	2.162970982
HIPK3	2.305087215	2.61995428
AP003733.1	0.030648413	0.040814426
INCENP	0.104715282	0.125444414
AP001925.1	3.497622695	3.398962578
RDX	3.021330283	2.824396109
DHH	13.40580023	16.88308423
LMBR1L	8.102637717	10.20901642
KRT82	0.146640094	0.486541227
KRT75	0.051271827	0.194347297
KRT6B	0.118769223	0.347175494
KRT6C	0.112821285	0.368585417
KRT6A	0.078428277	0.38319465
SLC16A7	2.007539194	2.07930596
FNDC3A	2.38475767	2.487311655
TSSK4	2.576205973	2.366561729
GOLGA8G	0.494859925	0.238520083
MORF4L1	2.168202053	2.156049443
HBQ1	0.471810309	0.429920944
RP11-297M9.1	2.232006855	2.240105922
ARL6IP1	2.550026605	2.3286886
SLC35G6	2.081618671	2.97510584
CBX1	3.308837941	2.696752573
GREB1L	2.178590112	2.561396771
NDUFS7	0.425512799	0.469398559
RTBDN	0.341173993	0.470333587
AC024580.1	3.899330479	3.224679756
REL	2.714835973	2.830569985
ANKRD36C	2.883660893	2.8811935
KCNJ13	2.921334517	2.16801154
MAPRE1	2.487375143	2.560045735
AL118506.1	2.571745892	2.558842453
B3GALT5	2.915357459	2.038098743
POM121L7	2.119590915	2.151236424
ECE2	0.457195116	0.405926594
CAMK2N2	0.495617152	0.491329341
C4orf48	0.42932756	0.448403529
FAM160A1	2.062530196	2.253470425
HLA-F	2.299996787	2.305506874
EZR	2.381701133	2.386192647
TNRC18	0.177217194	0.212649116
FBXL18	0.077967786	0.089800915
TRIM74	0.267592586	0.335382673
GNGT1	2.419031935	2.153626283
PEG10	2.949675248	2.608367231

Continued on next page

Table B3: *Continued from previous page*

Gene ID	Fold-Change $\frac{kd\_ncRNA-RB1}{control}$	Fold-Change $\frac{kd\_RB1}{control}$
RASA4B	0.181925973	0.166350175
POLR2J3	0.475593697	0.463867989
RASA4	0.036498707	0.038985032
RP11-514P8.6	0.164175709	0.165806702
UPK3BL	0.488685019	0.475438256
SPDYE2B	0.316198406	0.35361376
ARHGEF35	2.201013082	3.146591332
DNAJB6	2.320909515	2.218156815
HMBOX1	2.481281967	2.083994812
AL160274.1	3.141447865	2.224123841
PSAT1	3.229380337	2.893198017
RP11-508N12.4	4.72738864	4.034633898
GTF3C5	0.120619499	0.139591413
CEL	0.0028871	0.003322004
RALGDS	0.017582408	0.017262808
SURF6	0.182594494	0.211899928
MED22	0.067017556	0.072926663
XAGE5	0.008219026	0.013356565
MST4	2.75056185	2.175187511
LCA10	2.331475388	2.088458837

Publications

Musahl A.S., Huang X., Rusakiewicz S., Ntini E., Marsico A., Kroemer G., Kepp O., Ørom U.A. (2015). A long non-coding RNA links calreticulin-mediated immunogenic cell removal to RB1 transcription. Oncogene. [Musahl et al., 2015]

Marsico, A., Huska, M. R., Lasserre, J., Hu, H., Vucicevic, D., Musahl, A., Ørom, U.A. & Vingron. (2013) M. PROmiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. Genome Biol. [Marsico et al., 2013]