

From The Department of Oncology and Pathology
Karolinska Institutet, Stockholm, Sweden

CHARACTERIZATION OF NON-CODING RNAs IN CANCER

Per Johnsson



**Karolinska
Institutet**

Stockholm 2014

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by åtta.45 Tryckeri AB

© Per Johnsson, 2014

ISBN 978-91-7549-634-4



**Karolinska
Institutet**

**Institutionen för Onkologi-Patologi, Karolinska
Institutet**

CHARACTERIZATION OF NON-CODING RNAs IN CANCER

AKADEMISK AVHANDLING

som för avläggande av medicine doktorsexamen vid Karolinska
Institutet offentligen försvaras i Radiumhemmets föreläsningssal,
P1:01, Karolinska Universitetssjukhuset, Solna

Fredagen den 19 september, 2014, kl 10.00

av

Per Johnsson

Huvudhandledare:

Dan Grandér
Karolinska Institutet
Institutionen för Onkologi-Patologi

Bihandledare:

Martin Corcoran
Karolinska Institutet
Institutionen för Onkologi-Patologi

Andor Pivarcsi
Karolinska Institutet
Institutionen för Medicin

Fakultetsopponent:

Piero Carninci
RIKEN
Division of Genomic Technologies

Betygsnämnd:

Karl Ekwall
Karolinska Institutet
Institutionen för Biovetenskaper och
Näringslära

Marie Öhman
Stockholms Universitet
Institutionen för Molekylär Biovetenskap

Gonçalo Castelo-Branco
Karolinska Institutet
Institutionen för Medicinsk Biokemi och
Biofysik

Stockholm 2014

ABSTRACT

While originally thought to only code for proteins, it now stands clear that RNA is a multifunctional molecule involved in a great variety of molecular processes. Recent advances in genome-wide platforms have revealed underappreciated roles of non-protein coding RNAs (ncRNAs). Thousands of ncRNAs have been characterized using these novel techniques but functional investigations are still very limited.

Pseudogenes constitute one group of non-coding DNA. These genes represent duplications of protein-coding genes that have lost their protein-coding potential through various molecular events. Pseudogenes litter the human genome and it was not until recently it was found that many pseudogenes are actively transcribed. One of those is the *PTEN* pseudogene, PTENpg1.

PTEN is a tumor suppressor gene, which is frequently dysregulated in cancers by epigenetic inactivation. In **paper I**, we characterized a previously unknown long ncRNA, PTENpg1 antisense RNA (asRNA). We identified two different isoforms that are involved in regulating expression of *PTEN* at both the transcriptional and post-transcriptional level. While one isoform recruits the chromatin remodelers DNMT3a and EZH2 to the *PTEN* promoter, the other isoform is involved in microRNA-mediated regulation of *PTEN* mRNA.

In **paper II**, we investigated the role of PTENpg1 asRNA in drug resistance mechanisms to the BRAF inhibitor vemurafenib in melanoma. *BRAF* is a proto-oncogene commonly mutated in melanoma, resulting in its constitutive activation. Therapies that specifically target the mutated BRAF have been developed and vemurafenib is one of those drugs. However, the development of resistance during treatment is a major problem. We found that the PTENpg1 asRNA is involved in resistance to vemurafenib and that gain in expression of PTENpg1 asRNA promotes silencing of *PTEN* in melanoma cells. Moreover, we revealed that the expression of PTENpg1 asRNA is a promising marker for clinical outcome in melanoma patients.

In **paper III**, we investigated the expression of microRNAs in Swedish and Egyptian patients with ABC DLBCL (Acute B-cell diffuse large B-cell lymphoma). We found that differences in environmental exposure alter the expression of microRNAs and that miR-1234 targets the proto-oncogene *STAT3*. The expression of miR-1234 was elevated in the Egyptian patients and *STAT3* was consequently suppressed in these patients. Additionally, we also showed that ectopic expression of miR-1234 in cell line models mediates suppression of *STAT3*.

Taken together, the findings in this thesis reveal intriguing molecular functions of ncRNAs, which are relevant for cancer development. The characterization of PTENpg1 asRNA and miR-1234 may prove important for the development of therapies and for predicting clinical outcome in various cancers.

LIST OF PUBLICATIONS

Johnsson, P., Ackley, A., Vidarsdottir, L., Lui, W.O., Corcoran, M., Grandér, D., and Morris, K.V. (2013). A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nature structural & molecular biology*

Johnsson P., Vidarsdottir, L., Azimi, A., Serviss, A., Ingvar, C., Jönsson, G., Olsson, H., Frostvik Stolt, M., Hansson, J., Egyházi Brage, S., and Grandér, D., PTEN^{pgl} antisense RNA mediates PTEN suppression in vemurafenib resistance and predicts clinical outcome in melanoma patients. *Manuscript*

Hogfeldt, T., **Johnsson, P.**, Grandér, D., Bahnassy, A.A., Porwit, A., Eid, S., Osterborg, A., Zekri, A.R., Lundahl, J., Khaled, M.H., Mellstedt, H., and Moshfegh, A., (2014). Expression of microRNA-1234 related signal transducer and activator of transcription 3 in patients with diffuse large B-cell lymphoma of activated B-cell like type from high and low infectious disease areas. *Leukemia & lymphoma* 55, 1158-1165

PUBLICATIONS NOT INCLUDED IN THESIS

Johnsson, P., Morris, K.V., and Grandér, D. (2014). Pseudogenes: a novel source of trans-acting antisense RNAs. *Methods in molecular biology* 1167, 213-226.

Johnsson, P., Lipovich, L., Grandér, D., and Morris, K.V. (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochimica et biophysica acta* 1840, 1063-1071.

Serviss, J.T., **Johnsson, P.**, and Grandér, D. (2014). An emerging role for long non-coding RNAs in cancer metastasis. *Front. Genet.*

Kharaziha, P., De Raeve, H., Fristedt, C., Li, Q., Gruber, A., **Johnsson, P.**, Kokaraki, G., Panzar, M., Laane, E., Osterborg, A., Zhivotovsky, B., Jernberg-Wiklund, H., Grandér, D., Celsing, F., Björkholm, M., Vanderkerken, K., Panaretakis T. (2012). Sorafenib has potent antitumor activity against multiple myeloma in vitro, ex vivo, and in vivo in the 5T33MM mouse model. *Cancer research* 72, 5348-5362

Mojallal, M., Zheng, Y., Hultin, S., Audebert, S., van Harn, T., **Johnsson, P.**, Lenander, C., Fritz, N., Mieth, C., Corcoran, M., Lembo, F., Hallström, M., Hartman, M., Mazure, N.M., Weide, T., Grandér, D., Borg, J.P., Uhlén, P., Holmgren, L. (2014). AmotL2 disrupts apical-basal cell polarity and promotes tumour invasion. *Nat Commun* 5, 4557

TABLE OF CONTENTS

1. Introduction	1
2. Background	3
2.1 Non-coding RNAs	3
2.1.1 Long ncRNAs.....	3
2.1.2 The FANTOM Consortium.....	4
2.1.3 The ENCODE Consortium	6
2.1.4 Identification of lincRNAs using histone marks.....	7
2.1.5 Subclasses of lincRNAs	8
2.1.6 Mechanisms of regulation	9
2.1.6.1 antisense RNA	9
X-chromosome inactivation: Xist/Tsix	10
Trans-acting lincRNA: HOX Antisense Intergenic RNA	11
Imprinting: Insulin-like growth factor 2 receptor	12
Epigenetic remodeling: Antisense Non-coding RNA in the INK4 Locus	14
miRNA target site: Beta-secretase-1 asRNA	14
Splicing: Zinc finger E-box-binding homeobox 2.....	15
2.1.6.2 Non antisense RNA mediated regulation.....	15
lincRNA mediated regulation of phosphorylation	15
lincRNA-mediated regulation of mRNA stability	16
2.1.7 Transcribed RNA = functional RNA?	17
2.1.8 Effector proteins of lincRNAs.....	18
2.1.8.1 Polycomb repressive complex.....	18
2.1.8.2 DNA methyl transferases	20
2.2 Pseudogenes	23
2.2.1 Evolutionary conservation of pseudogenes.....	25
2.2.2 Functional investigations of pseudogenes.....	25
2.2.3 Competitive endogenous RNAs.....	27
2.2.4 Pseudogenes: a novel source of asRNAs?	27
2.2.5 Technical considerations during functional investigations of pseudogenes	28
2.2.5.1 Quantitative analysis using PCR.....	28
2.2.5.2 Targeting of pseudogenes.....	29
2.2.5.3 Overexpression	30
2.3 Cancer: A lincRNA perspective	31
2.3.1 The Hallmarks of cancer	31
2.3.1.1 Sustaining proliferative signaling.....	31
2.3.1.2 Evading growth suppression	31
2.3.1.3 Enabling replicative immortality.....	32
2.3.1.4 Activating invasion and metastasis	32
2.3.1.5 Inducing angiogenesis	33
2.3.1.6 Resisting cell death	33
2.3.2 Mutations of <i>EZH2</i> and <i>DNMT3a</i>	34
2.4 miRNAs	35
2.4.1 Biogenesis of miRNAs.....	35
2.4.2 miRNAs and cancer.....	36

2.5 PTEN.....	37
2.6 BRAF	39
2.7 STAT3.....	41
2.8 Lymphoma	42
3. Aims of this thesis	43
3.1 Paper I	43
3.2 Paper II.....	43
3.3 Paper III	43
4. Results and discussion	44
4.1 Paper I	44
4.1.1 Rationale	44
4.1.2 Characterization of PTENpg1 asRNA transcripts	44
4.1.3 PTENpg1 asRNA α	44
4.1.4 PTENpg1 asRNA β	45
4.1.5 Further characterization of PTENpg1 asRNA transcripts	45
4.1.6 PTENpg1 asRNA α versus PTENpg1 asRNA β	46
4.1.7 Evolutionary conservation of asRNA mediated regulation of <i>PTEN</i>	46
4.1.8 <i>Cis-</i> versus <i>trans</i> -acting lncRNAs.....	47
4.1.9 PTENpg1 asRNA, DNMT3a and other ncRNAs	48
4.1.10 Concluding remarks of paper I.....	51
4.2 Paper 2.....	53
4.2.1 Rationale	53
4.2.2 Characterization of PTENpg1 asRNA in resistance to vemurafenib	53
4.2.3 PTENpg1 asRNA expression in primary melanoma samples	54
4.2.4 Concluding remarks of paper II.....	55
4.3 Paper 3.....	56
4.3.1 Rationale	56
4.3.2 Characterization of miRNA expression	56
4.3.3 Concluding remarks of paper III	57
5. Concluding remarks	58
6. Acknowledgements	59
7. References.....	61

LIST OF ABBREVIATIONS

ABC DLBCL	Activated B-cell diffuse large B-cell lymphoma
Airn	Antisense Insulin-like growth factor receptor
ALL	Acute Lymphoblastic Leukemia
AML	Acute Myeloid Leukemia
ANRIL	Antisense non-coding RNA in the INK4 locus
ASO	Antisense oligo
asRNA	Antisense RNA
BACE1	Beta-secretase-1
BANCR	BRAF regulated lncRNA
CAGE	Cap analysis of gene expression
CBX	Chromodomain
CDK	Cyclin-dependent kinase
cDNA	Complementary DNA
ceRNA	Competitive endogenous RNA
ChIP	Chromatin immunoprecipitation
ChIP-seq	ChIP sequencing
ChIRP	Chromatin isolation by RNA purification
CPSF1	Cleavage and polyadenylation specific factor 1
DC	Dendritic cell
DDR	DNA-damage response
DLBCL	Diffuse large B-cell lymphoma
DNA	Deoxyribonucleic acid
DNMT	DNA methyl transferase
dsRNA	Double stranded RNA
E-cad	E-caderin
EBV	Epstein Barr virus
EED	Embryonic ectoderm development
EMT	Epithelial-mesenchymal-transition
ENCODE	Encyclopedia of DNA elements

EST	Expressed sequence tag
EZH	Enhancer of zeste homolog
FANTOM	Functional annotation of mouse
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GAS	Gamma interferon activation sequence
HDAC1	Histone deacetylase 1
HIF1 α	Hypoxia inducible factor alpha
HIV	Human immunodeficiency virus
HMGA1	High mobility group A1
HOTAIR	HOX antisense intergenic RNA
HVB	Hepatitis virus B
IGF2	Insulin-like growth factor 2
IGF2R	Insulin-like growth factor 2 receptor
IL-6	Interleukin 6
IRES	Internal ribosome entry site
JAK	Janus tyrosine kinases
kb	Kilobase
Let-7	Lethal-7
LincRNA	Large Intergenic noncoding RNA
LINE	Long interspersed elements
Lnc-DC	Long non-coding dendritic cells
LncRNA	Long non-coding RNA
MAPK	Mitogen-activated protein kinase
MDM2	Mouse double minute 2 homolog
MET	Mesenchymal-epithelial-transition
miRNA	MicroRNA
mRNA	Messenger RNA
ncRNA	Non-coding RNA
nNOS	Neuronal nitric oxide synthase
nt	Nucleotide
Onco-miR	Oncogenic microRNA

ORF	Open reading frame
PCAT-1	Prostate cancer associated transcript 1
PcG	Polycomb group
PI3K	Phosphatidylinositol 3-kinases
PIP ₂	Phosphatidylinositol-4,5-bisphosphate
Piwi-RNA	Piwi-interacting RNA
PKR	Protein kinase R
PMA	Phorbol myristate acetate
Pol II	Polymerase II
PolyA	Poly-adenylated
PRC	Polycomb repressive complex
PRE	Polycomb response elements
Pseudo-asRNA	Pseudogene antisense RNA
PTEN	Phosphatase and tensin homolog
qRT-PCR	Quantitative real-time-PCR
RepA	Repetitive element A
RepC	Repetitive element C
RISC	RNA induced silencing complex
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
rRNA	Ribosomal RNA
RTK	Receptor tyrosine kinase
shRNA	Short hairpin RNA
SINE	Short interspersed elements
siRNA	Small interference RNA
SMD	Staufen1 (STAU1) mediated mRNA decay
snoRNA	Small nucleolar RNA
STAT	Signal transducer and activator of transcription
SUZ12	Suppressor of zeste 12
T-UCR	Transcribed ultra conserved regions
TERC	Telomerase RNA component

TERT	Telomerase reverse transcriptase
TF	Transcription factor
TGS	Transcriptional gene silencing
ti-RNA	Transcriptional initiation RNA
tRNA	Transfer RNA
TSG	Tumor suppressor gene
TSS	Transcriptional start site
UTR	Untranslated Region
Xa	X-chromosome active
XCI	X-chromosome inactivation
Xi	X-chromosome inactive
Xic	X-chromosome inactivation center
XIST	X-inactive specific transcript
YY1	Yin Yang 1
ZEB2	Zinc-finger E-box-binding homeobox 2 gene

1. INTRODUCTION

After sequencing of the human genome was completed in 2001, it became evident that only a minor fraction of the human genome encoded for proteins^{1,2}. Previous estimates of the number of protein-coding genes had ranged from 50,000 up to 140,000 genes³, but genome-wide analysis now indicated that there are approximately 20,000 genes in the human genome⁴. This number was therefore dramatically lower than expected and did not reflect the assumption that the complexity of an organism was related to the number of protein-coding genes⁵ since less complex organisms such as chicken⁶ and mouse⁷ appeared to have approximately the same number of genes. Instead, it was proposed that the relative amount of non-protein-coding DNA could contribute to developmental complexity⁸ and that many of these non-coding regions can function through transcription into non-coding RNAs (ncRNAs).

Recent advances in genome-wide platforms have revealed the transcriptional landscape to be vastly more complex than previously thought^{9,10}. While only 1.2% of the human genome encodes for proteins¹¹, it is now known that the large majority, 70-80%, of the genome is being transcribed into ncRNAs¹²⁻¹⁴. Studies have revealed the presence of pervasive and widespread transcription and tens of thousands of ncRNAs have been identified^{9,10,14,15}. This would suggest that there are more genes coding for ncRNAs, possibly with regulatory functions, than there are genes encoding proteins. Prior to these studies, the presence of long non-coding RNAs (lncRNAs) was largely discounted and the few proven biologically important lncRNAs, such as *Xist*^{16,17} and *H19*¹⁸⁻²⁰, were considered rare cases.

The last decade has changed our view on the human genome and gene regulation. It is becoming increasingly clear that ncRNAs are indeed important and crucial regulatory elements for gene regulation and disease²¹⁻²³. The function of RNA, as we know it today, is far more extensive than anticipated only a decade ago^{1,2,9,10,24}. However, the vast majority of identified ncRNAs to date still remain to be functionally characterized and investigated. Continued efforts to study ncRNAs will increase our understanding of their functional importance and also improve our understanding of their role in disease and evolution. This thesis aims to expand our understanding of ncRNAs, their functional properties and the role they may play in cancer development and progression.

2. BACKGROUND

While originally being considered merely as information between DNA and proteins, it is now becoming increasingly clear that RNA has multiple functions. Technical development of genome-wide platforms such as RNA sequencing (RNA-seq)^{9,13} and chromatin immunoprecipitation sequencing (ChIP-seq)¹⁵ has greatly increased our understanding of gene regulation. Although rare cases such as ribosomal RNAs (rRNAs)²⁵ and transfer RNAs (tRNA)²⁶ were described already in the 1950s, it is not until now it is becoming clear that ncRNAs are involved in a wide range of cellular functions such as epigenetic regulation²¹, cell cycle regulation²⁷, cancer progression^{28,29} and X-chromosome inactivation^{16,17,30} among many others.

2.1 NON-CODING RNAs

ncRNAs are arbitrary divided into short ncRNAs (<200 nt) and lncRNAs (>200 nt). There are numerous short ncRNAs, including piwi-interacting RNAs (piwi-RNAs), small nucleolar RNAs (snoRNAs) and transcriptional initiation RNAs (tiRNAs)³¹. The relatively well-studied microRNAs (miRNAs), which are 19-21 nt long, are also found among the short ncRNAs. Thousands of miRNAs have been discovered and hundreds of those functionally investigated. In contrast, the lncRNAs, including the large intergenic ncRNAs (lincRNA)¹⁵, antisense RNAs (asRNAa)¹⁰ and intronic RNAs³², are not as thoroughly investigated. While the subclass of miRNAs is well defined, the role of lncRNAs appears to be more diverse. Whereas the miRNAs predominantly act as post-transcriptional regulators in the cytoplasm, lncRNAs act as both positive and negative regulators of gene expression, playing roles in chromatin structure, epigenetic remodeling and RNA stability.

2.1.1 Long ncRNAs

Already in the 1990s, pioneering work had proposed various functional roles for lncRNAs. Although very few, the examples ranged from imprinting of the lncRNA H19^{18,19}, X-chromosome inactivation (XCI) by the X-inactive specific transcript (Xist)^{16,17}, as well as asRNA mediated regulation of *Xist* by the Tsix transcript^{30,33}. These examples highlighted widespread function of lncRNAs, but were at the time mainly considered as special and rare cases.

In 2002, it started to become clear that the presence of lncRNAs could be more prevalent than previously thought. Kapranov *et al* carried out pioneering work, where transcriptionally active regions of human chromosomes 21 and 22 were investigated in a number of different tumor cell lines²⁴. By using oligonucleotide arrays, a surprisingly large number of transcriptionally active regions were detected, not mapping to any known protein-coding genes. Kapranov and colleagues reported that only ~10% of the transcriptional active regions on chromosomes 21 and 22 corresponded to protein-coding exons and that the number of

non-coding regions greatly exceeded the coding regions. Already at that point, the authors speculated that some of these newly identified transcripts could indeed function as ncRNAs²⁴.

The year after, Rinn *et al* designed a similar study and once again addressed the transcriptional landscape of human chromosome 22³⁴. In this study, the approach used also allowed for strand specific analysis. Numerous novel transcribed regions were found and many of them originated from introns of known protein-coding genes representing both novel exons as well as expressed introns. In addition, a significant fraction of these novel RNAs were found to be transcribed in the opposite direction, thus forming sense:antisense pairs of transcription. Overall, the study by Rinn *et al* characterized twice as many transcribed bases as previously reported on chromosome 22 and noted that several of the transcribed regions were conserved between human and mouse. In addition, the authors speculated that some of the overlapping genes could function as asRNAs, which may control levels of the gene encoded on the opposite strand³⁴.

The above-mentioned studies are some of the early examples that initially highlighted that the transcriptional landscape and complexity of the genome had been vastly underestimated. These studies barely scratched the surface and advances in genome-wide technologies would soon support these observations. Experiments carried out by the two large international consortia FANTOM (Functional Annotation Of Mouse) and ENCODE (The Encyclopedia of DNA elements) have during the last decade generated large amounts of data from numerous species, tissues and cell lines. The work by these two consortia has generated large libraries of data and provides a major resource for functional genome annotation that is freely available for scientists worldwide. This allows for independent analysis for finding novel RNA transcripts, transcriptional start sites (TSS), transcription factor (TF) binding sites and much more. This work has largely changed our view on biology and highly contributed to our understanding of the complexity of the eukaryotic genome.

2.1.2 The FANTOM Consortium

The FANTOM Consortium was initiated in the year 2000 and originally aimed to assign functional annotations to mouse full-length complementary DNA (cDNA)³⁵. The initial findings on chromosomes 21 and 22^{24,34} were indeed supported over the next few of years by the FANTOM1³⁶, FANTOM2^{37,38} and FANTOM3^{9,10} projects. The data generated by the FANTOM3 project was in many ways remarkable and made it clear that a large majority of the mouse genome is transcribed and the overall coverage of transcribed regions spanning the mouse genome was reported to be as much as 62.5%⁹. The FANTOM projects combine full-length cDNA isolation and mapping of 5' - ends of RNA transcripts with the CAGE technique (cap-analysis of gene expression)³⁹. The CAGE technique captures polymerase II (pol II) transcribed RNAs, which are capped on the 5' end. This allows for the mapping of transcriptionally active regions and the identification of pol II driven TSSs.

The data generated by FANTOM3 was subsequently also used to address the frequency of asRNA transcription¹⁰. asRNA transcription refers to the fact that DNA can be transcribed on both strands. Normally, an asRNA is defined as the transcript being transcribed from the opposite strand in relation to a protein-coding sense strand transcript. Therefore, the protein-coding sense strand ultimately share sequence complementarity with the asRNA transcript, forming a so-called sense:asRNA pair. The analysis of the FANTOM3 data reported that 72% of all transcribed regions of the mouse genome also had a related asRNA annotation¹⁰. Knockdown studies showed that asRNAs could be negative as well as positive regulators of the corresponding protein-coding sense gene. The majority of the asRNAs appeared to be poly-adenylated (polyA), but some of them were also reported as polyA negative¹⁰.

FANTOM4 focused on understanding the dynamics of TSSs upon monocytic differentiation of the acute myeloid leukemia (AML) cell line THP-1. In a time-course study during phorbol myristate acetate (PMA) induced differentiation, CAGE analysis characterized differentially expressed TSSs and active promoters. This allowed for the identification of important promoter regions, which could be used for prediction of key TF binding sites⁴⁰. This approach identified key factors that determine monocytic differentiation and showed that cell states are dictated by complex networks of both positive and negative regulators depending on numerous TFs⁴⁰. No single factor by itself was enough to dictate cell states and the data suggested coordinated and complex networks where the activity of several TFs is required for complete differentiation. The complex network of TFs was further explored within FANTOM4 by creating an atlas of TF-TF interactions⁴¹. In support of the previous findings, the TF-TF interactome revealed that single TFs do not determine tissue specificity⁴¹. Instead, many TFs were found to be expressed among many different tissues and tissue specificity was rather mediated by tissue specific TF-TF interactions⁴¹.

The data generated by the FANTOM4 project also identified the presence of tiRNAs³¹ and transcribed retrotransposones⁴². tiRNAs are ~18 nt long and the majority of them were shown to coincide with TSSs (from -60 to +120 nt) of actively transcribed CpG rich pol II promoters³¹. tiRNAs were found in human, chicken and *Drosophila* and a functional role in chromatin modifications and transcriptional initiation have been suggest, possibly through interacting with the protein CTCF⁴³. Finally, FANTOM4 discovered that retrotransposon elements such as LINE (Long interspersed elements) and SINE (Short interspersed elements), among others, are frequently transcribed⁴². Repetitive elements, in particular retrotransposons, comprise 30-50%^{44,45} of mammalian genomes and it was reported that as much as 30% of all CAGE tags initiated within repetitive elements in some tissues, thus having a large impact of the transcriptional output. It was suggested that retrotransposons that localize in close proximity to promoter regions could serve as alternative promoters thus generating alternative transcripts of both messenger RNAs (mRNAs) and ncRNAs. In addition, 25% of all protein-coding genes were found to possess retrotransposons in their 3'UTR (untranslated region). Interestingly, this subset of mRNAs was in general expressed at lower levels and it was speculated that this could be a consequence of the introduction of

miRNA binding sites. It was also suggested that retrotransposons in the 3'UTR could recruit *trans*-acting mechanisms with the potential to promote degradation of the mRNAs.

The most recent update from the FANTOM Consortium was announced in 2014. The FANTOM5 project published a series of papers (~20) aimed at mapping the TSSs of the majority of promoters⁴⁶ and enhancers⁴⁷ in the mammalian genome. Approximately 1,000 human and 400 mouse samples, including primary cells, tissues and cancer cell lines were mapped using the CAGE technique³⁹. A comprehensive promoter atlas of gene expression was generated and CAGE tags supported the presence of over 90% of all annotated protein-coding genes. In addition, the authors studied the evolutionary conservation of TSSs among human and mouse. It was reported that TSSs representing highly expressed promoter regions appear more conserved than others, while TSSs of ncRNAs were less conserved.

2.1.3 The ENCODE Consortium

Launched in 2003, the ENCODE project aimed to identify all functional elements in the human genome by using high-throughput methods⁴⁸. The pilot phase of the ENCODE project was set out to investigate 1% of the human genome and included the analysis of a number of human cell lines using methods such as RNA-seq, ChIP-seq and DNase I profiling. This initial phase allowed for the evaluation and development of techniques that would later be used for more comprehensive characterization of the remaining 99% of the human genome. The ENCODE pilot phase was published in 2007¹² and provided evidence for pervasive transcription as the majority of the investigated loci were associated with primary RNA transcripts (including unspliced RNAs). Many novel RNA transcripts were identified, both overlapping with protein-coding loci as well as within loci previously thought to be transcriptionally silent. Certain histone marks were shown to correlate remarkably well with transcriptionally active regions. Trimethylation of histone 3 lysine 4 (H3K4me3) was for example shown to be enriched at active promoters. The authors also identified 201 pseudogenes and 19% of those showed evidence of transcription. Moreover, it was also observed that a large fraction of the identified functional elements appeared to lack evolutionary conservation throughout mammalian evolution.

About five years after publishing the ENCODE pilot phase, the studies had been expanded to cover the remaining 99% of the human genome¹³. In 2012, a set of 30 papers were published by the ENCODE Consortium, presenting a great variety of findings such as RNA-seq, TF-binding, histone marks and sequence conservation of regulatory regions. Altogether, the data included 1.640 data sets from 147 different cell types. Although not possible to cover all findings within this thesis, some of the highlights, in particular with regards to lncRNAs, will be discussed.

Overall, 80.4% of the human genome was shown to have a biochemical function defined as: a genome segment either encoding for RNA or having a reproducible signature for chromatin structure or protein binding⁴⁹. Moreover, ~60-75% of the genome was reported to be

transcribed into RNA¹³ and the transcribed regions were very well predicted by histone marks⁵⁰. Among the lncRNAs, mass spectrometry analysis further supported that most were not translated into proteins (92%)⁵¹. Also, these transcripts were found predominantly localized to chromatin structures and to the nucleus¹⁴. lncRNAs were in general less conserved than protein-coding mRNAs, but promoter regions of lncRNAs were almost as conserved as promoters for protein-coding genes¹⁴. Interestingly, a third of the lncRNAs seemed to be primate-specific. The expression of lncRNAs also appeared to be more tissue specific than protein-coding mRNAs. While 25% of the mRNAs were expressed at < 1 copy per cell, the corresponding number for lncRNAs was 80%¹³. This does not necessarily imply that lncRNAs in general are expressed at very low levels, but could instead indicate that lncRNAs are expressed in very specific subset of cells or tissues. Altogether, ENCODE identified transcription of ~9,500 lncRNAs, including 5,058 lincRNAs and 3,124 asRNAs⁵². In addition, ENCODE also investigated the presence of pseudogenes in the human genome. 11,224 pseudogenes were identified and 863 of those showed evidence of transcription.⁵³ The distribution of genes which have been identified by ENCODE is annotated by the GENCODE Consortium⁵². The GENCODE v7 annotation of genes is illustrated in Figure 1.

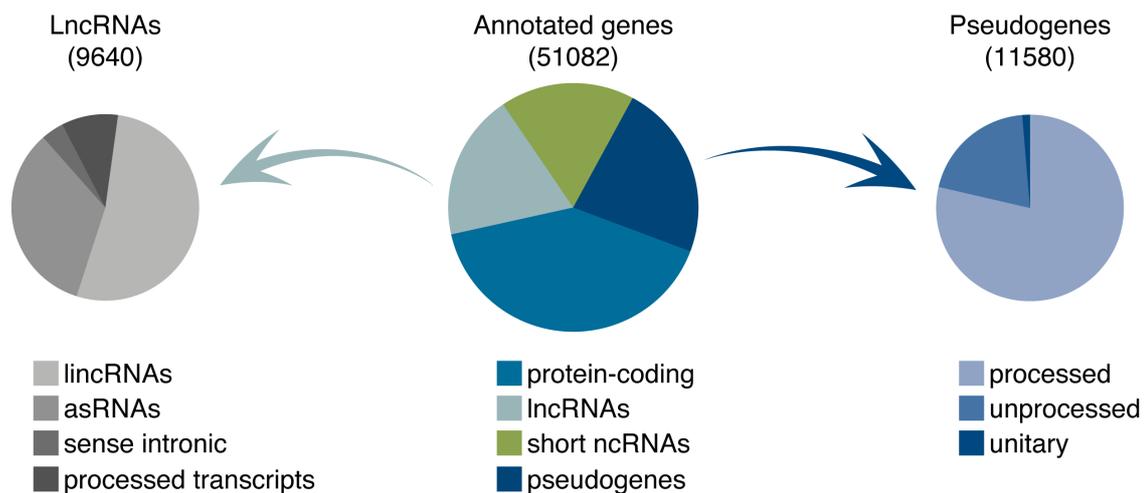


Figure 1 The number of human annotated genes in GENCODE v7¹⁴. Pseudogenes are discussed in more detail in section 3.2.

2.1.4 Identification of lincRNAs using histone marks

In addition to the work carried out by the FANTOM and ENCODE Consortia, alternative approaches have also been used in order to identify lncRNAs. By using ChIP-seq data it was found that actively transcribed regions could be identified by the chromatin marks H3K4me3 and H3K36me3 (trimethylation of histone 3 lysine 36). While active promoter regions are marked by H3K4me3, pol II transcribed regions are marked with H3K36me3 throughout the transcribed region (so called K4-K36 domains)^{12,54}. Searching for these chromatin structures could therefore identify previously uncharacterized expressed regions, possibly encoding novel lncRNAs.

In a study carried out by Guttman *et al*¹⁵, ChIP-seq data from mouse embryonic stem cells, mouse embryonic fibroblasts, mouse lung fibroblasts and neural precursor cells were analyzed for K4-K36 domains. By specifically looking for K4-K36 domains that did not overlap with known protein-coding genes, the authors discovered ~1,600 putatively novel lincRNAs. Microarrays covering these intergenic domains verified 70% to be actively transcribed. The subset of lincRNAs identified by H3-K36 domains showed clear sequence conservation. Although the lincRNA exons showed less conservation compared to protein-coding exons, the promoter regions were often equally conserved. Moreover, 70% of human orthologous regions also showed corresponding K4-K36 domains. Taken together, this implied that many of the identified lincRNAs could be related to biological function.

Although no concrete biological function was associated with the lincRNAs at that point, it was noted that many promoters harbored binding sites for Oct4, Nanog and p53. Indeed, activation of p53 by exposing mouse embryonic fibroblasts to DNA damage triggered the expression of 39 lincRNAs. Moreover, ChIP-seq data supported direct binding of Oct4 and Nanog to several of the promoters of the lincRNAs, consequently supporting lincRNAs to be regulated by key TFs¹⁵.

2.1.5 Subclasses of lincRNAs

lincRNAs do not have a single clear definition, except that they are longer than 200 nt⁵⁵. The lincRNAs are instead divided into subclasses based on genomic features mainly dependent on localization relative to protein-coding genes. The most well established subclasses of lincRNAs are the following (Figure 2)^{14,52}:

- Antisense RNAs (asRNA) (Figure 2A)
- Large intergenic RNAs (lincRNAs) (Figure 2B)
- Sense overlapping RNAs (Figure 2C)
- Sense intronic RNAs (Figure 2D)
- Processed transcripts (do not contain an open reading frame (ORF) and cannot be placed in any of the other categories)

In addition to these main categories, numerous other subclasses such as transcribed ultra conserved regions (T-UCR)⁵⁶ and transcribed pseudogenes⁵⁷ have also been proposed. Importantly, as far as we know today there is no evidence of a difference in function among these subclasses and the distinct classes of lincRNAs do not differentiate between specific functions⁵⁵. For example, asRNAs can act as both positive (concordant) and negative (discordant) regulators of their corresponding sense transcript^{21,58}. Moreover, lincRNAs are involved in regulation of locus specific chromatin remodeling⁵⁹, organization of multichromosomal regions⁶⁰ as well as post-translational regulation^{61,62}. Below follows a brief selection of pioneering investigations of functional properties of asRNAs and lincRNAs.

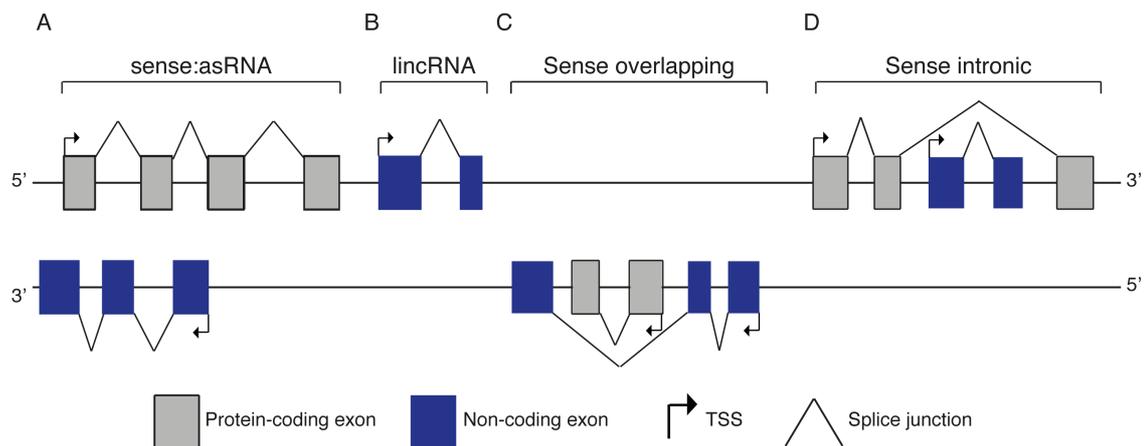


Figure 2 A schematic figure depicting the different subclasses of lncRNAs.

2.1.6 Mechanisms of regulation

2.1.6.1 antisense RNA

asRNAs are expressed from the opposite strand of a sense transcript, which is either a protein-coding mRNA or another ncRNA gene. The asRNA shares complementarity to the sense expressed transcript (with the exception of bidirectional promoters) and can overlap promoters, exons, 5'- and 3'- UTRs as well as introns (Figure 3).

The presence of asRNA transcription has been known for decades, first studied in bacteria (reviewed in⁶³) and later in eukaryotes (reviewed in⁶⁴). Well-studied examples such as inactivation of the X-chromosome³⁰ and the Insulin-like growth factor 2 receptor (*Igf2r*)⁶⁵ had served as groundbreaking examples of asRNA mediated regulation. However, it was not until the introduction of genome-wide approaches that it became clear that the presence of asRNA transcription is widespread¹⁰. The number of protein-coding genes with a corresponding asRNA transcript is estimated to range from 20-40%^{66,67}. The studies carried out by the ENCODE Consortium indicate that asRNAs, just as lncRNAs in general, preferentially accumulate in the nucleus^{13,14}.

The genomic arrangement of sense:asRNA transcription gives the impression that they act on each other through *cis*-acting mechanisms. Indeed, most of the examples studied to date suggest such a regulatory mechanism^{21,68-70}. However, this model may have been an oversimplified view of asRNA-mediated regulation and a number of recent reports challenge this view and show that asRNAs can also function through *trans*-regulatory mechanisms (Figure 3)^{59,71}.

Numerous mechanisms of action have been proposed for asRNA mediated regulation including epigenetic changes²¹, imprinting^{30,65}, altered splicing²³ and RNA stability^{68,69}. Moreover, asRNAs have also been implicated in disease and for example several cancer-associated genes such as p15²¹, p21⁷⁰ and p53⁶⁸ have been shown to be under the regulation of asRNAs. At the moment, characterization and understanding of asRNA-mediated

regulation relies on detailed molecular investigation of specific cases. Clear roles have emerged for some asRNAs and a selection of the most thoroughly investigated genes will be discussed in greater detail.

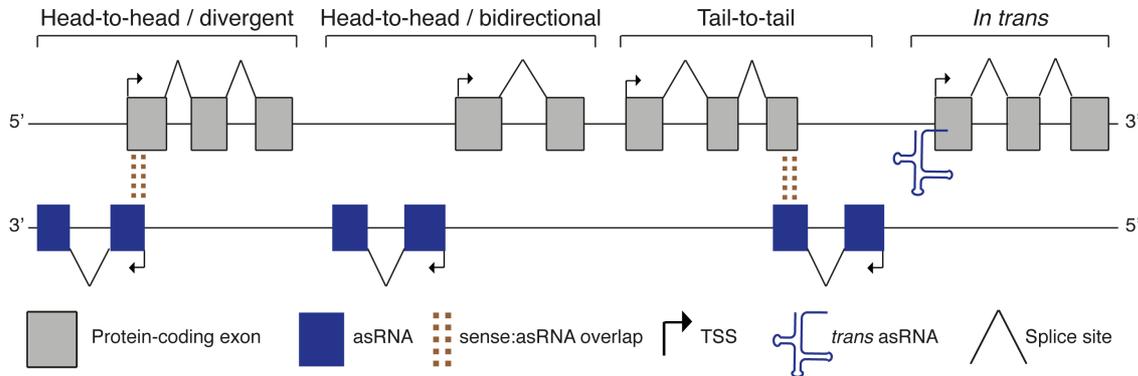


Figure 3 A schematic figure depicting different forms of asRNA transcripts.

X-chromosome inactivation: Xist/Tsix

The lncRNAs involved in XCI (X-chromosome inactivation) are probably among the best characterized and often serve as models for understanding lncRNA-mediated epigenetic regulation. In contrast to males, females have two X-chromosomes and undergo XCI to balance the expression of the nearly ~1,000 genes localized on the X-chromosome. This process is initiated at the X-chromosome inactivation center (*Xic*) and is dependent and regulated by a number of different lncRNAs; *Xist*, *Tsix*, *RepA*, *Jpx* and *Xite*. The *Xist* transcript (X-inactivate-specific transcript) evades inactivation and is transcribed on the inactive X-chromosome (Xi) and was one of the first lncRNAs to be identified^{72,73}. *Xist* is a nuclear localized ~17 kilobases (kb) *cis*-acting lncRNA^{16,17} which maps to the *Xic* and is required for initiation of the XCI⁷⁴. *Xist* coats the X-chromosome and recruits chromatin remodeling complexes to the targeted X-chromosome whereby chromosome-wide silencing is induced and spreads along Xi. More specifically, the 5' end of *Xist* contains a conserved region called Repetitive element A (*RepA*)⁷⁵ which binds and recruits the RNA binding protein Enhancer of zeste homolog 2 (*EZH2*) to Xi⁷⁶. *EZH2*, a histone methyl transferase and a subunit of the polycomb repressive complex 2 (*PRC2*), catalyzes the methylation of the repressive histone mark H3K27me3 (trimethylation of histone 3 lysine 27). Upon induction of *Xist-RepA*, the *PRC2* complex is recruited to Xi through the formation of *RepA-PRC2* complexes whereby the inactivation process is initiated.

While *Xist* dictates the formation of the Xi, an antisense transcript to *Xist*, *Tsix*, dictates the active X-chromosome (Xa) and prevents the action of *Xist*³⁰. *Tsix* is a 40 kb lncRNA which has been found to exist in both spliced and unspliced variants at approximately equal levels⁷⁷. It has been speculated that the different isoforms may hold different functions but this remains to be investigated⁷⁷. *Tsix* is initially expressed on both Xi and Xa. Upon the inactivation process, *Tsix* is induced on the future Xa, while it is repressed on the future Xi. *Tsix* has been shown to block XCI by acting negatively on *Xist* in several different ways. First, *Tsix*, like *Xist*, also interacts with *PRC2* and it is believed that *Tsix* acts as a decoy and

titrates PRC2 away from *Xist*⁷⁶. Upon suppression of Tsix, the level of free PRC2 increases and the RepA-PRC2 interaction is formed and consequently recruited to the chromatin. Second, Tsix is reported to form duplexes with Xist-RepA whereby the recruitment of RepA-PRC2 to Xa is prevented⁷⁸. Lastly, Tsix also appears to regulate transcription of *Xist*^{79,80}. Tsix has been shown to interact with the *de novo* DNA methyl transferase 3a (DNMT3a). The Tsix-DNMT3a complex is recruited to the *Xist* promoter and transcription is ultimately suppressed by the induction of DNA methylation⁸⁰.

Both the Xist and Tsix lncRNAs appear to act predominantly *in cis*. One could envision that Xist may diffuse from Xi and also initiate silencing of the Xa. However, it has been found that Xist is being tethered to the Xi by forming interactions with the Yin Yang 1 (YY1) protein⁸¹. YY1 is a bivalent protein capable of binding both DNA and RNA. YY1 interacts with Xi as well as the so-called Repeat C (RepC) element; a conserved C rich element of Xist repeated 14 times in tandem^{16,17}. It has been shown that YY1 is blocked from binding to Xa and the Xi-YY1-Xist complex is therefore tethered and forms exclusively on the Xi⁸¹. However, the underlying mechanism of differential binding of YY1 between Xi and Xa remains unidentified.

Finally, XCI is regulated by yet another lncRNA since it has been found that the lncRNA Jpx1 antagonizes the effect of Tsix^{82,83}. Jpx1 activates transcription of *Xist* on Xi by evicting CTCF from the *Xist* promoter⁸². CTCF is a transcriptional repressor⁸⁴ which initially binds to the *Xist* promoter thus blocking its transcription. Upon XCI, Jpx1 expression increases and Jpx1 binds CTCF. The Jpx1-CTCF complex is impaired in binding to *Xist* promoter on Xi and *Xist* transcription is consequently induced. Simultaneously, Tsix increases in expression on Xa subsequently preventing XCI.

Taken together, XCI clearly illustrates the complexity and diversity of lncRNAs and highlights that lncRNAs can mediate a wide variety of regulatory functions.

Trans-acting lncRNA: HOX Antisense Intergenic RNA

In addition to the lncRNAs involved in XCI, HOTAIR (HOX Antisense Intergenic RNA) is another lncRNA that has been extensively studied. Chang and colleagues investigated the transcriptional landscape of the four human *HOX* loci termed *HOXA*, *HOXB*, *HOXC* and *HOXD*. The *HOX* loci encode for 39 protein-coding genes, all of which are extensively regulated by chromatin modifications and functionally involved in specifying positional identities of cells. By using a tiling microarray approach, the presence of 231 transcribed loci which did not overlap with the protein-coding genes were revealed within these four loci⁵⁹. Indeed, the majority of the transcribed loci (231 of 407) were intergenic, non-coding and showed evolutionary conservation indicative of functional relevance. Among the identified lncRNAs, HOTAIR encoded on the *HOXC* locus has been subject of extensive analysis^{29,59,85,86}.

HOTAIR is a 2.2 kb lncRNA which is conserved to approximately 90% between human and mouse. Knockdown of HOTAIR does not affect the *HOXC* genes on chromosome 12 but instead reactivates several of the *HOXD* genes on chromosome 2. Mechanistically, HOTAIR recruits the PRC2 complex to the *HOXD* locus whereby transcriptional silencing is catalyzed by the formation of H3K27me3. Furthermore, silencing of the *HOXD* locus spreads over a nearly 40 kb large genomic region of the *HOXD* locus in a process which is still not well understood. Interestingly, the HOTAIR mediated recruitment of PRC2 is specific for the *HOXD* locus and does not affect any of the genes located on *HOXA*, *HOXB* or *HOXC*⁵⁹.

Subsequent studies have continued to dissect the function and molecular characteristics of HOTAIR. Genome-wide pulldown of HOTAIR using a tiling method called Chromatin Isolation by RNA Purification (ChIRP) identified 832 binding sites of HOTAIR⁸⁵. Similar to *HOXD*, these sites highly overlapped with the chromatin mark H3K27me3. HOTAIR appears to be highly enriched within short and specific regions containing sequence features with Guanine-Adenine-rich (poly-purine) elements, while PRC2 and H3K27me3 spread over broad domains. It has been speculated that HOTAIR is the initiation factor of silencing and recruitment of PRC2. This is indeed supported by the notion that HOTAIR is still bound to chromatin in the absence of PRC2, while PRC2 is not recruited to the chromatin in the absence of HOTAIR⁸⁵.

Further characterization of the HOTAIR transcript has revealed yet another functional property of HOTAIR, namely the presence of two different protein binding domains⁸⁶. While the 5' end of HOTAIR binds and recruits PRC2, the 3' end binds to another histone modification complex, CoREST/REST. The CoREST/REST complex contains the H3K4me2 demethylase LSD1⁸⁷. H3K4me2/3 correlates with actively transcribed regions⁸⁸ and its demethylation is required for proper repression of HOX genes^{89,90}. Consequently, HOTAIR functions as a molecular scaffold for two different modules, bringing PRC2 and CoREST/REST together. Methylation of H3K27me3 is therefore subsequently coordinated with the demethylation of H3K4me2^{86,89}.

In conclusion, HOTAIR and Xist are two extensively studied lncRNAs that have served as pioneering examples on functional investigations of lncRNAs. Through the recruitment of PRC2 to distinct loci and their involvement in chromatin remodeling, the functional properties of these lncRNAs are to some extent overlapping. It is also interesting to notice that while both Xist and HOTAIR initially recruit PRC2 to distinct loci, the repressive complex later spreads over larger regions. However, there are distinct characteristics as well. While Xist primarily functions *in cis*, HOTAIR functions *in trans* at many chromosomal regions.

Imprinting: Insulin-like growth factor 2 receptor

Imprinting is a phenomenon that refers to genes that are only expressed from either the maternal or the paternal allele. The *Igf2r* (Insulin-like growth factor 2 receptor) is a paternally imprinted locus encoding for the transmembrane receptor Igf2r, which functions by binding

of the Insulin-like growth factor II (*Igf2*) at the cell surface. Imprinting of the paternal *Igf2r* was first shown to require a CpG region within the second intron of *Igf2r* (Figure 4)⁹¹. This region was later revealed to be the promoter for the asRNA transcript *Airn* (antisense Insulin-like growth factor receptor)⁶⁵. The transcription of *Airn* is suppressed by DNA methylation on the maternal chromosome (Figure 4A), while it remains active on the paternally inherited chromosome (Figure 4B). *Airn* acts *in cis* on the paternal chromosome and induces transcriptional silencing by promoting DNA methylation of the *Igf2r* promoter^{92,93}. *Airn* is predominantly expressed as an unspliced and nuclear localized 108 kb lncRNA. However, several spliced isoforms that are exported to the cytoplasm have also been reported but their function remains unknown (Figure 4C)⁹⁴. Recent studies suggest that imprinting is mediated by the actual transcription of *Airn*, not the primary RNA transcript, consequently causing transcriptional interference that is thought to reduce the recruitment of pol II to the *Igf2r* promoter⁹³. It has also been reported that *Airn* interacts with G9a, a histone methyltransferase catalyzing the formation of the repressive histone mark H3K9me2. However, this appears to be dispensable for inactivation of *Igf2r*⁹⁵.

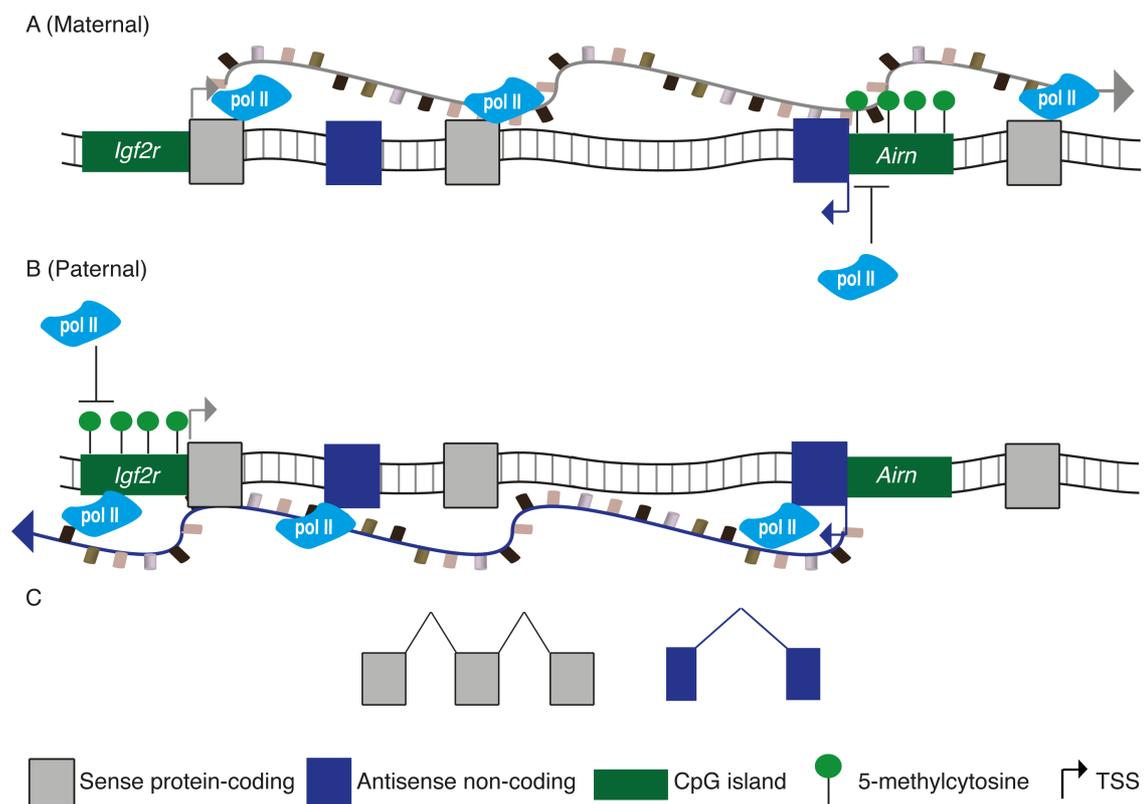


Figure 4 Imprinting of the *Igf2r* locus. **(A: Maternal)** *Igf2r* is actively being transcribed and the asRNA promoter (*Airn*) localized within the second intron of *Igf2r* is epigenetically inactivated. **(B: Paternal)** The *Airn* asRNA promoter is transcriptionally active, consequently being transcribed over the promoter of *Igf2r* thereby causing transcriptional interference of *Igf2r*. The paternal *Igf2r* is silenced by methylation of the promoter-associated CpG island. **(C)** *Airn* is also spliced but no function has yet been described for this isoform.

Epigenetic remodeling: Antisense Non-coding RNA in the INK4 Locus

The tumor suppressor gene cluster *INK4B-ARF-INK4A* is localized at the human chromosomal region 9p21. This region encodes three critical tumor suppressor genes, p15^{INK4B}, p16^{INK4A} and p14^{ARF}, which have been found inactivated through homozygous deletions as well as epigenetic mechanisms in many human tumors. p15^{INK4B} and p16^{INK4A} are two cyclin-dependent kinase (CDK) inhibitors involved in inhibiting cell cycle progression, whereas p14^{ARF} increases the levels of p53 by binding to the p53-inhibiting protein MDM2. Inactivation of this locus in tumors thus disrupts both the pRb and p53 tumor suppressor pathways. In addition to the protein-encoding transcripts *INK4B-ARF-INK4A*, this region has also been found to encode the asRNA transcript ANRIL (Antisense Non-coding RNA in the INK4 Locus)⁹⁶ spanning a region of ~125 kb. It contains 19 exons that are frequently alternatively spliced and numerous isoforms have been reported. ANRIL is transcribed from the *p14^{ARF}* bidirectional promoter and has been shown to regulate transcription of both *p15^{INK4B}* and *p16^{INK4A}*.

The capacity of ANRIL to induce chromatin remodeling at the *INK4B-ARF-INK4A* locus was originally described in 2008 by Yu *et al*²¹. High expression of ANRIL was found to correlate with low expression of p15^{INK4B} in AML as well as in acute lymphoblastic leukemia (ALL). Moreover, ectopic expression of ANRIL in several human cell line models as well as mouse embryonic cells caused suppression of *p15^{INK4B}* through the formation of heterochromatin and induction of H3K9me2, but was found independent of DNA methylation. Functional investigations of ANRIL have been addressed in a number of subsequent papers and a role for silencing of both *p15^{INK4B}* and *p16^{INK4A}* has been reported^{22,97}. ANRIL recruits the PRC2 complex, containing of the proteins EZH2, SUZ12 and EED, which initiates the silencing⁹⁷. SUZ12 interacts directly with ANRIL whereby EZH2 catalyzes the formation of H3K27me3. Next, PRC1 containing BMI1/MEL18, mPh1/2, RING1A/B, and Chromobox 7 (CBX7) is recruited to the site. The CBX7 protein recognizes the H3K27me3 mark whereby long-term silencing is established through the activity of RING1A, which mediates monoubiquitination at histone H2A Lysine 119 (see section 3.1.8.1 for further details of PRC1 and PRC2).

miRNA target site: Beta-secretase-1 asRNA

The Beta-secretase-1 asRNA (BACE1-AS) was identified by the FANTOM3 project as a sense:asRNA pair conserved between human and mouse¹⁰. Elevated levels of BACE1-AS, as well as the protein-coding sense transcript *BACE1*, were found in individuals with Alzheimer's disease and a role for the asRNA has been suggested in progression of the disease^{69,98}. Mechanistically, it was reported that BACE1-AS stabilizes the expression of the *BACE1* mRNA by forming RNA:RNA interactions⁶⁹. The sense:asRNA interaction interferes with a miRNA target site (miR-485-5p) in the *BACE1* transcript, whereby the miRNA mediated suppression of *BACE1* is lost. Noteworthy, this interaction and regulatory interplay does not occur in the 3' UTR of the *BACE1* transcript where miRNA targeting normally occurs. Instead, the competitive mRNA-asRNA-miRNA interplay was reported to take place within the protein-coding ORF of *BACE1*.

Splicing: Zinc finger E-box-binding homeobox 2

The zinc-finger E-box-binding homeobox 2 gene (*ZEB2*) is a transcriptional suppressor of the cancer-associated gene E-cadherin. The expression of *ZEB2* has been linked to several different forms of cancer including breast⁹⁹ and gastric cancer¹⁰⁰ and its expression has been found to associate with the presence of an asRNA, *ZEB2-AS1*^{23,101}. *ZEB2-AS1* overlaps the 5'UTR of *ZEB2* and the sense:asRNA overlap is conserved between mouse and human¹⁰¹. The 5' UTR of *ZEB2* contains an intron where an internal ribosome entry site (IRES) is located. This intron is normally spliced away but the splice site is masked upon the induction of *ZEB2-AS1*, consequently increasing the translation of *ZEB2* mRNA through initiation at the IRES element.

2.1.6.2 Non antisense RNA mediated regulation

The majority of studies on lncRNAs so far have focused on their regulation of chromatin structure and epigenetic remodeling, but a few recent reports have highlighted a functional role beyond chromatin. Two reports suggest lncRNAs to be involved in post-translational regulation of proteins by affecting their phosphorylation. While the lncRNA *PVT1* protects the oncoprotein *MYC* from being phosphorylated and degraded⁶², the lnc-DC (lncRNA Dendritic cells) protects *STAT3* (Signal transducer and activator of transcription 3) from being inactivated by de-phosphorylation⁶¹ (see details below). In addition, there is an emerging role for lncRNAs to regulate the stability of mRNAs by the formation RNA:RNA interactions at Alu rich elements (see below)¹⁰².

lncRNA mediated regulation of phosphorylation

The *MYC* oncogene is localized on the chromosomal region 8q24.21, a region frequently amplified in various human tumors. In addition to *MYC*, the protein-encoding *GSDMC* and *CCDC26* genes as well as the lncRNA gene *PVT1* are found within this region. While *MYC* has clear oncogenic properties, the functional relevance of the other genes has remained unclear. A fascinating interplay between *MYC* and the lncRNA *PVT1* was recently revealed⁶². A number of different mouse models with low copy number gain of the 8q24.21 region were generated: *gain(Myc)*, *gain(Pvt1, Ccdc26, Gsdmc)* and *gain(Myc, Pvt1, Ccdc26, Gsdmc)*. Interestingly, *gain(Myc)* did not show any induction of tumor formation and neither did *gain(Pvt1, Ccdc26, Gsdmc)*. However, *gain(Myc, Pvt1, Ccdc26, Gsdmc)* showed increased penetrance of tumor initiation compared to the other genotypes. This implied the presence of other genetic elements than *MYC* alone within the 8q24.21 region with oncogenic properties that amplify the pro-tumorigenic effect of *MYC*. The authors revealed *PVT1* to be this factor⁶². Phosphorylation of *MYC* on threonine 58 (Thr58) has been shown to regulate the stability of the *MYC* protein by decreasing its half-life¹⁰³. *PVT1* was found co-expressed and to interact with *MYC*, consequently protecting *MYC* from phosphorylation on Thr58 in a process still not completely identified. Interestingly, publicly available databases showed that >97% of tumors with increased 8q24 copy number had increased copy number of both the *MYC* and *PVT1* genes⁶².

A second example of post-translational modification mediated by lncRNAs was reported by Wang *et al*⁶¹. They investigated the transcriptome during differentiation of human dendritic cells (DC) and revealed a novel lncRNA named lnc-DC. Lnc-DC was found only modestly evolutionary conserved but interestingly induced upon DC differentiation in both human monocytes as well as in mouse bone marrow cells. Histone marks (e.g. H3K4me3) from the ENCODE data reinforced their findings and supported transcriptional activity at the genomic region encoding lnc-DC¹². Moreover, the induction of lnc-DC was found to be mediated by the transcription factor PU.1, which is a key regulator of DC differentiation^{104,105}. Knockdown of lnc-DC did not influence any of the nearby genes. Instead, biotin pulldown of lnc-DC revealed its association with STAT3. STAT3 is a TF, which in response to cytokines and growth factors is phosphorylated by receptor-associated kinases. Upon phosphorylation, STAT3 forms hetero- and homodimers and translocates to the nucleus where it activates transcription of target genes. Importantly, STAT3 is known to be involved in DC differentiation¹⁰⁶. Mechanistically, lnc-DC does not induce phosphorylation of STAT3 but instead suppresses de-phosphorylation and inactivation of STAT3. Lnc-DC binds to and blocks the function of the protein tyrosine phosphatase SHP-1, which normally de-phosphorylates STAT3. The level of active phosphorylated STAT3 is thus consequently increased (STAT3 and SHP-1 is described in greater detail in section 3.7).

lncRNA-mediated regulation of mRNA stability

Alu elements are among the most common sequences in the human genome. There are approximately 1 million copies of Alu elements covering 10% of the human genome¹⁰⁷. Alu elements belong to the SINE subtype of retrotransposones and are ~300 bp. The name Alu was given due to the presence of the recognition site for the Alu restriction enzyme (AGCT)¹⁰⁸. It is believed that Alu elements are derived from duplications of the 7SL gene, which is part of the ribosome complex. The Alu elements are found within genes, introns, 3'UTRs and intergenic regions and have been reported to be involved in the regulation of gene expression (reviewed in¹⁰⁹).

Staufen1 (STAU1) mediated mRNA decay (SMD) is a process where the STAU1 protein binds to double-stranded hairpin RNA structures in the 3'UTR of mRNAs. This interaction is involved in mRNA decay by de-stabilizing the targeted mRNAs¹¹⁰. Chenguang *et al* reported that STAU1 is not only recruited to hairpin RNA structures. It was found that Alu elements in the 3'UTR of mRNAs could form similar structures by interacting with Alu elements of lncRNAs. These double-stranded RNA:RNA interactions were shown to be sufficient to recruit STAU1 and to initiate the SMD pathway¹⁰².

The authors thoroughly investigated one of those lncRNAs in detail (named 1/2-sbsRNA1) and its interaction with the protein-coding mRNAs *SERPINE1* and *ANKRD57*. The interaction between the 1/2-sbsRNA1 and these protein-encoding mRNAs was indeed found to reduce amounts of the mRNAs, but not of the lncRNA. Although limited to a single lncRNA, it is tempting to speculate that this regulatory pathway is widespread. Knockdown

of STAU1 in HeLa cells increased the expression of ~1.6% of all expressed mRNAs. Among those, 13% contained a single Alu element in their 3'UTR¹⁰². In addition, the authors reported that 378 lncRNAs contained similar Alu elements but their interplay with Alu containing mRNAs remains to be investigated¹⁰².

2.1.7 Transcribed RNA = functional RNA?

As discussed above, thousands of lncRNAs have been and reported but only a handful of those have been assigned function and the functional relevance of all of these transcripts has become a matter of debate. While the recent ENCODE publications claim 80% of the genome to be functional (defined as either encoding for RNA or having a reproducible signature for chromatin structure or binding of protein)⁴⁹, others refer to this as highly overestimated¹¹¹. Some of the major concerns about the functional importance of lncRNAs range from lack of conservation¹¹² to leakiness and stochastic initiation of pol II transcription¹¹³. In addition, functional predictions of lncRNAs have proven difficult and functional investigations of individual lncRNAs are still very limited.

In a study by Mercer *et al*¹¹⁴, a subset of lncRNAs was investigated in terms of their expression and localization in mouse brain tissue. Many lncRNAs displayed highly specific expression patterns in different subcellular compartments, cell types and neuroanatomical regions. Such distinct pattern of lncRNA expression argues for a highly regulated transcriptional process rather than leaky and stochastic expression. Although this does not necessarily imply functionality of the transcripts, it may argue against lncRNA being an artifact of uncontrolled and leaky transcriptional initiation.

While evolutionary conservation most often indicates functional importance, absence of sequence conservation does not necessarily imply lack of function (reviewed in^{115,116}). Whereas mRNAs encode for proteins, lncRNAs are involved in a wide range of processes such as epigenetic remodeling²¹, post-translational modifications⁶² as well as interactions between RNA:protein⁸⁶, mRNA:lncRNA¹⁰² and miRNA:lncRNA⁵⁷. Due to the diversity in function of lncRNAs, it may be reasonable to believe that lncRNAs have different evolutionary constraints than protein-coding mRNAs. Here, the evolutionary conservation of lncRNAs will be discussed further by highlighting some of the previously discussed examples: *Xist/Tsix*, *Airn* and HOTAIR.

The importance of *Xist* and *Tsix* in XCI has been well documented in mice and the presence of *XIST* and *TSIX* has also been verified in the human genome¹⁶. The function of mouse *Xist* and human *XIST* as well as the exon-intron structure is very similar but the overall sequence identity is surprisingly low, being only around 60%¹¹⁷. However, it has been found that certain repetitive elements such as the RepA sequence, which binds and recruits PRC2, are relatively well-conserved^{17,118}. Thus, stretches of poorly conserved sequences link more conserved repetitive elements together into a functional lncRNA. It is therefore believed that

Xist/XIST is predominantly conserved on the level of structure, and not on the level of sequence.

In contrast to Xist/XIST, the function of human TSIX is surprisingly not orthologous to mouse¹¹⁹. Although TSIX is also present in the human genome, the transcript is not involved in asRNA-mediated suppression of *XIST*. The function of human TSIX in XCI remains to be explored in human cells.

Similar observations as TSIX have also been seen for the asRNA Airn¹²⁰. While Airn mediates imprinting of the paternal *Igf2r* in mice⁹²⁻⁹⁴, no such function has been reported in humans. Although AIRN is present in the human genome, its involvement in imprinting has not been established. Interestingly, human AIRN localizes at the same genomic locus as in mouse, but asRNA transcription over the *IGF2R* promoter, which is critical for mouse imprinting⁹³, has not been observed in human cells. A mouse ortholog has also been found for the lncRNA HOTAIR (mHotair)¹²¹. Whereas human HOTAIR consists of six exons, mHotair only encodes for two. Moreover, the 5' end of HOTAIR, which has been reported to bind to PRC2⁵⁹, is missing in mHotair and the sequence conservation is overall poor. In support of this, absence of mHotair only showed modest, if any, changes on the chromatin level in mouse¹²¹ while human HOTAIR has been associated with chromatin changes of hundreds of loci⁸⁵.

Taken together, although recent reports suggest many lncRNAs to be evolutionary conserved, the examples discussed here clearly illustrates that careful considerations should be taken before assuming that orthologous genes always have orthologous functions.

2.1.8 Effector proteins of lncRNAs

Although there is a huge repertoire of lncRNAs, their functional interplay with proteins remains relatively poorly studied. lncRNAs may interact directly with proteins, serving as guiding molecules and scaffolds. While HOTAIR serves as a molecular scaffold for the PRC2 and the LSD1/REST/coREST complexes, 1/2-sbsRNA1 guides STAU1 mediated mRNA decay. The proteins reported to interact with lncRNAs continue to grow in number. Here, two mediators of lncRNA function, PRC2 and DNMT3a, both involved in epigenetic regulation and relevant for the studies in this thesis, are discussed in greater detail.

2.1.8.1 Polycomb repressive complex

The polycomb group (PcG) of proteins were first identified in *Drosophila melanogaster* as repressors of *Hox* genes. *Hox* genes are a set of transcription factors that are conserved among all vertebrates with bilateral symmetry. The PcG proteins are involved in epigenetic silencing and classified into PRC1 and PRC2. PRC2 is generally considered to be the initiation complex while PRC1 is described as the maintenance complex.

The core components of PRC2 are EZH2, Suppressor of zeste 12 (SUZ12) and Embryonic ectoderm development (EED). EZH2 is the catalytic component of PRC2 and functions as a methyltransferase by catalyzing di- and trimethylation of H3K27 (H3K27me2/3)¹²². The PRC2 complexes containing the homolog EZH1 have low catalytic activity. Methylation of H3K27 is a stepwise process; H3K27me2 is performed on monomethylated H3K27 (H3K27me1) and H3K27me3 results from methylation of H3K27me2¹²³⁻¹²⁵. EZH2 only catalyzes the formation of H3K27me2 and H3K27me3 while the mechanism for the initiating H3K27me1 remains unclear.

PRC1 complexes consist of BMI1/MEL18, mPh1/2, RING1A/B and a chromodomain (CBX). The H3K27me3 histone mark is recognized by CBX whereby the maintenance complex PRC1 is recruited to the site. The ubiquitin ligase RING1B triggers ubiquitylation of Histone 2A at lysine 119 (H2AK119Ub), which consequently blocks transcription and induces long term silencing.

In *Drosophila melanogaster*, the presence of DNA elements, named Polycomb response elements (PRE), have been revealed to recruit PRC2 to targeted loci. In contrast, no similar elements have been found in mammals. Neither PRC1 nor PRC2 have any specificity in their DNA binding. Therefore, an unresolved question has for a long been how these chromatin-remodeling complexes are recruited to a targeted locus. The observation in 2007 that HOTAIR⁵⁹ and, soon thereafter in 2008, that the RepA element of Xist⁷⁶ interact with PRC2 pinpointed that RNA could be the guiding molecule that recruits the PRC2 complex to specific places in the genome. Indeed, these studies were followed up by genome-wide approaches for the identification of RNAs interacting with PRC2 and it was discovered that up to 20% of the previously identified lincRNAs interact with PRC2, possibly acting as guiding molecules¹²⁶. In support of their findings, targeting of lincRNAs in human cell lines generated similar phenotypes as targeting of PRC2¹²⁶.

A role for RNA being the guide molecule for PRC2 is thus emerging, and since the initial reports of HOTAIR⁵⁹ and Xist⁷⁶, numerous lincRNAs have now been reported to interact with and guide PRC2 in a similar manner. However, how the interaction between PRC2:RNA is formed and exactly how the RNA subsequently guides this complex to a targeted locus remains poorly understood. While some reports pinpoint a role for short hairpin structures within the RNA⁷⁶, others suggest PRC2 and, in particular, EZH2 have high affinity for RNA, but in contrast low specificity^{127,128}. Davidovich *et al* showed promiscuous binding of EZH2 to RNA with low sequence specificity, and therefore put into question whether RNA forms specific interactions with PRC2¹²⁸. Interestingly, PRC2 appeared to have higher affinity for long rather than short RNAs. Similarly, Kaneko *et al* showed preferential binding of PRC2 to 5' ends of RNAs, resulting in recruitment to actively transcribed promoters without triggering epigenetic silencing¹²⁷. These two studies clearly raise more questions about the interplay between PRC2:RNA, specifically how this interaction is formed and what dictates whether a locus undergoes epigenetic silencing or not. A model where RNA is ubiquitously recruiting

PRC2 to targeted locus, where PRC2 has the capacity to screen and sense whether the locus is supposed to be active or inactive has been proposed^{127,128}. The observations by Davidovich *et al*¹²⁸ and Kaneko *et al*¹²⁷ are very interesting but more studies will clearly be required in order to better understand the widespread recruitment of PRC2 by lncRNAs.

Finally, intriguing observations have revealed that the PRC2 complex and its subunit EZH2 interacts with DNMT3a¹²⁹. This is suggestive of interplay between histone modifications and DNA methylation where the two processes could act in concert.

2.1.8.2 DNA methyl transferases

Three DNA methyl transferases with catalytic activity have been identified; DNMT1^{130,131}, DNMT3a and DNMT3b¹³². DNMTs are involved in cellular DNA methylation by catalyzing the conversion of cytosine to 5-methylcytosine. DNMT3a and DNMT3b add methyl groups to unmodified DNA and are therefore called *de novo* DNMTs¹³³. In contrast, DNMT1 localizes to replication foci and preferentially methylates hemi-methylated DNA during cell division and is therefore considered as the maintenance DNMT. DNA methylation of CpG dinucleotides is an important epigenetic regulatory mechanism that is associated with gene silencing and has been linked to important processes such as XCI, development, genomic imprinting as well as cancer development (reviewed in¹³⁴⁻¹³⁶). DNMT1 and DNMT3b are both indispensable for embryonic development while knockout mice for DNMT3a die shortly after birth^{133,137}. Although essential, very little is known about the mechanisms by which DNMTs are governing the establishment and maintenance of DNA methylation.

CpG dinucleotides are generally underrepresented in mammalian genomes but found enriched at gene promoter regions and at the 5' end of genes, forming so-called CpG islands¹³⁸. Approximately 70% of all promoters contain CpG islands^{139,140}. The definition of a CpG island, although arbitrary, is generally considered a stretch of DNA where the presence of C+G is more than 55% over a region of 500 bp¹⁴¹. In general, the large majority of the non-promoter associated CpG dinucleotides are methylated. Interestingly, these CpGs dinucleotides are often hypomethylated upon cancer development while CpG islands at promoters are hypermethylated (reviewed in¹⁴²). However, the molecular mechanisms underlying this switch are not well understood.

The mechanisms for the recruitment of DNMTs to targeted loci are today mostly unknown and these proteins do not contain any domain conferring DNA specificity. One way of recruiting DNMTs to specific DNA loci may be through the capacity to read the histone code. As discussed above, DNMT3a and EZH2 have been reported to interact, thus allowing for synchronization in methylation of histones (H3K27me3) and methylation of DNA¹²⁹. Genome-wide analysis presents a strong inverse correlation between the localization of DNMT3a and H3K4me3, a histone mark representing active promoter regions^{143,144}. Additionally, it has also been found that DNMT3a associates with the histone deacetylase 1 (HDAC1)¹⁴⁵⁻¹⁴⁷ as well as the H3K9 methylase G9a¹⁴⁸. Altogether, it appears likely that

histone modifications and DNA methylation are tightly linked events and coordinated through a number of different molecular mechanisms.

A number of recent studies highlight a potential role for RNA in the regulation of locus specific DNA methylation. DNMT3a¹⁴⁹, DNMT3b¹⁴⁹ as well as DNMT1¹⁵⁰ have all been shown to interact with RNA. DNMT3a and DNMT3b were initially demonstrated to interact with small interference RNAs (siRNAs)¹⁴⁹. Likewise, in a process called transcriptional gene silencing (TGS), it was shown that small RNAs, double stranded as well as single stranded, could be directed against actively transcribed genes whereby transcription is suppressed through epigenetic silencing^{151,152}. This phenomenon had not been observed in human cells at the time and was intriguing since it demonstrated a possible role of RNA in mediating chromatin remodeling. Subsequent studies revealed DNMT3a to be recruited to the targeted locus¹⁵² in a process which initiated long-term silencing and was stable even after the withdrawal of the small hairpin RNA (shRNA)¹⁴⁷.

In a recent study by Di Ruscio *et al*, genome-wide analysis was carried out for the RNA interactome with DNMT1 and it was found that DNMT1 interacts with numerous promoter overlapping non-coding transcripts¹⁵⁰. Whereas the majority of studies on lncRNAs to date have focused on the recruitment of epigenetic remodelers to targeted locus, this study surprisingly showed that the interactions between DNMT1:RNA prevented DNMT1 from maintaining DNA methylation. The promoter overlapping RNAs functioned as decoys for DNMT1, and an inverse correlation of DNMT1:RNA and DNA methylation was noted¹⁵⁰. In contrast to these observations it has also been found that promoter overlapping ncRNAs can form DNA:RNA triplex structures at the promoter regions of rRNA genes¹⁵³. In this case, the DNA:RNA interaction is recognized by DNMT3b whereby *de novo* methylation is initiated, consequently triggering transcriptional suppression.

The model of DNMT3a/b serving as *de novo* DNMTs and DNMT1 as the maintenance DNMT has recently being challenged by a number of studies. DNMT1 has been shown to have *do novo* activity in cancer cells¹⁵⁴. It has also been presented that DNMT3a mediates transcriptional suppression also in the absence of its catalytic activity^{145,155}. This appears to take place through deacetylation of histones through interactions with HDAC1 as well as TFs^{145,155}. One reported example is the interaction between p53 and DNMT3a, which affected p53-dependent induction of the tumor suppressor gene p21¹⁵⁵. In this case, the suppression of p21 was active also upon mutations that inactivated the catalytic activity of DNMT3a. Thus, it appears that DNMT3a counteracts the transcriptional activation of p53 independently of DNA methylation. Besides, it is interesting to note that DNMT3a interacts with sequence specific TFs, which indirectly mediate sequence specific recruitment of DNMT3a^{145,155}. The function of DNMT3a that is independent of DNA methylation remains to be studied more thoroughly, but it is tempting to speculate whether there is a switch and co-factors needed in order to trigger DNA methylation. In this scenario, DNMT3a could be ubiquitously recruited to targeted loci without primarily triggering its catalytic activity, possibly in a similar

mechanism as recently presented for PRC2 and as discussed above^{127,128}. Finally, although methylated CpGs are considered to be a suppressive mark, it has also been found that gene-bodies of actively transcribed genes are enriched for methylation^{156,157} and it has been speculated that such methylation may prohibit unspecific and spurious initiation of transcription^{156,158}. Moreover, exons present higher levels of methylation than introns and a sharp contrast of methylation at the intron-exon junction has been reported and a function related to splicing has been proposed¹⁴³.

Taken together, although DNA methylation is a key epigenetic signature involved in gene regulation, the underlying mechanisms for induction, maintenance and control of sequence specific methylation is largely unknown. In the relatively few studies carried out to date, a role for RNA as a regulator of this process is emerging but more studies are needed in order to increase our understanding.

2.2 PSEUDOGENES

The term pseudogene was introduced in 1977 when Jacq and coworkers discovered a truncated version of the 5S DNA that still contained homology with the active gene¹⁵⁹. A pseudogene refers to a duplicated gene which has lost its protein-coding capacity through molecular events such as mutations causing frame shifts or premature stop codons. Pseudogenes have long been considered non-functional relics in the genome but a number of recent investigations have illuminated that they may have some exciting functional properties^{57,71}.

Depending on how the pseudogene was generated from its ancestral functional gene, also called the parental gene, the pseudogenes are divided into:

- Processed pseudogenes (Figure 5A)
- Unprocessed pseudogenes (Figure 5B)
- Unitary pseudogenes

Processed pseudogenes (Figure 5A) are derived from mRNAs which have been spliced, reverse transcribed and inserted into the genome at a new genomic locus. Consequently, promoter regions as well as other regulatory elements such as enhancers and introns of the parental gene have been lost. The pseudogene and its transcriptional activity are therefore highly dependent on whether it has integrated into a transcriptionally active or inactive genomic region. Due to the mRNA intermediate during the pseudogenization, processed pseudogenes often have poly-adenine features at their 3' end. This category of pseudogenes may also be partially spliced, thus still containing some remnants of parental introns.

Unprocessed pseudogenes (Figure 5B) are the result of gene duplications. In contrast to processed pseudogenes, unprocessed pseudogenes maintain introns as well as promoter elements and are therefore more likely to be transcriptionally active. Unprocessed pseudogenes may also be partially duplicated, meaning that only a distinct region of the parental gene has been duplicated.

Unitary pseudogenes refer to previously active genes that have become inactive through various mutations. Thus, the process of inactivation is similar to unprocessed pseudogenes without the prerequisite for the duplication process and the existence of a functional ancestral gene.

The number of pseudogenes in the human genome was initially estimated to be ~20,000¹⁶⁰. This estimate was largely based on the pilot phase of the ENCODE project covering chromosomes 21 and 22 (see section 3.1.3). The frequency of pseudogenes (201 in total) within these arbitrarily chosen regions was extrapolated to the entire human genome. However, the 1% of the human genome included in the ENCODE pilot phase turned out not

to be fully representative of the entire genome and the estimate has now been adjusted to around 14206 annotated pseudogenes according to the most recent update of GENCODE⁵³.

Among annotated pseudogenes, the vast majority identified are processed pseudogenes (10532) while 2942 are unprocessed and 161 are unitary pseudogenes. The excess of processed pseudogenes is believed to be the result of a burst in retrotranspositional activity in ancestral primates about 40 million years ago, which interestingly also coincides with the amplification of Alu elements^{161,162}. Although processed pseudogenes are more than three times as numerous as duplicated pseudogenes, this is not reflected with regards to transcriptional activity. As reported by the ENCODE Consortium, 876 processed and 531 unprocessed pseudogenes have at some point and in some tissue been detected to be transcribed into RNA⁵³. Noteworthy, while some genes only have a single pseudogene, some highly expressed housekeeping genes are associated with a large numbers of pseudogenes. For example, the ribosomal protein L121 has 143 pseudogenes and the glyceraldehyde-3-phosphate dehydrogenase GAPDH has 68 pseudogenes^{53,163,164}.

While originally considered only as ‘junk DNA’, previously underappreciated functions of pseudogenes are now emerging. However, the identification of pseudogenes is not only of interest due to their proven functional role in regulating its parental gene. The presence of pseudogenes may also interfere with investigations of the protein-coding counterpart due to the high sequence similarity and might for example interfere with analysis of transcriptional abundance as well as mutational analysis. Extra care and considerations should therefore be taken in the analysis of gene:pseudogene pairs.

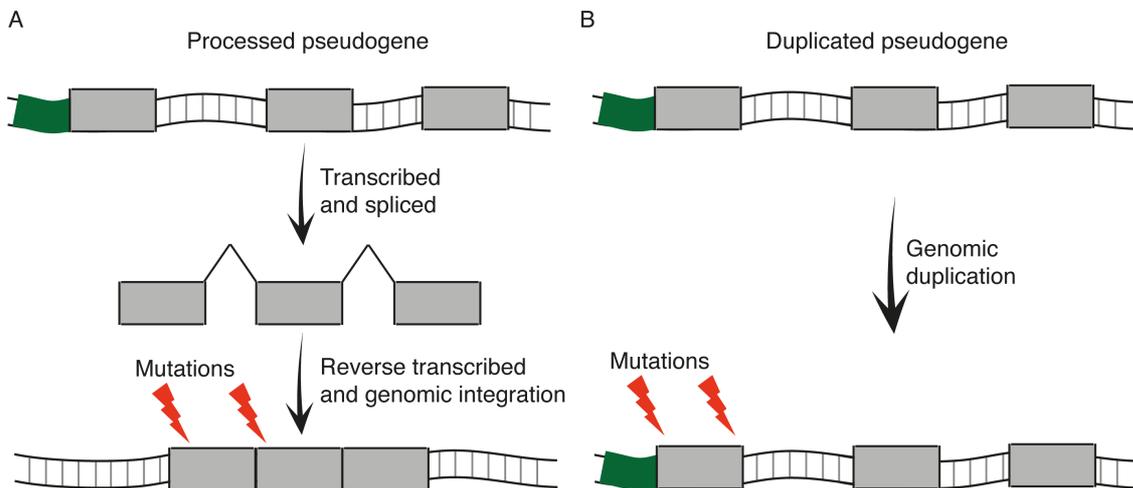


Figure 5 A schematic illustration of the generation of pseudogenes **(A)** Processed pseudogenes are reverse transcribed and integrated into new genomic locations. The promoter region of the protein-coding parental gene is lost upon this process and acquired mutations inactivate the protein-coding potential. **(B)** Unprocessed pseudogenes are duplicated from protein-coding parental genes. The promoter region is maintained and the duplicated pseudogene is therefore more likely to be transcriptionally active. The protein-coding potential is lost due to mutations.

2.2.1 Evolutionary conservation of pseudogenes

The common ancestors of mice and humans diverged approximately 75 million years ago. As discussed above, most human pseudogenes are the result of a burst in retrotranspositional activity that took place ~40 million years ago in ancestral primates^{161,162}. Interestingly, a similar event took place in mice ~36 million years ago¹⁶⁴. Thus, most pseudogenes (~60%) were created after the last human and mouse common ancestor, and are therefore lineage specific. In addition, studies focusing only on transcriptionally active human pseudogenes suggest that as few as 3% of those are conserved in mouse¹⁶⁵.

A critical point for investigations of pseudogenes will be to find good candidates to validate for function. Evolutionary conservation is often a valid starting point and a good indication of functional importance. Although few, 1,019 conserved pseudogenes have been identified between mice, non-human primates and humans with 195 of those also showing conservation at the sequence level⁵³. However, as noted, the conservation of pseudogenes among different species is generally low and may not be a valid approach for finding the most relevant candidates for investigations. An alternative approach could be to investigate the overall sequence similarity between the pseudogene and its protein-coding parental gene. However, this approach could also be misleading due the fact that the generation of most pseudogenes is a relatively recent evolutionary event. Thus, high sequence overlap between gene:pseudogene pairs may not necessarily reflect evolutionary constraints but rather expose the notion that these are recent evolutionary events.

In contrast, by identifying the corresponding 5'/3'UTR and coding-region, a pseudogene can be divided into different genomic regions and subsequent sequence analysis carried out over the different regions⁵³. By doing so, most pseudogenes were revealed to have comparable sequence identities in the different genomic regions. This would be expected if there is a lack of function and therefore no evolutionary constraints on that particular pseudogene. However, one subset of pseudogenes had higher sequence identity in the 3'UTR and another subset in the coding region. This suggests that mutations have been rejected and that some regions are under stronger evolutionary constraints than others in this gene set. In total, 998 pseudogenes showed such high sequence identity in the coding region compared to its parental gene while 36 showed higher sequence identity in the 3'UTR⁵³.

In summary, although the group of pseudogenes is divergent and poorly conserved among species, some show traces of evolutionary constraints suggesting functional importance. This group of pseudogenes generates a good platform for initial investigations.

2.2.2 Functional investigations of pseudogenes

Similarly to lincRNAs and asRNAs, the development of genome-wide platforms revealed an underappreciated complexity of pseudogenes. The fact that the ENCODE data revealed that many pseudogenes are being transcribed increased the interest in this subset of non-coding genes and inspired more functional studies⁵³. Although the presence of pseudogenes has been

known for ~30 years¹⁵⁹, functional investigations have been very limited. As early as 1984, the important tumor suppressor gene p53 was found to have a pseudogene¹⁶⁶. Other important regulatory proteins, many of which are well known for being involved in diseases, have indeed been reported to have pseudogenes including *PTEN* (described in 1998)¹⁶⁷, *BRAC1* (described in 2002)¹⁶⁸ and *OCT4* (described in 2005)¹⁶⁹.

In a study by Pain *et al*¹⁶⁹, *OCT4* was found to have six pseudogenes. Shortly thereafter, transcriptional investigations were undertaken and showed that at least two of those, OCT4pg1 and OCT4pg5, were transcriptionally active¹⁷⁰. Although no function was assigned at that point, the authors noted high expression in several cancer cell lines and also pointed out that the high abundance of pseudogene associated transcripts could interfere with studies of the protein-coding counterpart. Similarly, the *PTEN* pseudogene (*PTENpg1*, *PTENp1*, *PTENψ*) was also reported as highly expressed in some cell lines and human tissues¹⁷¹.

More than 10 years after its discovery, a functional role for the *PTENpg1* was reported⁵⁷. *PTENpg1* is a processed pseudogene which is truncated due to a missense mutation in the start codon for methionine, therefore preventing translation¹⁷¹. The pseudogene is highly homologous to *PTEN* with a sequence identity of >95%. The protein-coding *PTEN* is regulated by numerous miRNAs^{172,173} and in a significant study by Poliseno *et al*⁵⁷, it was found that *PTENpg1* had retained the same miRNA binding sites as *PTEN*. *PTENpg1* was found to possess a regulatory function through its ability to compete for miRNA binding, thus acting as a decoy for *PTEN* related miRNAs. High expression of *PTENpg1* thus soaked up and sponged miRNAs¹⁷⁴, consequently releasing *PTEN* from miRNA-mediated suppression. Contrarily, loss of expression of *PTENpg1* released miRNAs, which instead targeted *PTEN* causing suppression. In support of their findings, the expression of *PTEN* and *PTENpg1* correlated in human tissues as well as prostate tumor samples. Moreover, copy number losses of the *PTENpg1* locus in colon cancer correlated with decreased expression of *PTEN*. Deletion of the *PTENpg1* locus has also been reported in human melanoma¹⁷⁵. Unexpectedly, *PTEN* was found co-deleted in the majority of melanoma samples with lost *PTENpg1*, therefore suggesting a role for *PTENpg1* beyond acting as decoy for *PTEN* targeted miRNAs.

The studies on *PTENpg1* clearly illustrate a role for pseudogene-expressed transcripts in gene regulation and cancer development. The studies on *PTENpg1* were followed by genome-wide approaches for the identification of transcribed pseudogenes in cancer. Using a total of 293 samples, 248 cancer and 45 benign samples, a pipeline for the identification of pseudogene-associated transcripts was developed and over 1,500 actively transcribed pseudogenes were found¹⁷⁶. Some pseudogenes were highly tissue specific while others were ubiquitously expressed in most samples investigated. Although most pseudogenes were found expressed in both cancer and benign samples, an interesting subset of 248 showed specificity for cancer cells only¹⁷⁶. Noteworthy and as pointed out by the authors, the pipeline developed within this study did not allow identification of pseudogenes with very high sequence overlap with its parental protein-coding gene¹⁷⁶. *PTENpg1* was noted to be one such example.

2.2.3 *Competitive endogenous RNAs*

The findings by Poliseno *et al*⁵⁷ suggest a complex regulatory network where RNAs can be biologically active by “talking” to each other through their ability to bind and compete for miRNAs independently of the protein-coding capacity. mRNAs had previously been considered as passive targets of miRNAs but this model, named competitive endogenous RNAs (ceRNA)¹⁷⁷, suggests cross-talk between different RNA molecules. The ceRNA model may not only be applied to pseudogenes but could also be valid for other RNAs such as lincRNAs¹⁷⁸, asRNAs¹⁷⁹ as well as mRNAs¹⁸⁰. In addition, it is also suggested that pseudogenes may act as decoys for RNA binding proteins. As an example, the HMGA1 gene (high mobility group A1), which is associated with insulin resistance and type 2 diabetes, is stabilized by the α CP1 protein through protein:RNA interactions in the 3'UTR of HMGA1. Upon induction of its pseudogene, HMGA1-p, α CP1 translocates to the pseudogene whereby the stability of HMGA1 mRNA decreases¹⁸¹.

2.2.4 *Pseudogenes: a novel source of asRNAs?*

Although several reports suggest thousands of pseudogenes to be transcriptionally active^{53,176}, no such comprehensive data is available for pseudogenes transcribed in the asRNA direction. The investigations of such transcripts mainly rely on case-by-case studies. As proven by a limited number of studies^{71,182,183}, asRNAs expressed from pseudogenes, pseudo-asRNAs, are indeed potential regulators of their parental gene. Pseudo-asRNAs have greater flexibility and chances to evolve compared to *cis*-asRNAs, since there are no constraints from the protein-coding sense gene. They can therefore emerge into new important regulatory transcripts and exhibit an independent transcriptional regulation as compared to their ancestral gene. Upon integration into a new genomic location, a processed pseudogene could for example be fused with new genomic elements leading to the formation of transcripts with modified functional properties. Furthermore, it is also possible that a protein-coding gene originally lacking *cis*-asRNA transcription may acquire pseudo-asRNA transcription through the evolution of a new pseudogene-associated asRNA promoter¹⁸⁴. In an attempt to identify pseudo-asRNAs, expressed sequence tags (ESTs) were used and indeed, 87 pseudogenes were found to have pseudo-asRNAs¹⁸⁴. Among those, 15 parental genes showed evidence of asRNAs, suggesting that the sense:asRNA pair had existed prior the pseudogenization process and had been duplicated together. Similar to sense transcribed pseudogenes⁵³, different regions of the pseudo-asRNA were noted to harbor different degrees of sequence conservation. The sequence overlapping with the parental gene was shown to contain the lowest degree of mutations, suggesting there is selection for maintaining complementarity to the parental gene.

The first evidence of pseudo-asRNAs was presented in 1992 when Zhou *et al* showed the Human DNA Topoisomerase I Pseudogene to be associated with asRNA transcription¹⁸⁵. A few years later, in 1997, similar discoveries were reported also in mouse for the FGFR-3 pseudogene¹⁸⁶. However, no function was assigned for either of these transcripts.

The first functional characterization of a pseudo-asRNA was carried out on the neuronal nitric oxide synthase gene (nNOS) in the mollusk *Lymnaea stagnalis*¹⁸⁷. nNOS was found having a pseudogene, pseudo-nNOS, which was expressed in the central nervous system and involved in regulating the expression of nNOS. pseudo-nNOS was not expressed as an asRNA from the opposite DNA strand but instead contained a 150 bp long inverted repeat in the sense orientation. This repeat was found to form a duplex with nNOS and to prevent its translation by a mechanism still not identified¹⁸⁷.

AsRNAs derived from pseudogenes can also serve as substrates for the generation of endogenous siRNAs. RNA-seq data from mouse oocytes revealed the presence of siRNAs originating from pseudo-asRNAs^{188,189}. Unexpectedly, the dsRNA formation that serves as template for Dicer processing was not mediated by interactions between the pseudogene sense and asRNA transcripts. Instead, it appeared that the pseudo-asRNA formed a Dicer substrate with the mRNA of its protein-coding parental gene. The authors speculated whether these interactions can only take place in mouse oocytes due to the lack of dsRNA inducible protein kinase R (PKR)¹⁹⁰. A similar mechanism has also been proposed in human cells. Mapping of small RNAs from deep sequencing data identified a subset which aligned to pseudogenes¹⁹¹.

Finally, pseudo-asRNAs can also guide epigenetic remodeling complexes. One of the OCT4 pseudogenes, OCT4-pg5, was found having an asRNA transcript which acted as a negative regulator for the protein-coding gene *OCT4* as well as the OCT4-pg1⁷¹. Comparable to other *cis*-acting asRNAs⁹⁷, OCT4-pg5 caused epigenetic silencing through the recruitment of EZH2 and subsequently formation of H3K27me3 at the targeted loci⁷¹.

2.2.5 Technical considerations during functional investigations of pseudogenes

As described above, transcribed pseudogenes hold great promise as important regulatory molecules in gene regulation. Thousands have been validated as transcriptionally active, yet functional studies have to be undertaken in order elucidate their function. However, the investigation of pseudogenes, and in particular pseudo-asRNA transcripts, are associated with a great number challenges that should be considered during experimental investigations.

2.2.5.1 Quantitative analysis using PCR

When profiling the expression of a pseudogene, sequence homology between the pseudogene and its parental gene should be carefully considered. Expression profiling using either quantitative real-time-PCR (qRT-PCR) and/or semi-qRT-PCR is often a rational starting point in order to investigate expression and tissue specificity. While qRT-PCR may be the primary method of choice, primer design is often limited due to size limitations (~100-200 bp) in qRT-PCR analyses. In contrast, semi-qRT-PCR gives greater flexibility with regards to primer design as longer PCR products may be used. Preferably, mismatches between the gene/pseudogene should be placed in the 3' end of the primer in order to avoid unspecific amplification. The identity of the PCR product should also be confirmed for example by

cloning and subsequent sequencing. An alternative to sequencing would be analysis by restriction enzymes, assuming there are unique restriction sites in either the parental gene or the pseudogene.

In addition, primer design for sense:asRNA pairs could also face challenges. Aiming for non-overlapping sequences and strand specific splice junctions is the primary method of unique primer design, but not always feasible considering that most processed pseudogenes lack introns. However, in the presence of pseudo-asRNAs it is possible that the transcript has evolved to have unique splice junctions.

Although RNA-seq data provides several advantages allowing for an unbiased analysis of RNA transcripts, the data should preferentially be strand-specific. It should also be considered that some pseudogene-derived RNAs might be polyA negative and therefore not detected by using standard procedures for RNA-seq.

In conclusion, quantitative analysis of pseudogenes, sense as well as asRNA transcripts, may be a challenging task and needs to be addressed on a case-by-case basis.

2.2.5.2 Targeting of pseudogenes

It is often necessary to suppress the pseudogene of interest in order to carry out functional investigations, e.g. by using siRNAs, shRNAs or antisense oligos (ASOs). Similarly to the primer design, a major issue is the high sequence homology to its parental gene. The relatively few mismatches between the gene and the pseudogene should be used with great consideration. If using siRNAs or shRNAs, having mismatches in both the Ago2 cleavage site (nucleotides ~8-12) and in the seed sequence (nucleotides 2-7) is desirable to impart specificity. In order to minimize the risk of facing off target effects, it is also beneficial to design several siRNAs/shRNAs for each target. However, this may not always be possible if the sequence overlap is very high. An alternative approach may instead be to identify cell lines and model systems that are negative in the expression of the pseudogene of interest. By transfecting the siRNAs/shRNAs into those cell lines, no change in phenotype would be expected.

Moreover, the analysis of sense:asRNA pairs may require strand specific targeting. Preferably, targeting the sense or asRNA at a point where no overlap is present is optimal. However, this may not always be achievable if certain isoforms are expressed only at the sense:asRNA overlap. In such scenarios, single stranded ASOs may be used. Several designs for ASOs are available. In general, ASOs are chemically modified, single stranded DNA molecules which form DNA:RNA hybrids with the target sequence. This interaction activates the RNase H pathway, which is predominantly active in the nucleus¹⁹². Like siRNAs/shRNAs design, mismatches between the gene and pseudogene are optimally placed in the mid section of the ASO. It may be beneficial using ASOs when targeting nuclear expressed transcripts since it is believed that Ago2 is predominantly active in the cytoplasm. However, siRNA

mediated targeting of nuclear expressed transcripts has been successful and there are today numerous reports showing Ago2 to be active also in the nucleus^{151,193,194}.

2.2.5.3 Overexpression

Overexpression of the transcript of interest may be very informative and many different cloning constructs are available for this purpose. While some uses pol II promoters, e.g. the CMV promoter, others use pol III promoters, e.g. the U6 promoter. Pol II transcribed transcripts are polyA positive and likely to be expressed in the cytoplasm. In contrast, the pol III transcribed transcripts lack a polyA tail and transcription is stopped after a stretch of thymines. Moreover, the pol III promoter predominantly generates nuclear expressed transcripts¹⁹⁵. Consequently, depending on the desire to express different transcripts, localized in the nucleus or cytoplasm, polyA positive or negative, this may require different vectors.

Overexpression also gives the possibility of generating escape mutants. More precisely, the target site for the siRNA/shRNA/ASO may be mutated whereby the sequence is no longer targeted. This represents a good approach for controlling for off target effects.

Finally, if no effect is observed upon overexpression, this does not necessarily mean that the lncRNA is non-functional; rather, several aspects should be considered in such a scenario. First, ectopic overexpressing may not generate the correct localization or isoforms, and some lncRNAs may function only when expressed *in cis*. Moreover, there could be a secondary bottleneck factor that restricts the function of the lncRNA. For example, although the lncRNA is induced there may not be sufficient amounts of the targeted mRNA or associated effector protein, e.g. EZH2 or DNMT3a. This could potentially be circumvented with simultaneous induction of both the lncRNA as well as the protein.

2.3 CANCER: A lncRNA PERSPECTIVE

lncRNAs are currently being explored in cancer and several recent reports clearly support an important role for lncRNAs in neoplasia. Novel discoveries are constantly reported and an open mind is required during functional analysis, as it must be remembered that RNA is a multi-functional molecule. Several examples have described the ability of lncRNAs to regulate tumor suppressor genes (TSGs) and proto-oncogenes through various mechanisms. TSGs represent genes that protect cells from developing cancer, often by having repressive effects on the cell cycle and by promoting apoptosis. Contrarily, proto-oncogenes have the potential to promote cancer progression. Proto-oncogenes can become oncogenes, for example by acquiring activating mutations or increasing their expression and could consequently trigger cell growth and suppress apoptosis. The following represent examples that highlight the functional importance of lncRNAs in cancer development, frequently by disrupting the expression of TSGs and/or oncogenes.

2.3.1 The Hallmarks of cancer

Although cancer is a heterogeneous group of diseases, some features are unifying. In 2000, Hanahan and Weinberg summarized these properties into the six hallmarks of cancer¹⁹⁶. This model was later expanded based on our increased understanding of cancer¹⁹⁷. The original six hallmarks of cancer are: (1) sustaining proliferative signaling, (2) evading growth suppression, (3) enabling replicative immortality, (4) activating invasion and metastasis, (5) inducing angiogenesis, and (6) resisting cell death. lncRNAs have been reported to be involved in each and every one of these processes, and examples will herein be described.

2.3.1.1 Sustaining proliferative signaling

A very distinct characteristic of cancer cells is the need to constantly proliferate, even in the absence of external stimuli. This process is normally tightly controlled but can be overridden, for example through constitutive activation of growth promoting signaling pathways.

In a study screening for differential expression of lncRNAs in prostate tumors, the lncRNA PCAT-1 (prostate cancer associated transcript 1) was found induced in a subset of localized metastatic high-grade cancers¹⁹⁸. Differential expression of PCAT-1 has also been reported in colorectal cancer¹⁹⁹. Alterations of PCAT-1 by knockdown and overexpression confirmed involvement in proliferation¹⁹⁸. Functionally, PCAT-1 is suggested to be involved in repair of double-strand DNA breaks possibly by interacting with the 3'UTR of *BRCA1* and thereby trigger suppression of *BRCA1*²⁰⁰.

2.3.1.2 Evading growth suppression

In line with hallmark 1, tumor cells have to evade growth suppression. There are numerous cellular proteins with inhibitory effects on growth and proliferation. The TSGs *PTEN* and *p53* are two such proteins which are frequently inactivated by mutations, deletions and epigenetic changes. Not surprisingly, numerous lncRNAs have been reported to be involved in altering the expression of these genes. *PTEN* is regulated by the PTENpg1 sense transcript involving

both miRNA- and lncRNA-mediated regulation⁵⁷. Additionally, the CDK inhibitors p15^{INK4B}/p16^{INK4A} are suppressed by the previously mentioned asRNA ANRIL, in leukemia²¹ and prostate cancer²². Similarly, the CDK inhibitor p21 is also regulated by asRNA transcription⁷⁰. However, the p21-asRNA has not yet been reported dysregulated in cancer and its involvement in disease remains to be elucidated.

The critical tumor suppressor gene p53 is lost or mutated in over 50% of all human cancers and is probably one of the most thoroughly studied proteins²⁰¹. Very surprisingly, the *p53* mRNA was recently found to maintain protein coding-independent functions. MDM2 is a negative regulator of p53, which upon translocation to the nucleus binds to and triggers the degradation of p53 through its E3 ubiquitin ligase activity^{202,203}. However, upon genotoxic stress, the *p53* mRNA interacts with MDM2, an interaction which affects the nuclear translocation of MDM2. Consequently, this blocks the MDM2 mediated degradation of p53 and instead promotes the expression of p53²⁰⁴. In addition, *p53* is also regulated at the post-transcriptional level by its asRNA *WRAP53* (WD40 encoding RNA antisense to p53)⁶⁸ and miR-125²⁰⁵.

2.3.1.3 Enabling replicative immortality

The chromosome ends, termed telomeres, shorten upon each cell cycle and normal somatic cells can therefore only undergo a limited number of cell cycles²⁰⁶. Cancer cells have to overcome telomere shortening in order to gain replicative immortality and the vast majority of tumors do so by the gain of telomerase activity²⁰⁷. The telomerase complex contains the Telomerase Reverse Transcriptase (TERT) as well as the ncRNA component TERC (Telomerase RNA Component)²⁰⁸. TERT catalyzes reverse transcription of the telomere using the ncRNA TERC as a template.

2.3.1.4 Activating invasion and metastasis

The ability of cancer cells to undergo metastasis requires the capability to colonize distant tissue. This is a multistep process that often includes morphological changes and alterations in cell-cell interactions. This transition is often referred to as epithelial-mesenchymal-transition (EMT) and its reverse process mesenchymal-epithelial-transition (MET) (reviewed in²⁰⁹). The process often involves intravasation of cancer cells into blood and lymphatic tissue, followed by extravasation from vessels into target tissue and subsequently formation of the metastases. Several important factors have been identified in this cascade of events, one of which will be discussed with regards to lncRNAs: *E-caderin* (*E-cad*, also referred to as *CDH1*).

E-cad is involved in cell-cell junctions and is frequently downregulated in cancer and is in particular an important part of the EMT process (reviewed in²¹⁰). Several layers of lncRNA-mediated regulation have been reported for this protein. First, *E-cad* has a *cis* asRNA transcript that acts as a negative regulator of its transcription⁷⁰. Imbalance between the sense:asRNA transcription may therefore induce epigenetic silencing. Moreover, the ZEB2

protein is a transcriptional repressor of *E-cad*. Alternative splicing of *ZEB2* is mediated by the *ZEB2-AS1* (previously discussed in section 3.1.6.1)²³. Upon initiation of EMT, the expression of *ZEB2-AS1* is induced, which consequently increases translation of *ZEB2* and thus triggers transcriptional suppression of *E-cad*²³.

In addition to the lncRNAs involved in regulation of *E-cad*, HOTAIR has also been reported to be involved in metastasis (see section 3.1.6.1 for HOTAIR)²⁹. HOTAIR is frequently induced in primary breast tumors and metastasis and its expression is revealed to be a prognostic marker. The increased expression of HOTAIR mediates genome-wide epigenetic reprogramming in breast cancer metastasis²⁹. Moreover, enforced expression of HOTAIR increases invasiveness and metastasis while the depletion of HOTAIR reduces cancer invasiveness. Recently, expression of HOTAIR has also been associated with colorectal cancer²¹¹, hepatocellular carcinoma²¹² and gastric cancer²¹³.

2.3.1.5 Inducing angiogenesis

Upon growth and expansion of cancer cells, the formation of new blood vessels is needed in order to supply the tumor cells with oxygen and nutrients. The hypoxia inducible factor alpha (*HIF1 α*) is a transcription factor for hypoxia activated genes and it also stabilizes and increases the expression of p53. Several studies have reported the presence of an asRNA to *HIF1 α* , asRNA-*HIF1 α* (also called α HIF1), overlapping the 3'UTR of *HIF1 α* ²¹⁴⁻²¹⁶. It is thought that asRNA-*HIF1 α* is a negative regulator and uncovers AU-rich elements in the 3'UTR of the *HIF1 α* -mRNA, consequently increasing degradation of the protein-coding *HIF1 α* ²¹⁶.

2.3.1.6 Resisting cell death

In line with hallmarks I and II, cancer cells also need to avoid cell death in order to survive various types of stress and expand. In a study by Khaitan *et al*²¹⁷, the expression of lncRNAs was investigated in a stage III melanoma cell line compared to melanocytes and normal skin. In total, 77 lncRNAs with differential expression were identified and the differential expression of four of those was further confirmed in independent primary metastatic melanoma samples. One of the lncRNAs, *SPRY4-IT1*, was characterized in greater detail and knockdown studies revealed its inhibitory effect on apoptosis. *SPRY4-IT1* is derived from the intron of the protein-coding gene *SPRY4*. Interestingly, *SPRY4* is suggested to have tumor suppressive functions by impairing the formation of GTP-RAS, resulting in suppression of MAPK signaling (see section 3.6 for more information about MAPK signaling)^{218,219}. This is in clear contrast to the tumor promoting effect of *SPRY4-IT1* and the complex interplay between *SPRY4* and *SPRY4-IT1* remains to be further investigated. Although the molecular mechanism of *SPRY4-IT1* is still unknown, the data suggests *SPRY4-IT1* serves as an oncogene by suppressing apoptosis and likewise supports a role for lncRNAs in this hallmark of cancer.

2.3.2 Mutations of *EZH2* and *DNMT3a*

As noted above, several lncRNAs have been reported to be involved and dysregulated during the progression of cancer. In addition, recent findings also highlight that the function of some of the key regulatory proteins in lncRNA-mediated regulation may be impaired through cancer associated mutations. *DNMT3a*²²⁰⁻²²² as well *EZH2*²²³⁻²²⁶ are indeed reported mutated in subsets of cancer.

Mutations of *DNMT3a* are reported in 20% of patients with AML^{220,222} and biochemical studies show that the majority of these mutations result in reduced catalytic activity of *DNMT3a*^{220,221}. 182 genomic regions with reduced methylation were initially described in those patients²²⁰ and subsequent studies found nearly 889 differentially expressed genes in patients with mutated *DNMT3a*²²⁰. The *HOX* family of genes has been associated with AML pathogenesis²²⁷ and was indeed found induced in those patients^{220,228}. Noteworthy, the second *de novo* DNA methyl transferase, *DNMT3b*, and the maintenance *DNMT1* was not found mutated in any of those patients²²².

Increased expression of *EZH2* has been reported in several cancers including prostate²²³- and breast cancer²²⁴ and elevated levels are associated with poor clinical outcome. Mutations within the catalytic domain of *EZH2* have also been reported²²⁵. These mutations result in gain of function and are similarly associated with poor prognosis²²⁶. Surprisingly, mutations generating premature stop codons and truncated proteins have been reported and are also related to poor prognosis²²⁹. The fact that both activating as well as inactivating mutations are associated with poor clinical outcomes highlights the complex functional role of *EZH2*.

Taken together, further studies are needed to better understand the functional importance of mutated *DNMT3a* and *EZH2*. However, it is intriguing to speculate whether their interaction with lncRNAs may be involved in such cancer-related events.

2.4 miRNAs

The first miRNA was reported in *C. elegans* in 1993 when Ambros and colleagues found the short RNA lin-4²³⁰. Although not defined as a miRNA at the time, lin-4 shared sequence complementarity and caused suppression of the protein-coding lin14 mRNA. Lin-4 was for long considered as a unique case and it was not until the year 2000 that the presence of a second miRNA was reported, lethal-7 (let-7)²³¹. Let-7 was identified as important for developmental timing in *C. elegans* and was found highly conserved from nematode to human²³². Today, ~20 years after the first miRNA was described, over 1,000 miRNAs have been identified and 30-60% of all human mRNAs are suggested to be under the regulatory control of miRNAs^{233,234}.

2.4.1 Biogenesis of miRNAs

miRNAs are derived from primary transcripts that are transcribed by pol II, 5' capped and contain polyA tails and are thus very similar to most protein-coding mRNAs²³⁵. miRNAs may localize within exons and introns of proteins-coding mRNAs as well as lncRNAs²³⁶. Consequently, this allows for some miRNAs to be co-expressed together with their host gene. Initially, the miRNA precursor is transcribed as a pri-miRNA containing a hairpin loop structure. This structure is recognized by the RNase III endonuclease Drosha²³⁷ and its cofactor DGCR8²³⁸. The pri-miRNA is cleaved into a pre-miRNA containing a 2 nt 3'-overhang leaving the hairpin loop structure intact. The nuclear localized protein Exportin-5 recognizes the pre-miRNA and shuttles it to the cytoplasm²³⁹. Next, the RNase III endonuclease Dicer cleaves the pre-miRNA close to the loop structure, whereby an RNA duplex of 20-22 nt is formed²⁴⁰. Dicer provides loading of the duplex into the RISC (RNA induced silencing complex) containing the Ago2 protein^{241,242}. The mature miRNA gets loaded into Ago2 while the 'guide' strand gets degraded. The mature miRNA guides the RISC complex to the targeted mRNAs. The miRNA target recognition takes place predominantly in the 3'UTR of mRNAs, although targeting of promoter regions^{151,243}, ORFs^{69,244} as well as 5'UTRs²⁴⁵ has been reported.

Target recognition is guided by the so-called seed sequence. The seed sequence corresponds to nucleotides 2-7 of the miRNA 5'end and forms RNA:RNA interactions with the targeted mRNA^{233,246}. The complementarity between the miRNA and mRNA dictates the functional outcome: imperfect pairing generally leads to destabilization²⁴⁷ and translational suppression²⁴⁸ while perfect pairing, although rare, allows for Ago2 mediated cleavage of the mRNA^{249,250}.

In addition to the above described pathway of miRNA biogenesis, a number of miRNAs are also processed independently of Drosha. Mirtrons is a class of miRNAs situated within introns that are spliced directly into RNA hairpin structures. Drosha cleavage is therefore bypassed but Dicer processing is still required upon translocation to the cytoplasm^{251,252}. Moreover, the presence of capped pre-miRNAs which coincide with TSSs has also been

described. Similarly to mirtons, 5' capped pre-miRNAs are independent of Drosha cleavage but efficiently loaded into Ago2 after being processed by Dicer²⁵³.

Although predominantly thought to be active in the cytoplasm, several reports suggest a role for the miRNA pathway also in the nucleus. A role for nuclear Ago2 and Dicer is emerging, possibly serving as a link and interplay between transcriptional and post-transcriptional regulation^{35,151,254-257}. Small RNAs directed against promoters are known to induce TGS in an Argonaute-dependent manner¹⁵¹, DNA double-strand breaks activates the DNA-damage response (DDR) in a process that is dependent on Dicer processed RNA²⁵⁸. Finally, it has also been found that Dicer interacts with pol II and degrades endogenous dsRNA structures thereby preventing activation of the interferon-response pathway³⁵. Although intriguing, more studies will be needed in order to fully elucidate the nuclear function of the miRNA pathway.

Noteworthy, the discovery of the miRNA pathway allows for the design of siRNAs and shRNA. siRNA/shRNA induced knockdown of specific mRNAs and ncRNAs is a technique which has revolutionized molecular biology and is widely used for functional investigations²⁵⁹.

2.4.2 miRNAs and cancer

Like many protein-coding genes, miRNAs are often categorized as tumor suppressor miRNAs and tumor promoting miRNAs (onco-miRs). Since the vast majority of all mRNAs are under the regulatory control of miRNAs, it is not surprising that many miRNAs have indeed been related to disease and cancer development.

The first tumor suppressor miRNAs to be identified were the miR15a/miR16-1 cluster on chromosomal region 13q14, a region frequently deleted in chronic lymphocytic leukemia^{28,260}. The miR15a/miR16-1 cluster controls cell cycle regulation and apoptosis by targeting of *CYCLIN D1/E1*²⁷ and *BCL2*²⁶¹. The first onco-miRs to be identified were the miR-17/96 cluster, consisting of: miR-17, miR-18a, miR-19a, miR20a, miR19b-1 and miR92a-1^{262,263}. In a region frequently amplified in B-cell lymphoma, the miRNA host gene *C13orf25/MIR717HG* was found. Shortly after its discovery, ectopic expression revealed its distinct function as an onco-miR²⁶⁴.

The miR15a/miR16-1 and miR-17/96 clusters were among the first miRNAs to be identified as involved in cancer development. Since then, hundreds of miRNAs have been characterized, reported mutated and dysregulated in cancer and numerous miRNA:mRNA regulatory interplays have been experimentally validated.

2.5 PTEN

The phosphatase and tensin homolog (*PTEN*) was identified as a tumor suppressor gene frequently deleted and mutated on the chromosomal region 10q23^{265,266}. About 15 years after its discovery, *PTEN* is now identified as one of the most commonly inactivated tumor suppressor genes, frequently deleted²⁶⁷, mutated^{267,268} as well as transcriptionally inhibited by epigenetic suppression^{269,270}. Complete loss of *PTEN* is generally found in advanced tumors and metastasis²⁶⁷. However, subtle variations in *PTEN* expression have been reported to be sufficient in order to affect cancer susceptibility²⁷¹. Interestingly, complete loss of *PTEN* in early tumor development induces p53 dependent senescence and surprisingly does not provide a proliferative advantage unless concurrent with the loss of p53²⁷². Complete loss of *PTEN* is therefore initially less tumorigenic than partial inactivation.

The best-characterized function of PTEN is its involvement as a negative regulator of the prosurvival and oncogenic PI3K/AKT pathway (Figure 6). Upon growth stimulation of cells (Figure 6A), the class I phosphatidylinositol 3-kinases (PI3K) catalyzes the phosphorylation of phosphatidylinositol-4,5-bisphosphate (PIP₂) into PIP₃ (Figure 6 B-C) Next, PIP₃ recruits AKT and PDK1 (phosphoinositide-dependent kinase-1) to the inner membrane (Figure 6D) whereby phosphorylation of AKT (pAKT) is promoted (Figure 6E) and a downstream signaling cascade affecting cell survival and proliferation is initiated (Figure 6F) (reviewed in²⁷³). pAKT targets numerous downstream proteins such as MDM2^{274,275}, p21^{Cip1/WAF1}²⁷⁶ and p27^{Kip1}²⁷⁷⁻²⁷⁹. pAKT-induced phosphorylation of MDM2 translocates to the nucleus where it binds to and triggers degradation of p53. In contrast, pAkt-induced phosphorylation of the negative cell cycle regulators p21^{Cip1/WAF1} and p27^{Kip1} impairs their nuclear translocation and attenuates their cell-cycle inhibitory effect.

PTEN antagonizes the activity of PI3K by dephosphorylating PIP₃ into its inactive form PIP₂ (Figure 6G)²⁸⁰. Upon loss of PTEN expression, e.g. during cancer development, PIP₃ thus accumulates whereby the PI3K/AKT pathway is hyper-activated and subsequently promotes cell growth and survival.

PTEN is highly regulated on both the transcriptional, post-transcriptional and post-translational level by numerous overlapping mechanisms, indicating the importance of keeping the expression at stable levels. The tumor suppressor gene p53 has been reported to directly bind the *PTEN* promoter, causing transcriptional activation²⁸¹. In contrast, *PTEN* translation is suppressed by several miRNAs. miR-21 is one such miRNA and it is frequently induced in a number of different tumors, resulting in destabilization of *PTEN* mRNA and translational suppression²⁸². *PTEN* was also the first example being reported as regulated by miRNA sponging. *PTEN* and its pseudogene, PTENpg1, are involved in ceRNA networks where miRNA crosstalk between different mRNAs and lncRNAs serves as a fine-tuning mechanisms of PTEN expression (see section 3.2.3 for more details)⁵⁷.

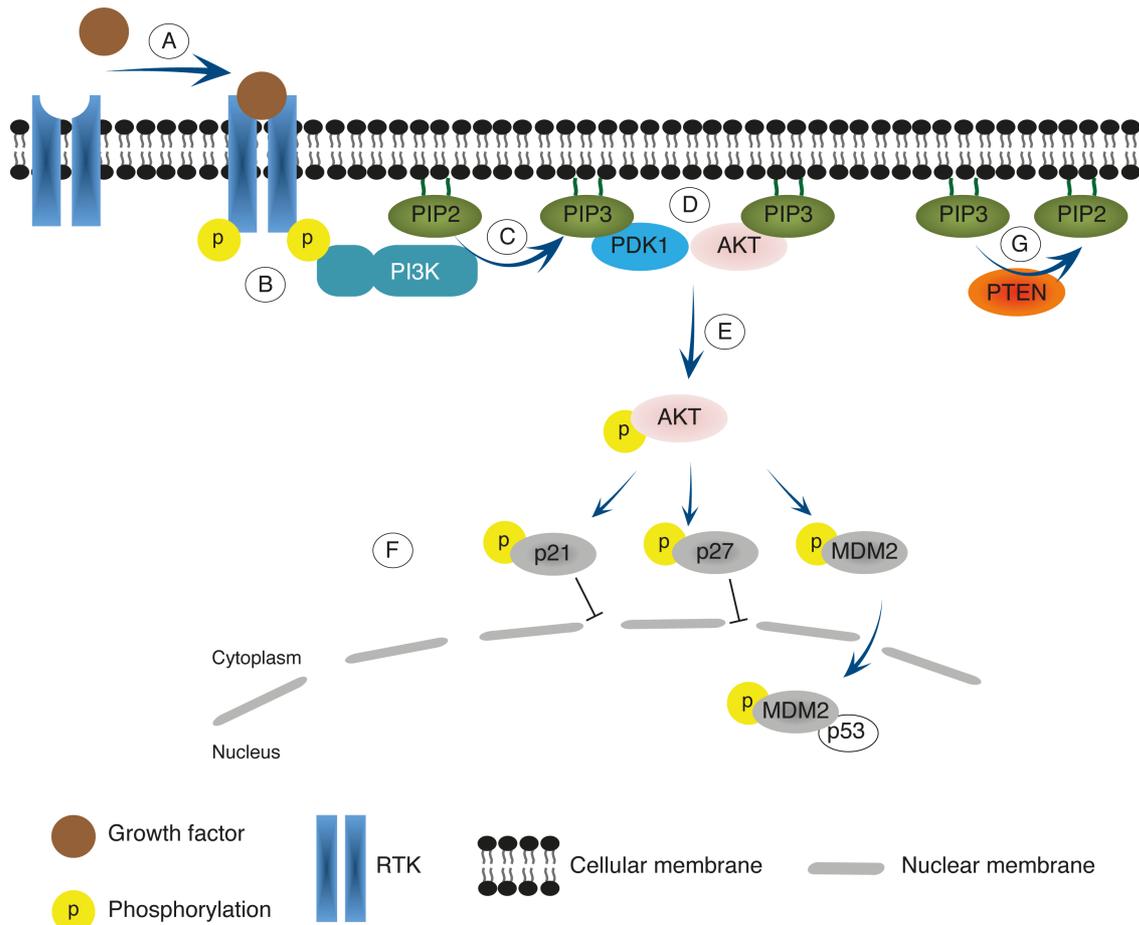


Figure 6 Schematic illustration of the *PI3K/AKT* pathway (A) The *PI3K/AKT* pathway is triggered upon growth stimulation of cells. (B) *PI3K* becomes activated by *RTK* and subsequently (C) catalyzes phosphorylation of *PIP*₂ into *PIP*₃. (D) *PIP*₃ recruits *AKT* and *PDK1* to the membrane whereby (E) phosphorylation of *AKT* is triggered. (F) *pAKT* initiates a downstream signaling cascade, affecting key regulatory proteins such as *p21*, *p27* and *MDM2*. (G) *PTEN* antagonizes the activity of *PI3K* and dephosphorylates *PIP*₃ into *PIP*₂.

2.6 BRAF

In 2002, the understanding of BRAF and its involvement in cancer development was dramatically increased. Sequencing efforts discovered that *BRAF* was frequently mutated in melanoma and many other cancers²⁸³. Overall, 8% of all cancers harbor mutated *BRAF*, but it occurs in as many as 50% of metastatic melanoma cases²⁸⁴, where the large majority of mutations are represented by a valine (V) to glutamic acid (E) substitution at position 600 (*BRAF*^{V600E}). The oncogenic BRAF is active through mimicking its active, phosphorylated form resulting in increased proliferation and survival. However, mutated *BRAF* is also found in nevi, which are quiescent and thus benign²⁸⁵. Mutation of *BRAF* by itself is therefore insufficient to drive melanoma tumorigenesis^{285,286}. Other genetic events, such as inactivation of *PTEN*, have been suggested as necessary in order to initiate melanoma carcinogenesis²⁸⁷.

The discovery of *BRAF*^{V600E} led to the initiation of numerous drug discovery efforts to target this hyperactive kinase (reviewed in²⁸⁸). Using targeted approaches to screen for small molecules resulted in the discovery of vemurafenib, which selectively targets *BRAF*^{V600E} and inhibits its oncogenic properties. Remarkably, ~80% of metastatic melanoma patients with *BRAF*^{V600E} initially respond well to vemurafenib treatment^{289,290}. However, development of resistance is a major drawback and the onset of drug resistance normally emerges within a year²⁹¹. Several mechanisms for acquired vemurafenib resistance have been proposed such as the involvement and concurrent loss of *PTEN*^{287,292}.

The involvement of BRAF in the RAS-RAF-MEK-ERK pathway (frequently also referred to as the mitogen-activated protein kinase, MAPK, pathway) makes it an important cellular regulator of growth factor signaling (Figure 7). Receptor tyrosine kinases (RTKs), such as the epidermal growth factor receptor, get activated upon growth factor stimulation (Figure 7A) and a cascade of phosphorylation events is initiated. The membrane bound RAS serves as the molecular switch and convert from its inactive form RAS-GDP to its active form RAS-GTP (Figure 7B)²⁹³. When activated by RTKs, RAS-GTP promotes phosphorylation and kinase activation of BRAF (Figure 7C), which forms homodimers (Figure 7D)^{294,295}. The BRAF homodimers directly phosphorylate MEK (Figure 7E), which next phosphorylates ERK (extracellular signal-regulated kinase) (Figure 7F). When phosphorylated, ERK translocates to the nucleus and promotes proliferation through changes in transcription (Figure 7G)²⁹⁶. The presence of *BRAF*^{V600E} mimics the phosphorylated form of BRAF and ultimately results in a constitutively active BRAF that function independently of signaling through RTKs (Figure 7H).

Ectopic expression of *BRAF*^{V600E} has been reported to regulate hundreds of lncRNAs, most of which remain to be functionally investigated²⁹⁷. However, one of them, lncRNA-BANCR (BRAF regulated lncRNA) is suggested to be involved in cell migration²⁹⁷.

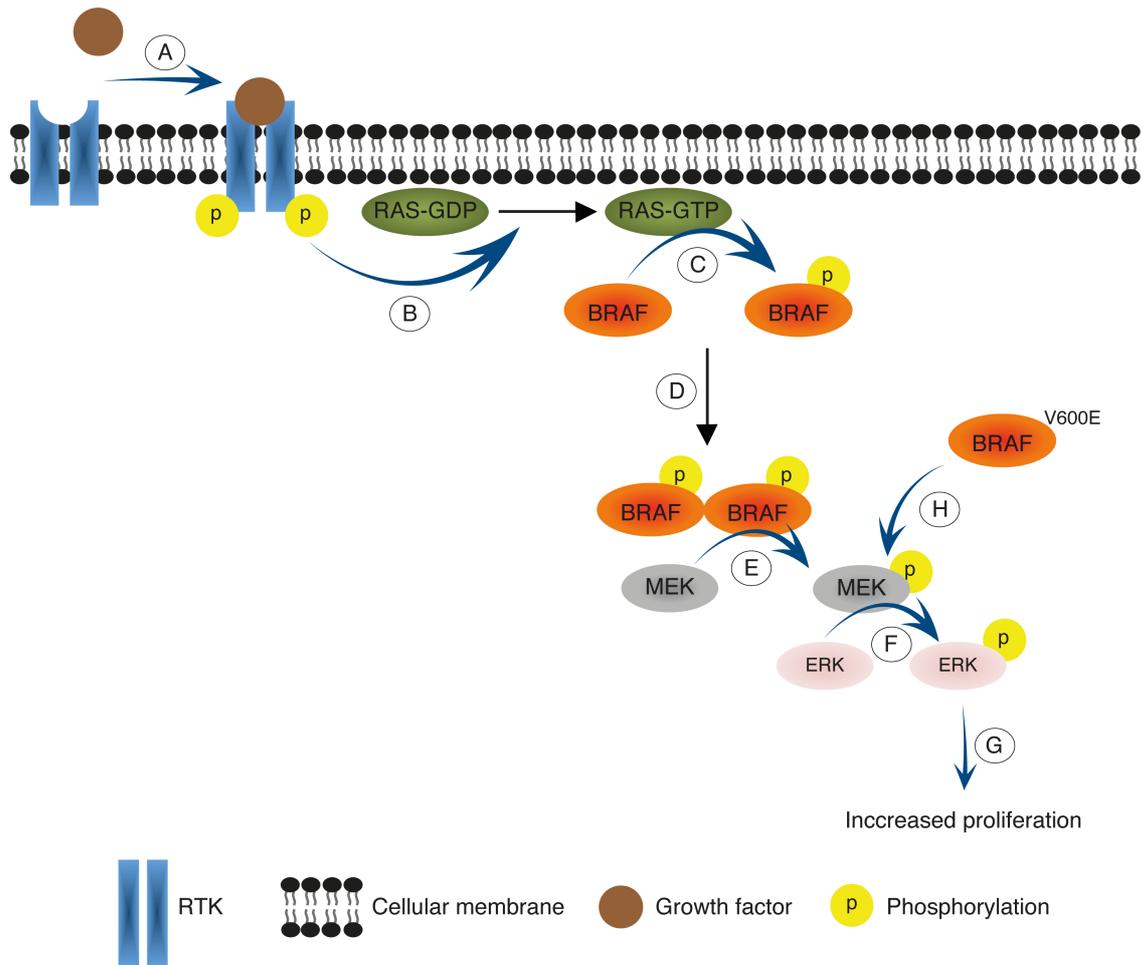


Figure 7 Schematic illustration of the *RAS-RAF-MEK-ERK* pathway (A) The *RAS-RAF-MEK-ERK* pathway is activated upon growth stimulation of cells and a signaling cascade is initiated. (B) Membrane bound RAS-GDP becomes activated and converted into RAS-GTP. (C) RAS-GTP promotes phosphorylation of BRAF, which (D) consequently forms homodimers. (E) The pBRAF homodimers phosphorylates MEK, which (F) subsequently phosphorylates ERK. (G) pERK promotes cell proliferation through changes in transcription. (H) Mutated BRAF^{V600E} is active also in the absence of signaling through RTKs.

2.7 STAT3

The STAT family of proteins consists of seven members: STAT 1, 2, 3, 4, 5a, 5b and 6. This family of proteins is important cellular regulators that transfer extracellular signals from the membrane to the nucleus²⁹⁸⁻³⁰⁰. STAT proteins are specifically phosphorylated and activated by the protein family of Janus tyrosine kinases (JAK) and this pathway is therefore frequently referred to as the JAK-STAT pathway. STAT3 will here be discussed in more detail.

In the JAK-STAT pathway, binding of a cytokine such as interleukin 6 (IL-6), to its receptor activates autophosphorylation of the gp130 transmembrane receptor (reviewed in³⁰¹). This leads to the activation of several JAKs and subsequently phosphorylation (on tyrosine residue 705) of STAT3. Phosphorylation of STAT3 allows the formation of homo- and heterodimers, which is essential for its translocation to the nucleus where it executes its function as a transcription factor. STAT3 binds to DNA sequences called GAS elements (DNA gamma interferon activation sequence) within promoters, thereby regulating transcription^{302,303}. STAT3 is involved in transcriptional regulation of genes involved in proliferation and survival, for example cyclins, Survivin and the Bcl-2 family of proteins³⁰⁴. Constitutive activation of STAT3 is reported in several tumors and often correlates with poor prognosis^{305,306}.

In addition, STAT3 has been reported to be involved in transcriptional repression by interacting with the chromatin suppressors DNMT1 and HDAC1³⁰⁷. STAT3 recruits DNMT1 and HDAC1 to the promoter of the tyrosine phosphatase *SHP-1*, which is a negative regulator of JAK-STAT signaling³⁰⁷⁻³⁰⁹. In a similar mechanism, STAT3 also mediates epigenetic silencing of the tumor suppressor genes *p53* and *p16^{INK4A}* in melanoma³¹⁰.

Control of STAT3 expression and activation is mediated by several different mechanisms including several regulatory functions of ncRNAs. As previously discussed (see section 3.1.6.2), activation of STAT3 is promoted by the action of the lncRNA lnc-DC through its inhibitory effect on SHP-1⁶¹. In complex with lnc-DC, SHP-1 loses its ability to dephosphorylate and inactivate STAT3 signaling. Moreover, STAT3 promotes the expression of miR-21 in colon adenocarcinomas. miR-21 can target *PTEN* thereby promoting cancer and cellular transformation^{311,312}. STAT3 has also been found to induce the expression of the miRNA-17/92 cluster³¹³, a cluster frequently activated as well as amplified in numerous cancers^{262,314}. In contrast, *STAT3* is under miRNA-mediated suppression by miRNA-17 and miRNA-20a, both encoded by the miRNA-17/92 cluster, thus allowing for a negative feedback mechanism³¹⁵. Finally, the highly evolutionary conserved miRNA, let-7a, has also been reported to regulate *STAT3* expression in the liver carcinoma cell line HepG2³¹⁶.

2.8 LYMPHOMA

The lymphocytes include the Natural killer cells, T-cells and B-cells and all arise from the bone marrow. Lymphoma is a common name of neoplasms of lymphoid precursor cells and the biology and morphology of lymphomas is highly heterogenous. Lymphomas are divided into Hodgkin's Lymphomas and Non-Hodgkin's Lymphoma (NHL). NHL is also divided in High grade and Low grade malignant lymphomas. High grade NHL is often curable with combined chemo-immunotherapy. They are also characterized by fast proliferation and rapid tumor growth. In contrast, Low grade NHL is represented by a slow disease progression and slow proliferation. These patients are often not curable but can live long with the disease and are usually treated when symptom occurs.

Lymphomagenesis is a process that has been linked to genetic, environmental as well as infectious factors. HIV, Epstein-Barr virus (EBV), human herpes virus 8 (HHV-8), human T-lymphocytic virus type 1 (HTLV-1) and hepatitis B/C are all examples of viruses that have been linked to the disease³¹⁷⁻³²⁰.

Diffuse large B-cell lymphoma (DLBCL) is a subgroup of the High grade lymphomas, and represents approximately 40% of all diagnosed lymphoma world-wide³²¹. Based on gene expression, the DLBCL can further be categorized into three different subtypes; (1) germinal center B-like, (2) activated B-cell (ABC) and (3) primary mediastinal B-cell lymphoma³²². The ABC DLBCL subgroup has poor prognosis^{323,324}. The frequency of DLBCL differs between different regions; In Egypt, 50% of diagnosed NHL represents DLBCL, while only 25-35% in the western countries³²⁵.

Expression profiling studies of comparing lymphoma samples from Swedish and Egyptian ABD DLBCL patients have reported differential expression of *STAT3* and *STAT5b*, possibly suggesting that different microbial infections or other environmental factors may be involved in the tumoregenesis³²⁶. Moreover, *STAT3* has been suggested to be an oncogene with therapeutic potential for ABC DLBCL^{327,328}.

3. AIMS OF THIS THESIS

The overall aim of this thesis was to functionally characterize the role of ncRNAs in gene regulation and their involvement in cancer development. Specifically, I aimed to increase our understanding of ncRNA-mediated regulation of the cancer-associated genes *PTEN* and *STAT3*. Increased understanding of these processes may reveal how to modify these pathways and allow for development of novel drugs.

3.1 PAPER I

- The aim of paper I was to investigate if the *PTEN* pseudogene, *PTENpg1*, expressed asRNA transcripts. The asRNA transcripts were functionally investigated and their role in regulating the expression of the corresponding tumor suppressor gene *PTEN* was examined.

3.2 PAPER II

- The aim of paper II was to follow up the findings in paper I. We aimed to investigate the role of *PTENpg1* asRNA in melanoma with a focus on the involvement of *PTENpg1* asRNA in the mechanisms of drug resistance to the BRAF inhibitor vemurafenib. Furthermore, we investigated the expression of *PTENpg1* asRNA in primary melanoma patient samples.

3.3 PAPER III

- The aim of paper III was to investigate the expression of miRNAs in patients with DLBCL and to explore whether different environmental exposures could affect miRNA expression.

4. RESULTS AND DISCUSSION

4.1 PAPER I

A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells

4.1.1 Rationale

PTEN is a critical tumor suppressor gene, which is frequently inactivated during cancer development. In addition to mutations and deletions, *PTEN* is also suppressed by epigenetic inactivation by largely unknown mechanism(s). As previously mentioned, even subtle suppression of *PTEN* levels may promote tumorigenesis²⁷¹. The involvement of lncRNA-mediated regulation of *PTEN* at the post-transcriptional level had recently been reported, describing the action of the PTENpg1 sense transcript functioning as a ceRNA⁵⁷. Studies of other pseudogenes, namely OCT4-pg5, had shown that this group of non-coding genes may additionally encode asRNAs that could have the capability to regulate the transcriptional activity of its parental protein-coding gene. Based on these observations, we set out to investigate whether the PTENpg1 gene may also encode for asRNA transcripts involved in the regulation of *PTEN*. Such findings could improve our understanding of the fine tuned regulation of *PTEN* and its dysregulation during tumor progression and further shed light on the regulatory function of the PTENpg1 gene.

4.1.2 Characterization of PTENpg1 asRNA transcripts

In this first paper, we set out to investigate whether the previously described PTEN pseudogene^{57,167,171}, PTENpg1 also encodes an asRNA. Our initial observations using the UCSC genome browser indicated one aligned EST (BX374997), prompting us to investigate the transcriptional landscape of this locus in greater detail. Supporting the observations on the UCSC genome browser, asRNA transcripts were frequently detected and found differentially expressed among several human cell lines. By expanding our analysis, the presence of two distinct transcripts was revealed: PTENpg1 asRNA α and PTENpg1 asRNA β . These two transcripts overlap each other but differ in their TSS and cellular localization. The TSS of PTENpg1 asRNA α is initiated further upstream compared to the β isoform, consequently leading to a greater shared overlap of the α transcript with both the *PTEN* promoter and PTENpg1 sense. We also noted that the PTENpg1 associated transcripts, PTENpg1 sense, PTENpg1 asRNA α and PTENpg1 asRNA β were concordantly expressed.

4.1.3 PTENpg1 asRNA α

PTENpg1 asRNA α is the isoform sharing the greatest sequence complementarity with the *PTEN* promoter. Correlation analysis demonstrated that high expression of PTENpg1 asRNA α associated with low expression of *PTEN*. This was suggestive of PTENpg1 asRNA α being a negative regulator of *PTEN* expression, possibly acting through a similar mechanism as

described for the OCT4-pg5 asRNA⁷¹. shRNAs and siRNAs were designed to target the PTENpg1 asRNA α isoform. Suppression of the PTENpg1 asRNA α also resulted in increased expression of *PTEN*. The induction was restricted to *PTEN* and no change in expression was observed for PTENpg1 sense. Mechanistically, we revealed that PTENpg1 asRNA α localized to the *PTEN* promoter recruits DNMT3a and EZH2 consequently resulting in transcriptional suppression of *PTEN*. Accordingly, we observed loss of DNMT3a, EZH2 and H3K27me3 at the *PTEN* promoter upon depletion of PTENpg1 asRNA α . Recruitment of DNMT3a and EZH2 appeared to be dependent on the region of PTENpg1 asRNA α containing sequence homology to the *PTEN* promoter. Although the exact interactions at the *PTEN* promoter should be investigated in greater detail we envision at least three possible scenarios; **(1)** The formation of DNA:RNA interactions, **(2)** the formation of a DNA:RNA triplex or **(3)** the formation of RNA:RNA interactions. In the case of RNA:RNA interactions, the presence of *PTEN* associated promoter overlapping RNAs is a prerequisite. Finally, it would be of great interest to further characterize the PTENpg1 asRNA α transcript and in greater detail investigate what region that interacts with DNMT3a and what region that provides specificity for the recruitment to the *PTEN* promoter.

4.1.4 PTENpg1 asRNA β

We found the PTENpg1 asRNA β isoform to stabilize expression levels of PTENpg1 sense through the formation of RNA:RNA interactions. Elevated levels of PTENpg1 sense have previously been shown to increase miRNA sponging of *PTEN*-related miRNAs and thus release *PTEN* from miRNA-mediated suppression⁵⁷. Our observations of the function of PTENpg1 asRNA β isoform are in support of those findings. Moreover, we showed that PTENpg1 sense is a polyA negative transcript. Mechanistically, the stability of PTENpg1 sense seems to increase through the formation of RNA:RNA interactions with PTENpg1 asRNA β , which is polyA positive.

4.1.5 Further characterization of PTENpg1 asRNA transcripts

In order to fully evaluate the number of different transcripts at the PTENpg1 locus, a northern blot analysis would be of great interest. Although a rigorous effort was spent on such analysis, we did not succeed in this due to technical limitations as mentioned below, and possibly to the relatively low expression of these transcripts. Due to the site of integration of the pseudogene, the PTENpg1 asRNA is a composite of two seemingly different genomic elements: one region with high sequence complimentary with the *PTEN* gene and one region with highly repetitive elements, predominantly LINE elements. No successful northern blot probe design could be generated within the LINE repetitive region due to the high abundance of LINE elements throughout the human genome¹⁰⁸. In addition, the region homologous to *PTEN* is relatively rich in cytosine and guanine and also unsuitable for probe design. Additional challenges included generating strand specific probes for the asRNA that would not result in hybridization with sense transcripts. Taken together, although northern blot analysis of the PTENpg1 locus would have been very informative, we could not carry out such analysis successfully.

4.1.6 PTENpg1 asRNA α versus PTENpg1 asRNA β

In contrast to the PTENpg1 asRNA β isoform, we did not observe any RNA:RNA interactions between PTENpg1 sense and PTENpg1 asRNA α , although these two transcripts share an even longer sequence overlap. We believe this may be the consequence of secondary structures present within the 5' end of PTENpg1 asRNA α . Such secondary structures could hide the homology-containing nucleotides from forming RNA:RNA interactions with PTENpg1 sense. Similar interactions as described here have also been reported for the tumor suppressor gene *p53* and its antisense transcript, *WRAP53*. *WRAP53* stabilizes the *p53* mRNA by forming an RNA:RNA interaction at the 5' end of the transcripts⁶⁸. Similar to PTENpg1, *p53* has several annotated asRNAs (*WRAP53* α , β and γ), yet only one isoform is involved in the regulation of *p53*⁶⁸.

Our observation that PTENpg1 asRNA have several different isoforms is also in line with many other studies on lncRNAs such as *Tsix*^{77,329} and *Airn*⁹³. Both *Tsix* and *Airn* have been reported to have differentially spliced transcripts, where the unspliced and nuclear isoform is found to be the dominant functional one. No function has so far been assigned for either spliced *Tsix* or spliced *Airn*. Notably, the distinct functions we report for the PTENpg1 asRNA α and β isoforms are one of the first examples where lncRNA isoforms from the same gene have been assigned different and distinct functions. lncRNAs are frequently differentially spliced in a similar manner as mRNAs and more examples of lncRNA isoforms with distinct functions are likely to be revealed in the near future.

4.1.7 Evolutionary conservation of asRNA mediated regulation of *PTEN*

Similar to other processed pseudogenes, PTENpg1 is lineage specific, possibly due to the previously mentioned burst in retrotranspositional activity in ancestral primates^{161,162}. It is therefore unlikely that any corresponding PTEN pseudogene will be found in more distant species. Indeed, bioinformatics approaches in our lab have not been able to identify a *Pten* pseudogene in mouse, therefore supporting PTENpg1 to be lineage specific (unpublished data). It is tempting to speculate whether, for example, mouse has developed other means of regulating *Pten*. Investigating the involvement of lncRNAs in transcriptional and post-transcriptional regulation of *Pten* in mouse would be of great interest. Such studies could reveal exciting information about the evolution of lncRNAs and would shed light on whether convergent evolution of lncRNAs could have occurred. Other species may have evolved and developed other systems for the regulation of *Pten*, possibly involving lncRNAs not related to PTENpg1. In line with this, it could be of interest to generate a transgenic mouse and/or mouse cell lines containing the PTENpg1 locus. This could reveal whether the PTENpg1 pathway is active in non-primate organisms, which lacks such a regulatory pathway and would highlight if mouse and human in essence have overlapping regulatory pathways that is dictated and recruited by different repertoires of lncRNAs.

It is also intriguing to speculate whether *PTEN* may have *cis*-expressed asRNAs. Within this study, we did not specifically look for such transcripts but they may exist. It would be

remarkable to find both *cis*- and *trans*-acting asRNAs regulating the same protein-coding gene. From the evolutionary point of view, this could either indicate that the asRNA was present already before the pseudogenization of *PTEN*. Alternatively, *PTEN* asRNA as well as *PTENpg1* asRNA could have evolved independently after the pseudogenization and developed two overlapping modes of regulation: one *cis*-acting and one *trans*-acting. This question could be addressed by looking for *PTEN* related asRNA transcription in other species. Primates, where *PTENpg1* is present could be contrasted with mouse, where *PTENpg1* is absent. By doing so, a time line for the evolution of asRNA-mediated regulation of *PTEN* would be determined.

Collectively, evolutionary studies of *PTENpg1*, and in particular its asRNA transcripts, are of great interest for future studies and may prove informative with regards to the evolutionary path of lncRNAs.

4.1.8 *Cis*- versus *trans*-acting lncRNAs

The *PTENpg1* encoded lncRNAs maintain transcriptional as well as post-transcriptional regulation of *PTEN*. This takes place through: **(1)** Interactions between miRNAs:lncRNAs the latter acting as a miRNA decoy, **(2)** interactions between lncRNAs:lncRNAs where *PTENpg1* sense and asRNA β interact and **(3)** interactions between lncRNA:proteins, where *PTENpg1* asRNA α interacts with DNMT3a. All of these functional properties have been characterized also for other *cis*-acting asRNAs: *BACE1-AS*⁶⁹ and *ciRS-7*¹⁷⁹ affect miRNA targeting, *p53* asRNA *WRAP53* regulates RNA stability⁶⁸ and *p15*^{INK4B}/*p16*^{INK4A} asRNA *ANRIL* mediates chromatin-remodeling^{21,22,97}. What is then the evolutionary benefit of having all the functions of *PTENpg1* at a separate locus?

One notion is that the generation of the *PTENpg1* pseudogene allows for greater plasticity and an entire independent regulatory system based on lncRNAs. This network has evolved in the absence of constraints from any protein-coding sense counterpart and thus allows for greater flexibility. The fact that this system is built on RNAs and not proteins allows for a relatively energy efficient mechanism to control gene expression that has inherent plasticity and flexibility, as RNA can fold into various conformational configurations and interact with DNA, RNA and/or proteins. Furthermore, the *PTENpg1* locus offers all of the above-discussed regulatory units and it is hard to envision the rapid and efficient evolution of such a sophisticated network under the restriction of a protein-coding sense gene. Additionally, at the pseudogene, the large majority of the *PTEN* promoter and regulatory elements were lost during the pseudogenization process, suggesting that the expression of *PTEN* and *PTENpg1* can be regulated by different TFs. This would consequently allow for independent feedback mechanisms offering greater variability compared to *cis*-acting asRNA networks. However, no studies have yet addressed the transcriptional regulation of *PTENpg1* and future investigations will have to explore this notion in greater detail.

Finally, although we show that PTENpg1 asRNA functions *in trans*, it would be interesting to investigate chromatin 3D structures of the *PTEN* and PTENpg1 loci and explore whether these loci are in close proximity to one another. Chromosomal looping has indeed been reported to be involved in some forms of lncRNA-mediated regulation^{60,330}.

4.1.9 PTENpg1 asRNA, DNMT3a and other ncRNAs

One of the main regulatory proteins we found to be involved in the PTENpg1-*PTEN* pathway was DNMT3a. We revealed that the PTENpg1 asRNA α acts in complex with DNMT3a to recruit DNMT3a to the *PTEN* promoter. Although a number of studies support a role for ncRNAs in guiding DNMT3a to a targeted locus^{149,152}, no comprehensive analysis of DNMT3a-ncRNA interactions has to date been carried out and no observations had to our knowledge been reported prior to ours that DNMT3a interacts with lncRNAs.

The methylome in cancer cells undergoes a switch whereby many promoter-associated CpG islands become methylated and a role for DNMT3a during cancer progression has been suggested. DNMT3a is the major *de novo* methyltransferase in human cells. Mutations of *DNMT3a* have been reported in cancer and its involvement in drug resistance mechanisms has also been demonstrated^{220-222,331}. While there is no doubt that DNMT3a is a major regulatory protein, mechanisms for its recruitment to targeted loci had until recently remained largely unknown. Our findings that DNMT3a interacts with the lncRNA PTENpg1 asRNA α serves as a pioneering example of a regulatory interplay between lncRNAs and DNMT3a. Our observations have indeed been supported by subsequent studies showing that an asRNA to HIV recruits DNMT3a to the HIV promoter³³². These observations are strikingly similar to those reported by our lab.

DNMT1 was recently found to interact with thousands of lncRNA transcripts¹⁵⁰. Contrary to our findings, these lncRNAs were acting as decoys for DNMT1 and blocked DNMT1 from inducing DNA methylation at targeted loci. Moreover, the DNMT1-associated lncRNAs were reported to primarily interact with sense and *cis*-expressed promoter-overlapping transcripts. Similar investigations into the transcripts that interact with DNMT3a on a genome-wide level would be of great interest. RIP-seq analysis of DNMT3a associated RNAs would thus be valuable in order to further shed light on the regulation and recruitment of DNMT3a to distinct loci.

Although further investigation is needed for both DNMT1 and DNMT3a in order to fully elucidate the role of lncRNAs in their regulation, it is tempting to speculate that DNMTs may be coordinated by such interactions. While lncRNAs, more precisely asRNAs, appear to recruit DNMT3a, similar interactions with sense expressed RNA transcripts seem to decoy and block DNMT1 function. These observations are suggestive of an intriguing interplay that could coordinate *de novo* and maintenance methylation. In such a model, DNMT1 is thus tethered and put 'on hold' at the targeted locus, although not catalytically active. Once DNMT3a has initiated *de novo* methylation, DNMT1 could become catalytically active and

trigger maintenance of DNA methylation upon loss of the DNMT1:lncRNA interaction. Alternatively, the DNMT1:lncRNA interaction is sustained and long-term silencing cannot be initiated, consequently leading to reactivation of the gene. Such a mechanism would permit for ‘sensing’ of DNA methylation and allow for transient and temporary methylation to be placed at the loci. Similar bimodal mechanisms have been described. XCI is initiated by the lncRNA Xist. However, after its initiation, Xist is no longer needed and the silencing is maintained in the absence of Xist. The initiation, maintenance and spreading of XCI appear to be separate mechanisms. Furthermore, the different PcG complexes (PRC2 and PRC1), appear to function in a similar way. lncRNAs are able to recruit the initiation complex, PRC2, to targeted loci whereby H3K27me3 is induced. The maintenance complex, PRC1, eventually takes over and the suppressive mark is maintained long-term. Again, the initiation and maintenance of silencing are separated by two different mechanisms. Such separation of the initiation and maintenance mechanisms allows for the lncRNAs to essentially serve as the “switch” between initiation and maintenance.

A model for the involvement of lncRNAs in DNMT1- and DNMT3a-mediated regulation of gene expression is presented in Figure 8.

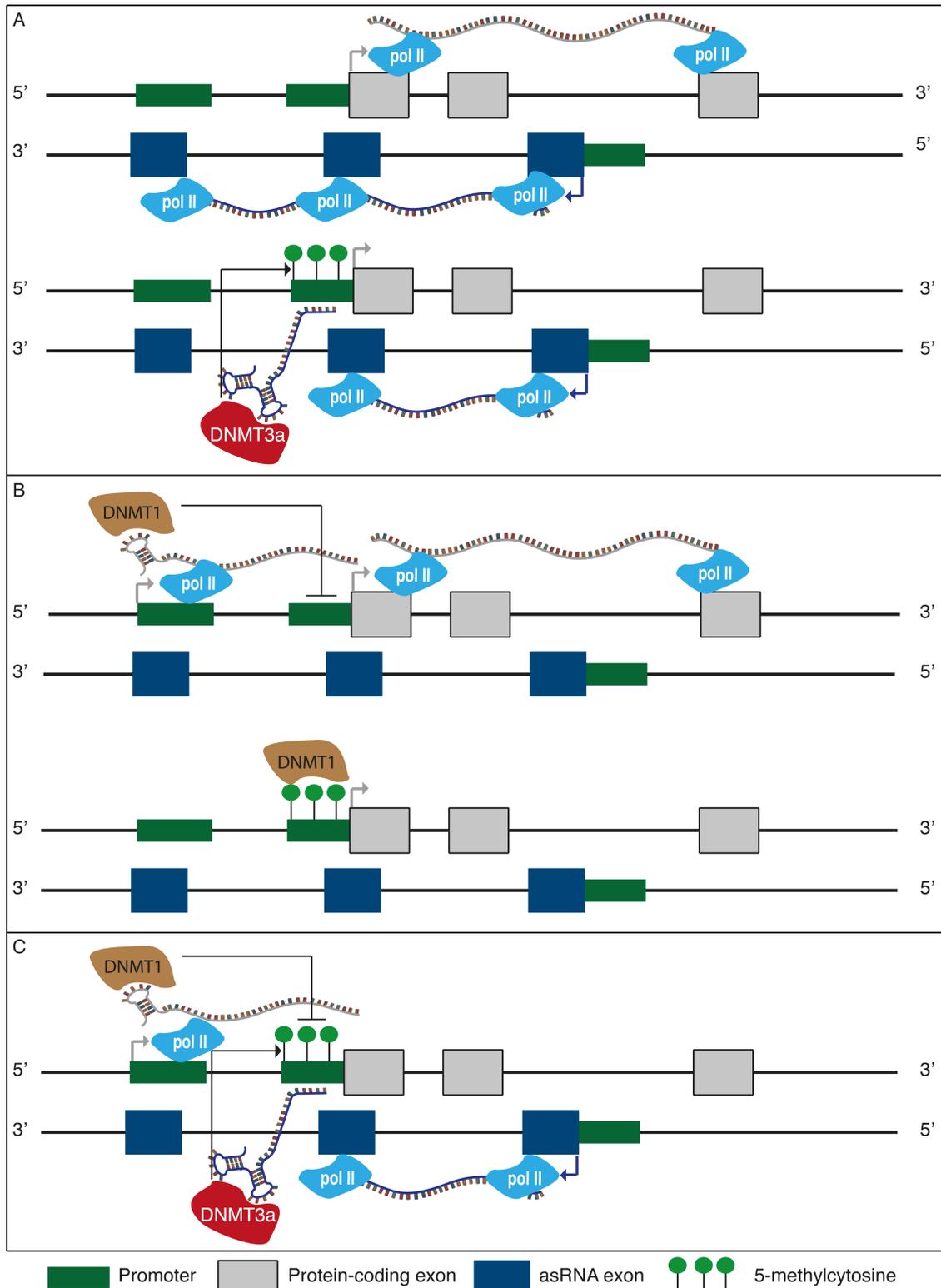


Figure 8 A schematic model for the involvement of lncRNAs in DNMT3a- and DNMT1-mediated gene regulation (**A, DNMT3a**) *Top*: A sense:asRNA pair is expressed. *Bottom*: The asRNA recruits DNMT3a to the sense gene promoter whereby gene suppression is initiated by DNA methylation. (**B, DNMT1**) *Top*: A promoter overlapping lncRNA is expressed from an upstream promoter element. DNMT1 is tethered to the locus but its activity is blocked by the interaction with the lncRNA. *Bottom*: The expression of the upstream lncRNA is suppressed whereby DNMT1 becomes active. (**C**) A suggested interplay between DNMTs and lncRNAs: DNMT3a initiates transient methylation. Long-term silencing is prevented as long as the DNMT1:lncRNA interaction is maintained. lncRNAs therefore serve as both inhibitors and activators of methylation.

4.1.10 Concluding remarks of paper I

The findings presented within this study to some extent raise more questions than have been answered. While our results clearly show the PTENpg1 asRNA α to be the dominant regulator of *PTEN* expression, it remains true that the PTENpg1 locus is indeed deleted in several different forms of cancer^{57,175}. Based on the model proposed in this paper, such deletions would lead to the induced expression of *PTEN*. This would presumably not be beneficial for the tumor cells as increased levels of *PTEN* would induce cell cycle arrest and apoptosis (reviewed in³³³). However, loss of PTENpg1 in melanoma has been found concurrent with co-deletions of *PTEN*, suggesting that the deletion of PTENpg1 also has other effects than through its action on *PTEN*¹⁷⁵. In this particular case, at least three different scenarios can be envisioned. **(1)** As supported by the ceRNA model, loss of PTENpg1 may result in loss of miRNA-sponging for other targets not related to *PTEN*. **(2)** As supported by ChIRP analysis, lncRNAs can bind hundreds of different loci to induce chromatin-remodeling⁸⁵. In such a scenario, PTENpg1 asRNA could have other targets not associated with *PTEN*. **(3)** The functional impact of PTENpg1 transcripts depends on other cellular factors, such as the expression of DNMT3a and miRNAs. For example, low expression levels of DNMT3a and/or EZH2 in certain tissues could lead to the PTENpg1 asRNA being unable to function as a transcriptional repressor of *PTEN*. Likewise, the same scenario holds true for the ceRNA model. miRNAs are expressed in a tissue-specific manner and the PTEN-PTENpg1 miRNA decoy mechanism can only take place if the *PTEN*-related miRNAs are expressed in the tissue of interest. Future studies will have to evaluate these contradictory observations in order to fully understand the regulatory functions of the PTENpg1 transcripts.

In summary, we show that the PTENpg1 expressed transcripts function in a ncRNA network consisting of lncRNAs as well as miRNAs. This network regulates both translation and transcription of *PTEN*. By characterizing this regulatory interplay, our understanding of *PTEN* inactivation will increase and hopefully reveal how manipulation of this pathway could be used for therapeutic development. Considering the tissue specificity of lncRNAs, these regulatory molecules hold great promise for targeted approaches with reduced risk of severe off target effects. The data presented in this paper may also have broader implications as models for ncRNA actions in cells.

Based on the findings in this paper, a model for PTENpg1-mediated regulation of *PTEN* is presented in Figure 9.

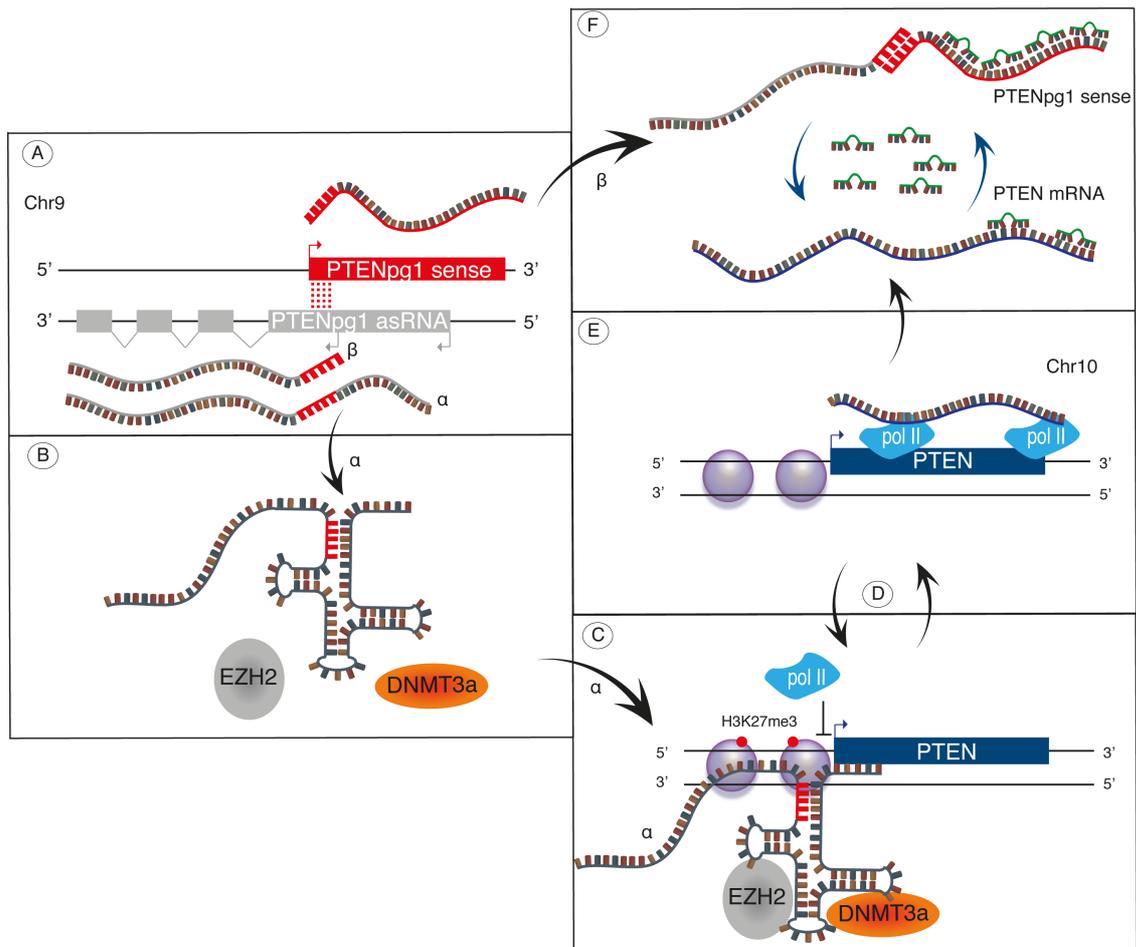


Figure 9 Model of *PTENpg1* mediated regulation of *PTEN* (A) *PTENpg1* is transcribed on chromosome 9 into three different transcripts; *PTENpg1* sense, asRNA α and asRNA β . (B) The 5' end of *PTENpg1* asRNA α folds into a secondary RNA structure with the ability to interact with DNMT3a and EZH2. The RNA sequence, which in the *PTENpg1* asRNA β isoform, interacts with *PTENpg1* sense (marked in red) is unavailable for such interactions due to the secondary folding. (C) The RNA:protein complex containing *PTENpg1* asRNA α , DNMT3a and EZH2 is recruited to the *PTEN* promoter where the formation of the chromatin suppressive mark H3K27me3 is catalyzed. (D-E) By targeting of the *PTENpg1* asRNA α transcript, the expression of *PTEN* is reactivated. (F) The *PTENpg1* asRNA β forms RNA:RNA interactions with *PTENpg1* sense. This interaction stabilizes *PTENpg1* sense, consequently affecting its decoy function of *PTEN* related miRNAs and thus the translational output of *PTEN*.

4.2 PAPER 2

PTENpg1 antisense RNA mediates PTEN suppression in vemurafenib resistance and predicts clinical outcome in melanoma patients

4.2.1 Rationale

We investigated the molecular mechanism of PTENpg1 encoded lncRNAs in great detail in the first paper using human cell lines. However, the involvement of PTENpg1 asRNA in disease remained to be elucidated. We next set out to investigate the involvement of the PTENpg1 asRNA in melanoma and we hypothesized a role for PTENpg1 asRNA in drug resistance to the BRAF inhibitor vemurafenib.

The hypothesis was largely based on a number of recent reports, when taken together are suggestive of a functional role and interplay between *PTEN* and PTENpg1 asRNA in progression of melanoma and in resistance mechanisms:

- *PTEN* is inactivated by epigenetic alterations in melanoma²²⁵.
- *PTEN* is suppressed in melanoma resistant to BRAF inhibitors²⁹².
- BRAF^{V600E} regulates a large number of lncRNAs²⁹⁷.
- DNMT3a is involved in melanoma tumorigenesis³³⁴.
- DNMT3a regulates chemotherapeutically induced apoptosis³³⁵.
- *PTEN* is suppressed by epigenetic alterations upon resistance to the tyrosine kinase inhibitor Imatinib³³¹.

4.2.2 Characterization of PTENpg1 asRNA in resistance to vemurafenib

To characterize the involvement of PTENpg1 asRNA in resistance to vemurafenib, we exposed the A375 melanoma to increasing amounts of vemurafenib and generated a number of resistant melanoma daughter cell lines. Interestingly, we found *PTEN* to be suppressed in all cell lines and strikingly, the expression of the PTENpg1 asRNA α isoform was induced. In support of our hypothesis, epigenetic alterations were also found at the *PTEN* promoter in resistant cells and a concurrent enrichment of EZH2 and the suppressive histone mark H3K27me3 was verified. No changes in total cellular levels were found for either DNMT3a, EZH2 or H3K27me3 and we therefore concluded that the suppressive marks were specifically induced at the *PTEN* promoter.

Next, we showed that ectopic expression of *PTEN* re-sensitized cells to vemurafenib treatment and thereby confirmed the involvement of *PTEN* in resistance. This also prompted us to investigate the expression of *PTEN* upon treatment with vemurafenib. We noticed that *PTEN* was induced simultaneously as *EZH2* was decreased in sensitive cells upon vemurafenib treatment. Interestingly, no such changes in expression were observed in the resistant cells. Moreover, ectopic expression of PTENpg1 asRNA made the sensitive cells more resistant to treatment. In summary, these data argue for a role of *PTEN* and PTENpg1 asRNA in resistance mechanisms to vemurafenib.

We argued that PTENpg1 asRNA α could be an interesting target due to its specific induction in the resistant cells. In paper I, siRNAs were originally used for targeting the PTENpg1 asRNA. However, siRNAs did not show efficient knockdown of PTENpg1 asRNA α in the melanoma cell lines. Instead, we turned to single stranded ASOs, more specifically gapmers (Exiqon), that are supposed to efficiently target nuclear expressed transcripts. Indeed, the knockdown of PTENpg1 asRNA α was improved, but satisfactory knockdown was not achieved and has to be further developed and optimized.

The modest reduction of PTENpg1 asRNA α levels by one gapmer, gap- α 3, induced apoptosis in the resistant cell line. Although all gapmers generated approximately the same degree of knockdown, two out of three gapmers did not induce apoptosis. Noteworthy, gap- α 3 targets the most upstream 5' end of PTENpg1 asRNA α . We therefore speculate that there might be low abundant upstream transcripts that are targeted more efficiently by gap- α 3. Taken together, this is supportive of PTENpg1 asRNA α being a promising therapeutic target. Most importantly, there is a distinct contrast between the sensitive and resistant cells as no induction of apoptosis was observed in the sensitive cells upon knockdown. This may indicate a therapeutic window for specific targeting of resistant melanoma cells, although these findings need to be further verified in future studies.

Additionally, we show that *PTEN* expression is induced upon simultaneous knockdown of *EZH2* and *DNMT3a*. Once again, the effects are specific for the resistant cells and no effect on *PTEN* is observed in the sensitive cells. The need for concurrent loss of *EZH2* and *DNMT3a* to see this effect on *PTEN* is a distinct feature of these resistant melanoma cells that we did not observe in paper I. This could either indicate that the individual knockdown is not efficient enough in these cells, or that they act in parallel pathways. Co-knockdown of PTENpg1 asRNA α in combination with either *DNMT3a* or *EZH2* could be informative and more precisely reveal the molecular events underlying suppression of *PTEN* upon resistance to vemurafenib.

The fact that expression of *PTEN* was induced upon knockdown of *EZH2* and *DNMT3a* prompted us to investigate whether resistance to vemurafenib is affected following knockdown. Indeed, the resistant cells became more sensitive to vemurafenib upon depletion of *EZH2* and *DNMT3a*. Drugs targeting *EZH2* as well as DNA methylation are available and it would be of great interest to try these in combination with vemurafenib especially during the development of treatment resistance³³⁶.

4.2.3 PTENpg1 asRNA expression in primary melanoma samples

In order to further dissect the role of PTENpg1 asRNA in melanoma we investigated its expression in two independent cohorts of primary melanoma samples from stage III first regional lymph node metastasis. Strikingly, we found that high expression of PTENpg1 asRNA correlated with poor survival. All samples investigated to date in this study were taken prior to oncological treatment, indicating that induced expression of PTENpg1 asRNA

was not the result of treatment. We speculate that tumors with high expression of PTENpg1 asRNA are more prone to develop resistance or form metastasis. In order to better understand these observations, it would be of interest to analyze the expression of *PTEN* and PTENpg1 asRNA before and after treatment with vemurafenib. Presently, we do not have access to such samples but they are being collected and these studies will be undertaken in the near future.

Moreover, generating stable cell lines with inducible expression of PTENpg1 asRNA α could also shed light on the involvement of PTENpg1 asRNA in melanoma. These cells could be used for xenograft experiments in mice in combination with vemurafenib and prove useful for further characterization of the PTENpg1 locus.

4.2.4 Concluding remarks of paper II

The findings in paper II are in line with our initial findings and are highly suggestive of a role of PTENpg1 asRNA in drug resistance to vemurafenib as well as in melanoma development. However, there is no doubt that more studies will have to be carried out in order to fully understand these issues.

It should be noted that our findings are in contrast to previous observations on the PTENpg1 locus in melanoma^{57,175}. We observe a concurrent induced expression of the PTENpg1 sense and PTENpg1 asRNA in our cell line models, but this does not overlap with induced expression of *PTEN*, which would be expected by the ceRNA model¹⁷⁷. The ceRNA model suggests a function where the PTENpg1 sense binds and acts as a decoy for *PTEN* related miRNAs. Based on this model, increased expression of PTENpg1 sense is therefore expected to induce the levels of *PTEN*. Instead, and in support of our observations in paper I, our data suggests the PTENpg1 asRNA transcript is the dominant regulator of *PTEN* expression through its suppressive function of *PTEN* transcription. The expression of PTENpg1 sense remains to be investigated in our clinical samples. At this point, we have not been able to address this question due to technical limitations. The samples taken from patients suffer from a certain degree of DNA contamination, which makes quantitative analysis of processed pseudogenes challenging.

4.3 PAPER 3

Expression of microRNA-1234 related signal transducer and activator of transcription 3 in patients with diffuse large B-cell lymphoma of activated B-cell like type from high and low infectious disease areas.

4.3.1 Rationale

Differential expression of *STAT3* had previously been reported in Swedish (endemic low infectious region) and Egyptian (endemic high infectious region) patients with ABC DLBCL. *STAT3* was found upregulated in the Swedish patient samples compared to the Egyptian³²⁶. miRNAs maintain regulatory functions of gene expression and *STAT3* has been reported to be under the regulatory control of miRNAs³¹⁵.

In paper III, we therefore hypothesized that differential expression of miRNAs, possibly reflecting various environmental factors including HBV, HCV as well as EBV³¹⁷⁻³¹⁹, may contribute to the differential expression of *STAT3*. We therefore set out to investigate the expression profile of miRNAs in lymphoma samples from patients with ABC DLBCL from these two different geographical areas. Such differences may associate with various pathogenic mechanisms in ABC DLBCL and furthermore also be relevant for targeted therapies.

4.3.2 Characterization of miRNA expression

The profiling revealed 21 miRNAs with distinct expression between the cohorts. Due to previous findings showing that *STAT3* is differentially expressed in lymphoma cells between these two groups of patients³²⁶, we specifically focused on miRNAs predicted to target *STAT3*. Activation of *STAT3* is an important event for cytokine and growth-factor induced proliferation, differentiation, survival and inflammation and is indeed frequently induced in many types of tumors including lymphomas. Among the miRNAs with differential expression, miR-1234 was found to target *STAT3*. Strikingly, low expression of miR-1234 in the samples from the Swedish patients correlated with increased expression of *STAT3*, while the opposite was observed in the samples from the Egyptian patients (high miR-1234, low *STAT3*). Moreover, ectopic expression of miR-1234 in cell line model systems verified its suppressive effect on *STAT3* expression. Taken together, the data suggested that miR-1234 regulates *STAT3* levels.

miR-1234 is a mirtron localized in one of the introns of the Cleavage and polyadenylation specific factor 1 (CPSF1) host gene²⁵². A coordinated and concordant expression of CPSF1 and miR-1234 is therefore likely but remains to be investigated. The CPSF1 protein has been established to be involved in cleavage and polyadenylation synthesis of pre-mRNAs³³⁷. Beyond a suggested role for miR-1234 in metastasis of breast cancer, not much is known about the function of this miRNA³³⁸. Our study is indeed to our knowledge the first to suggest

that miR-1234 acts to target *STAT3*, but further investigations are required in order to explore the function of miR-1234 and its other putative mRNA targets.

4.3.3 Concluding remarks of paper III

The mechanisms of miR-1234 regulation remain to be studied in greater detail. The regulation of the CPSF1/miR-1234 promoter by factors associated with inflammation and infectious disease load should be a valid starting point in order to better understand the interplay of *STAT3* and miR-1234. Our preliminary results indicate the presence of HBV in Egyptian patients but not in Swedish patients but these observations will need further characterization.

Taken together, increased understanding of molecular processes involved in the development of ABC DLBCL will be important in order to better understand the underlying causes to the disease. Pharmacological approaches are currently being undertaken in order to modulate *STAT3* activity (reviewed in³³⁹). Differences in *STAT3* expression may therefore be of great interest and serve as predictive biomarkers for efficient response to such drugs. Understanding the regulation of *STAT3* and its underlying mechanisms may therefore prove informative in selection of treatment and prediction of response and overall outcome.

5. CONCLUDING REMARKS

The most important findings in this thesis are the following:

- PTENpg1 asRNAs regulate expression of the protein-coding gene *PTEN*.
- DNMT3a is recruited to the targeted *PTEN* locus by a lncRNA.
- PTENpg1 asRNA is involved in the resistance to the BRAF inhibitor vemurafenib and serves as a putative prognostic marker for the outcome in melanoma patients.
- A differential expression of miR-1234, a regulator of *STAT3* expression, is observed in tumor samples from ABC DLBCL exposed to various environmental factors.

As illustrated by the work presented in this thesis, ncRNAs maintain crucial roles in gene regulation and there is no longer any doubt about their involvement in disease development. One of the greatest challenges in today's cancer therapies is avoiding side effects caused by targeting of healthy tissue. In this aspect, ncRNAs hold great promise due to the advantage of greater tissue specificity when compared to proteins and their manipulation may therefore result in less damage to normal tissue¹³. Moreover, targeting of lncRNAs may also be advantageous based on specific nucleotide pairing.

Many subgroups of lncRNAs have been proposed over the last ~10 years. Most importantly, this thesis highlights a role for pseudogene-encoded asRNAs in control of their protein-coding ancestor. Pseudogenes have for long been considered as 'junk DNA' but there are now growing evidence of an emerging role of both sense and asRNA expressed from pseudogenes. The classification of pseudogenes as 'junk DNA' should be reconsidered since it is now standing clear that at least some of these non-coding genes are expressed and maintain important regulatory functions. Recent reports suggest thousands of pseudogenes being transcriptionally active as sense transcripts^{53,176}, but whether these are also generally expressed as asRNAs remain unknown. Genome-wide approaches for the identification of pseudogene-encoded asRNAs will initially be needed in order to increase our understanding of this novel category of regulatory lncRNAs. The discovery of PTENpg1 asRNA presented herein should stimulate such genome-wide profiling as well as functional characterization.

Furthermore, this thesis sheds light on several of the functional aspects of ncRNAs and we present regulatory functions consisting of RNA:RNA as well as RNA:protein interactions. One of the major challenges in understanding the function of ncRNAs is their great heterogeneity. While mRNAs predominantly serve as templates for translation, the multifunctional aspects of ncRNAs are widespread and there is as of yet no universal code to decipher for the functions of lncRNAs. Revealing the 'functional code' of lncRNAs will indeed be a great challenge in the near future. The development of genome-wide technologies made it possible to characterize thousands of ncRNAs and functional investigations are now needed in order to better understand the functional properties they maintain. Such discoveries may both lead to the development of novel disease biomarkers, as well as serve as a platform for drug development targeting these systems.

6. ACKNOWLEDGEMENTS

I would like to express my gratefulness to the following people:

My main supervisor **Dan**, your hands off, non-micro-management styled supervision has been a lot of fun to experience during my PhD. It has been a pleasure to spend these years in your group and I think you have given me the greatest opportunity to grow into an independent researcher. Our journey together has been a great experience, although challenging from time to time (as it should be). Our field trips to Miami and NY were amusing; NY-shopping is more convenient when we have the same taste in clothes... Most importantly, you are a great person who always looks after the members in your group and make sure everyone is enjoying their time. Thank you for all these years, it has been a great pleasure to work together!

Kevin, the first time we met you showed me how to “duck dive” on the floor in the lab. The second time we met you showed me the QIAcube and the qRT-PCR. Needless to say, your second impression influenced me more since I still have not done my first duck dive but countless QIAcube/qRT-PCR analysis... You thought that you did not like jello(yellow)-shots before you met the “San Diego Swedes” but I think we have proven you wrong. Your never-ending energy and passion for science is remarkable! Your desire to constantly find out how things (RNA) work has been truly motivating to experience and when it has been a struggle, it’s cheerful to hear “if it was easy, someone would have done it already”. Thanks for good and fun times and thanks also to **Paula** and **Harper** for letting me stay on your sofa and enjoy the ocean view!

Martin and **Andor**, my co-supervisors for sharing your knowledge and being available when needed; Martin for teaching me everything useful and non-useful about primer design and Andor for showing me how to extract RNA from deep-frozen tissues.

Katja, for being a truly amusing and kind person with a lot of patience! The lab would not run as smoothly without you; there is always a colorful Russian pill in your drawers when headache is present. Thanks for everything over these years!

Present and former colleagues of the Grandér group for all your support and fun castle kick-offs; **Linda**, **Jason**, **Matheus**, **Iryna**, **Mikael**, **Moritz**, **Elin**, **Giuseppe**, **Lotte**, **Katja**, **Tiago**, **Santiago**, **Caroline**, and all the master students over the years!

Therese, who would have thought that our short collaboration would be such a great success! Thanks for including me into your project!

Many thanks to the Farnebo group (including **Marianne**); **Alex, Hanif, Fredrik, Elisabeth, Soniya** and **Christos**, for letting me join your social activities. It's been a lot of fun!

Thanks also to **Cinzia, Dudi** and **Sylvia** on the "4th floor". You are next!

Thanks to colleagues at the Kevin Morris lab, branch TSRI: **Amanda, Sheena, Stuart, Mike, Anne-Marie**, (did we ever find the best California burrito?), **Peter** and **Sharon**.

-**Sheena**; I have never been as terrified as I was when I joined you to buy a car... Let me know next time you are buying a car and I will find a good deal for you!

-**Amanda**; Thanks for all the Snickers (we actually do have them in Sweden as well). I hope you got to see the zebras in **Sheena's** backyard during your trip to South Africa.

-**Stuart**; I like your way of making coffee, but not your taste in music. Go ABBA!

-**Anne-Marie**, I am slowly learning how to use "the" and "," in a proper way. With your supervision I started to "scratch on the top of the iceberg" of lncRNAs!

Also many thanks to colleagues of the Kevin Morris branch UNSW for helping me out in Sydney; **Matt, Nick, Caio, Albert, Dave, Jessica** and **Rosie**.

Mike and **Jessica**, you thought me one invaluable, lifelong, lesson in Sydney; how to peel a mango! Thanks also for introducing us to the daily life in Sydney.

My greatest thanks to my family for all their support: **Mamma, Pappa, Olov, Lars, Fredrika, Linn, Märta, Oscar, Clara** and **Tilda**.

-**Lars** who introduced me to the exciting world of doing science. I cannot believe that I am still in the same business after been chopping fish food for days at SLU.

Also to the "Mallis 2014 gang", Henriksson/Rosell families: **Hans, Elisabeth, Erika, Johan, Adrian** and lilla **Sigrid**.

Martin, I finally got to meet the man behind the yellow colors outside the football field when we watched *Fucking Åmål* together... It's been fun all the way from Bälinge IP (although I do not remember that goal you scored once and keep talking about) to Katedralskolan, TSRI and now PhD (finally I will join you...). Thank you for all good times!

Also thanks to: **Gustaf, Calle, David, Arash, Kamil, Oscar, Daniel** and **Pär**, and many more!!

Most important, my love **Sofia**. I was getting coffee when you passed with your samples in a bucket of ice when we first met each other. The best life-changing 3 seconds in my life! You thought I was doing a post-doc already at that time (I am almost there now, 5 years later). This thesis is for you. It would not have been possible without your loving, encouraging and endless support. "You light my morning sky, With burning love". I love you!

7. REFERENCES

1. Venter, J.C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
2. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
3. Roest Crolius, H. et al. Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence. *Nat Genet* **25**, 235-8 (2000).
4. Goodstadt, L. & Ponting, C.P. Phylogenetic reconstruction of orthology, paralogy, and conserved synteny for dog and human. *PLoS Comput Biol* **2**, e133 (2006).
5. Hahn, M.W. & Wray, G.A. The g-value paradox. *Evol Dev* **4**, 73-5 (2002).
6. International Chicken Genome Sequencing, C. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716 (2004).
7. Mouse Genome Sequencing, C. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-62 (2002).
8. Taft, R.J., Pheasant, M. & Mattick, J.S. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**, 288-299 (2007).
9. Carninci, P. et al. The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-63 (2005).
10. Katayama, S. et al. Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564-6 (2005).
11. Collins, F.S., Lander, E.S., Rogers, J., Waterston, R.H. & Conso, I.H.G.S. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).
12. Consortium, E.P. et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
13. Djebali, S. et al. Landscape of transcription in human cells. *Nature* **489**, 101-8 (2012).
14. Derrien, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22**, 1775-89 (2012).
15. Guttman, M. et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223-227 (2009).
16. Brown, C.J. et al. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527-42 (1992).
17. Brockdorff, N. et al. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515-26 (1992).
18. Brannan, C.I., Dees, E.C., Ingram, R.S. & Tilghman, S.M. The product of the H19 gene may function as an RNA. *Mol Cell Biol* **10**, 28-36 (1990).
19. Zhang, Y. & Tycko, B. Monoallelic expression of the human H19 gene. *Nat Genet* **1**, 40-4 (1992).
20. Leighton, P.A., Saam, J.R., Ingram, R.S., Stewart, C.L. & Tilghman, S.M. An Enhancer Deletion Affects Both H19 and Igf2 Expression. *Genes & Development* **9**, 2079-2089 (1995).
21. Yu, W. et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* **451**, 202-6 (2008).
22. Yap, K.L. et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* **38**, 662-74 (2010).

23. Beltran, M. et al. A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev* **22**, 756-69 (2008).
24. Kapranov, P. et al. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916-9 (2002).
25. Palade, G.E. A small particulate component of the cytoplasm. *J Biophys Biochem Cytol* **1**, 59-68 (1955).
26. Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I. & Zamecnik, P.C. A soluble ribonucleic acid intermediate in protein synthesis. *J Biol Chem* **231**, 241-57 (1958).
27. Lerner, M. et al. DLEU2, frequently deleted in malignancy, functions as a critical host gene of the cell cycle inhibitory microRNAs miR-15a and miR-16-1. *Exp Cell Res* **315**, 2941-52 (2009).
28. Calin, G.A. et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* **99**, 15524-9 (2002).
29. Gupta, R.A. et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071-6 (2010).
30. Lee, J.T., Davidow, L.S. & Warshawsky, D. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet* **21**, 400-4 (1999).
31. Taft, R.J. et al. Tiny RNAs associated with transcription start sites in animals. *Nat Genet* **41**, 572-8 (2009).
32. Mattick, J.S. Introns: evolution and function. *Curr Opin Genet Dev* **4**, 823-31 (1994).
33. Sado, T., Wang, Z., Sasaki, H. & Li, E. Regulation of imprinted X-chromosome inactivation in mice by Tsix. *Development* **128**, 1275-86 (2001).
34. Rinn, J.L. et al. The transcriptional activity of human Chromosome 22. *Genes Dev* **17**, 529-40 (2003).
35. White, E., Schlackow, M., Kamieniarz-Gdula, K., Proudfoot, N.J. & Gullerova, M. Human nuclear Dicer restricts the deleterious accumulation of endogenous double-stranded RNA. *Nat Struct Mol Biol* **21**, 552-9 (2014).
36. Kawai, J. et al. Functional annotation of a full-length mouse cDNA collection. *Nature* **409**, 685-90 (2001).
37. Carninci, P. et al. Targeting a complex transcriptome: the construction of the mouse full-length cDNA encyclopedia. *Genome Res* **13**, 1273-89 (2003).
38. Okazaki, Y. et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563-73 (2002).
39. Shiraki, T. et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A* **100**, 15776-81 (2003).
40. Consortium, F. et al. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat Genet* **41**, 553-62 (2009).
41. Ravasi, T. et al. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**, 744-52 (2010).
42. Faulkner, G.J. et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* **41**, 563-71 (2009).
43. Taft, R.J., Hawkins, P.G., Mattick, J.S. & Morris, K.V. The relationship between transcription initiation RNAs and CCCTC-binding factor (CTCF) localization. *Epigenetics Chromatin* **4**, 13 (2011).
44. Margulies, E.H. et al. Comparative sequencing provides insights about the structure and conservation of marsupial and monotreme genomes. *Proc Natl Acad Sci U S A* **102**, 3354-9 (2005).

45. Levy, S. et al. The diploid genome sequence of an individual human. *PLoS Biol* **5**, e254 (2007).
46. Consortium, F. et al. A promoter-level mammalian expression atlas. *Nature* **507**, 462-70 (2014).
47. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-61 (2014).
48. Consortium, E.P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636-40 (2004).
49. Consortium, E.P. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
50. Dong, X. et al. Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol* **13**, R53 (2012).
51. Banfai, B. et al. Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res* **22**, 1646-57 (2012).
52. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-74 (2012).
53. Pei, B.K. et al. The GENCODE pseudogene resource. *Genome Biology* **13**(2012).
54. Mikkelsen, T.S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60 (2007).
55. Morris, K.V. & Mattick, J.S. The rise of regulatory RNA. *Nat Rev Genet* **15**, 423-37 (2014).
56. Calin, G.A. et al. Ultraconserved regions encoding ncRNAs are, altered in human leukemias and carcinomas. *Cancer Cell* **12**, 215-229 (2007).
57. Poliseno, L. et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033-U90 (2010).
58. Faghihi, M.A. et al. RNAi screen indicates widespread biological function for human natural antisense transcripts. *PLoS One* **5**(2010).
59. Rinn, J.L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by Noncoding RNAs. *Cell* **129**, 1311-1323 (2007).
60. Hacisuleyman, E. et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol* **21**, 198-206 (2014).
61. Wang, P. et al. The STAT3-binding long noncoding RNA Inc-DC controls human dendritic cell differentiation. *Science* **344**, 310-3 (2014).
62. Tseng, Y.Y. et al. PVT1 dependence in cancer with MYC copy-number increase. *Nature* (2014).
63. Wagner, E.G. & Simons, R.W. Antisense RNA control in bacteria, phages, and plasmids. *Annu Rev Microbiol* **48**, 713-42 (1994).
64. Vanhee-Brossollet, C. & Vaquero, C. Do natural antisense transcripts make sense in eukaryotes? *Gene* **211**, 1-9 (1998).
65. Wutz, A. et al. Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* **389**, 745-749 (1997).
66. Chen, J. et al. Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res* **32**, 4812-20 (2004).
67. Engstrom, P.G. et al. Complex Loci in human and mouse genomes. *PLoS Genet* **2**, e47 (2006).
68. Mahmoudi, S. et al. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. *Mol Cell* **33**, 462-71 (2009).
69. Faghihi, M.A. et al. Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol* **11**, R56 (2010).
70. Morris, K.V., Santoso, S., Turner, A.M., Pastori, C. & Hawkins, P.G. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet* **4**, e1000258 (2008).

71. Hawkins, P.G. & Morris, K.V. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription* **1**, 165-175 (2010).
72. Brown, C.J. et al. A Gene from the Region of the Human X-Inactivation Center Is Expressed Exclusively from the Inactive X-Chromosome. *Nature* **349**, 38-44 (1991).
73. Borsani, G. et al. Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**, 325-9 (1991).
74. Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131-137 (1996).
75. Wutz, A., Rasmussen, T.P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nature Genetics* **30**, 167-174 (2002).
76. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. & Lee, J.T. Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. *Science* **322**, 750-756 (2008).
77. Shibata, S. & Lee, J.T. Characterization and quantitation of differential Tsix transcripts: implications for Tsix function. *Hum Mol Genet* **12**, 125-36 (2003).
78. Ogawa, Y., Sun, B.K. & Lee, J.T. Intersection of the RNA interference and X-inactivation pathways. *Science* **320**, 1336-41 (2008).
79. Sado, T., Hoki, Y. & Sasaki, H. Tsix silences Xist through modification of chromatin structure. *Dev Cell* **9**, 159-65 (2005).
80. Sun, B.K., Deaton, A.M. & Lee, J.T. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol Cell* **21**, 617-28 (2006).
81. Jeon, Y. & Lee, J.T. YY1 Tethers Xist RNA to the Inactive X Nucleation Center. *Cell* **146**, 119-133 (2011).
82. Sun, S. et al. Jpx RNA activates Xist by evicting CTCF. *Cell* **153**, 1537-51 (2013).
83. Tian, D., Sun, S. & Lee, J.T. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* **143**, 390-403 (2010).
84. Filippova, G.N. et al. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol* **16**, 2802-13 (1996).
85. Chu, C., Qu, K., Zhong, F.L., Artandi, S.E. & Chang, H.Y. Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Molecular Cell* **44**, 667-678 (2011).
86. Tsai, M.C. et al. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science* **329**, 689-693 (2010).
87. Lunyak, V.V. et al. Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* **298**, 1747-52 (2002).
88. Schneider, R. et al. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol* **6**, 73-7 (2004).
89. Di Stefano, L., Ji, J.Y., Moon, N.S., Herr, A. & Dyson, N. Mutation of Drosophila Lsd1 disrupts H3-K4 methylation, resulting in tissue-specific defects during development. *Curr Biol* **17**, 808-12 (2007).
90. Bernstein, B.E. et al. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169-81 (2005).
91. Stoger, R. et al. Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell* **73**, 61-71 (1993).
92. Sleutels, F., Zwart, R. & Barlow, D.P. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810-3 (2002).

93. Latos, P.A. et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* **338**, 1469-72 (2012).
94. Seidl, C.I., Stricker, S.H. & Barlow, D.P. The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J* **25**, 3565-75 (2006).
95. Nagano, T. et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717-20 (2008).
96. Pasmant, E. et al. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res* **67**, 3963-9 (2007).
97. Kotake, Y. et al. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* **30**, 1956-62 (2011).
98. Faghihi, M.A. et al. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med* **14**, 723-30 (2008).
99. Elloul, S. et al. Snail, Slug, and Smad-interacting protein 1 as novel parameters of disease aggressiveness in metastatic ovarian and breast carcinoma. *Cancer* **103**, 1631-43 (2005).
100. Rosivatz, E. et al. Differential expression of the epithelial-mesenchymal transition regulators snail, SIP1, and twist in gastric cancer. *Am J Pathol* **161**, 1881-91 (2002).
101. Nelles, L., Van de Putte, T., van Grunsven, L., Huylebroeck, D. & Verschueren, K. Organization of the mouse Zfx1b gene encoding the two-handed zinc finger repressor Smad-interacting protein-1. *Genomics* **82**, 460-9 (2003).
102. Gong, C.G. & Maquat, L.E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470**, 284-+ (2011).
103. Yeh, E. et al. A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. *Nat Cell Biol* **6**, 308-18 (2004).
104. Merad, M. PU.1 takes control of the dendritic cell lineage. *Immunity* **32**, 583-5 (2010).
105. Guerriero, A., Langmuir, P.B., Spain, L.M. & Scott, E.W. PU.1 is required for myeloid-derived but not lymphoid-derived dendritic cells. *Blood* **95**, 879-85 (2000).
106. Laouar, Y., Welte, T., Fu, X.Y. & Flavell, R.A. STAT3 is required for Flt3L-dependent dendritic cell differentiation. *Immunity* **19**, 903-12 (2003).
107. Batzer, M.A. & Deininger, P.L. Alu repeats and human genomic diversity. *Nat Rev Genet* **3**, 370-9 (2002).
108. Houck, C.M., Rinehart, F.P. & Schmid, C.W. A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol* **132**, 289-306 (1979).
109. Hasler, J. & Strub, K. Alu elements as regulators of gene expression. *Nucleic Acids Res* **34**, 5491-7 (2006).
110. Kim, Y.K. et al. Stauf1 regulates diverse classes of mammalian transcripts. *EMBO J* **26**, 2670-81 (2007).
111. Graur, D. et al. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol Evol* **5**, 578-90 (2013).
112. Brosius, J. Waste not, want not--transcript excess in multicellular eukaryotes. *Trends Genet* **21**, 287-8 (2005).
113. Struhl, K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nature Structural & Molecular Biology* **14**, 103-105 (2007).

114. Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F. & Mattick, J.S. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci U S A* **105**, 716-21 (2008).
115. Pang, K.C., Frith, M.C. & Mattick, J.S. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in Genetics* **22**, 1-5 (2006).
116. Johnsson, P., Lipovich, L., Grander, D. & Morris, K.V. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta* **1840**, 1063-71 (2014).
117. Chureau, C. et al. Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Res* **12**, 894-908 (2002).
118. Nesterova, T.B. et al. Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Research* **11**, 833-849 (2001).
119. Migeon, B.R., Lee, C.H., Chowdhury, A.K. & Carpenter, H. Species differences in TSIX/Tsix reveal the roles of these genes in X-chromosome inactivation. *Am J Hum Genet* **71**, 286-93 (2002).
120. Oudejans, C.B. et al. Allelic IGF2R repression does not correlate with expression of antisense RNA in human extraembryonic tissues. *Genomics* **73**, 331-7 (2001).
121. Schorderet, P. & Duboule, D. Structural and Functional Differences in the Long Non-Coding RNA Hotair in Mouse and Human. *Plos Genetics* **7**(2011).
122. Cao, R. et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-43 (2002).
123. Simon, J.A. & Kingston, R.E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* **10**, 697-708 (2009).
124. Schuettengruber, B. & Cavalli, G. Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* **136**, 3531-42 (2009).
125. Zee, B.M. et al. In vivo residue-specific histone methylation dynamics. *J Biol Chem* **285**, 3341-50 (2010).
126. Khalil, A.M. et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* **106**, 11667-72 (2009).
127. Kaneko, S., Son, J., Shen, S.S., Reinberg, D. & Bonasio, R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat Struct Mol Biol* **20**, 1258-64 (2013).
128. Davidovich, C., Zheng, L., Goodrich, K.J. & Cech, T.R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat Struct Mol Biol* **20**, 1250-7 (2013).
129. Vire, E. et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871-4 (2006).
130. Yen, R.W. et al. Isolation and characterization of the cDNA encoding human DNA methyltransferase. *Nucleic Acids Res* **20**, 2287-91 (1992).
131. Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol* **203**, 971-83 (1988).
132. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* **19**, 219-20 (1998).
133. Okano, M., Bell, D.W., Haber, D.A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-57 (1999).
134. Jurkowska, R.Z., Jurkowski, T.P. & Jeltsch, A. Structure and function of mammalian DNA methyltransferases. *Chembiochem* **12**, 206-22 (2011).

135. Denis, H., Ndlovu, M.N. & Fuks, F. Regulation of mammalian DNA methyltransferases: a route to new mechanisms. *EMBO Rep* **12**, 647-56 (2011).
136. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* **10**, 295-304 (2009).
137. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-26 (1992).
138. Bird, A.P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-13 (1986).
139. Larsen, F., Gundersen, G., Lopez, R. & Prydz, H. CpG Islands as Gene Markers in the Human Genome. *Genomics* **13**, 1095-1107 (1992).
140. Saxonov, S., Berg, P. & Brutlag, D.L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**, 1412-7 (2006).
141. Takai, D. & Jones, P.A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* **99**, 3740-5 (2002).
142. Jones, P.A. & Baylin, S.B. The epigenomics of cancer. *Cell* **128**, 683-92 (2007).
143. Hodges, E. et al. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res* **19**, 1593-605 (2009).
144. Meissner, A. et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-70 (2008).
145. Fuks, F., Burgers, W.A., Godin, N., Kasai, M. & Kouzarides, T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J* **20**, 2536-44 (2001).
146. Datta, J., Ghoshal, K., Sharma, S.M., Tajima, S. & Jacob, S.T. Biochemical fractionation reveals association of DNA methyltransferase (Dnmt) 3b with Dnmt1 and that of Dnmt 3a with a histone H3 methyltransferase and Hdac1. *J Cell Biochem* **88**, 855-64 (2003).
147. Hawkins, P.G., Santoso, S., Adams, C., Anest, V. & Morris, K.V. Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells. *Nucleic Acids Res* **37**, 2984-95 (2009).
148. Feldman, N. et al. G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* **8**, 188-94 (2006).
149. Jeffery, L. & Nakielny, S. Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J Biol Chem* **279**, 49479-87 (2004).
150. Di Ruscio, A. et al. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371-6 (2013).
151. Morris, K.V., Chan, S.W.L., Jacobsen, S.E. & Looney, D.J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289-1292 (2004).
152. Weinberg, M.S. et al. The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. *RNA* **12**, 256-262 (2006).
153. Schmitz, K.M., Mayer, C., Postepska, A. & Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev* **24**, 2264-9 (2010).
154. Jair, K.W. et al. De novo CpG island methylation in human cancer cells. *Cancer Res* **66**, 682-92 (2006).
155. Wang, Y.A. et al. DNA methyltransferase-3a interacts with p53 and represses p53-mediated gene expression. *Cancer Biol Ther* **4**, 1138-43 (2005).
156. Ball, M.P. et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* **27**, 361-8 (2009).

157. Laurent, L. et al. Dynamic changes in the human methylome during differentiation. *Genome Res* **20**, 320-31 (2010).
158. Suzuki, M.M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465-76 (2008).
159. Jacq, C., Miller, J.R. & Brownlee, G.G. A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell* **12**, 109-20 (1977).
160. Harrison, P.M. et al. Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Res* **12**, 272-80 (2002).
161. Ohshima, K. et al. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biology* **4**(2003).
162. Zhang, Z.L., Carriero, N. & Gerstein, M. Comparative analysis of processed pseudogenes in the mouse and human genomes. *Trends in Genetics* **20**, 62-67 (2004).
163. Balasubramanian, S. et al. Comparative analysis of processed ribosomal protein pseudogenes in four mammalian genomes. *Genome Biol* **10**, R2 (2009).
164. Liu, Y.J. et al. Comprehensive analysis of the pseudogenes of glycolytic enzymes in vertebrates: the anomalously high number of GAPDH pseudogenes highlights a recent burst of retrotrans-positional activity. *BMC Genomics* **10**, 480 (2009).
165. Khachane, A.N. & Harrison, P.M. Assessing the genomic evidence for conserved transcribed pseudogenes under selection. *Bmc Genomics* **10**(2009).
166. Zakut-Houri, R. et al. A single gene and a pseudogene for the cellular tumour antigen p53. *Nature* **306**, 594-7 (1983).
167. Dahia, P.L. et al. A highly conserved processed PTEN pseudogene is located on chromosome band 9p21. *Oncogene* **16**, 2403-6 (1998).
168. Puget, N. et al. Distinct BRCA1 rearrangements involving the BRCA1 pseudogene suggest the existence of a recombination hot spot. *Am J Hum Genet* **70**, 858-65 (2002).
169. Pain, D., Chirn, G.W., Strassel, C. & Kemp, D.M. Multiple retropseudogenes from pluripotent cell-specific gene expression indicates a potential signature for novel gene identification. *J Biol Chem* **280**, 6265-8 (2005).
170. Suo, G. et al. Oct4 pseudogenes are transcribed in cancers. *Biochem Biophys Res Commun* **337**, 1047-51 (2005).
171. Fujii, G.H., Morimoto, A.M., Berson, A.E. & Bolen, J.B. Transcriptional analysis of the PTEN/MMAC1 pseudogene, Psi PTEN. *Oncogene* **18**, 1765-1769 (1999).
172. Tian, L., Fang, Y.X., Xue, J.L. & Chen, J.Z. Four microRNAs promote prostate cell proliferation with regulation of PTEN and its downstream signals in vitro. *PLoS One* **8**, e75885 (2013).
173. Huse, J.T. et al. The PTEN-regulating microRNA miR-26a is amplified in high-grade glioma and facilitates gliomagenesis in vivo. *Genes Dev* **23**, 1327-37 (2009).
174. Ebert, M.S., Neilson, J.R. & Sharp, P.A. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods* **4**, 721-6 (2007).
175. Poliseno, L. et al. Deletion of PTENP1 pseudogene in human melanoma. *J Invest Dermatol* **131**, 2497-500 (2011).
176. Kalyana-Sundaram, S. et al. Expressed Pseudogenes in the Transcriptional Landscape of Human Cancers. *Cell* **149**, 1622-1634 (2012).
177. Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P.P. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146**, 353-8 (2011).
178. Cesana, M. et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147**, 358-69 (2011).

179. Hansen, T.B. et al. Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384-8 (2013).
180. Tay, Y. et al. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* **147**, 344-57 (2011).
181. Chiefari, E. et al. Pseudogene-mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and type 2 diabetes. *Nat Commun* **1**, 40 (2010).
182. Watanabe, T. et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**, 539-43 (2008).
183. Tam, O.H. et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**, 534-8 (2008).
184. Muro, E.M. & Andrade-Navarro, M.A. Pseudogenes as an alternative source of natural antisense transcripts. *Bmc Evolutionary Biology* **10**(2010).
185. Zhou, B.S., Beidler, D.R. & Cheng, Y.C. Identification of antisense RNA transcripts from a human DNA topoisomerase I pseudogene. *Cancer Res* **52**, 4280-5 (1992).
186. Weil, D., Power, M.A., Webb, G.C. & Li, C.L. Antisense transcription of a murine FGFR-3 pseudogene during fetal development. *Gene* **187**, 115-22 (1997).
187. Korneev, S.A., Park, J.H. & O'Shea, M. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J Neurosci* **19**, 7711-20 (1999).
188. Tam, O.H. et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**, 534-U8 (2008).
189. Watanabe, T. et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**, 539-U9 (2008).
190. Stein, P., Zeng, F.Y., Pan, H. & Schultz, R.M. Absence of non-specific effects of RNA interference triggered by long double-stranded RNA in mouse oocytes. *Developmental Biology* **286**, 464-471 (2005).
191. Chan, W.L., Yang, W.K., Huang, H.D. & Chang, J.G. pseudoMap: an innovative and comprehensive resource for identification of siRNA-mediated mechanisms in human transcribed pseudogenes. *Database (Oxford)* **2013**, bat001 (2013).
192. Wahlestedt, C. et al. Potent and nontoxic antisense oligonucleotides containing locked nucleic acids. *Proc Natl Acad Sci U S A* **97**, 5633-8 (2000).
193. Berezhna, S.Y., Supekova, L., Supek, F., Schultz, P.G. & Deniz, A.A. siRNA in human cells selectively localizes to target RNA sites. *Proc Natl Acad Sci U S A* **103**, 7682-7 (2006).
194. Gagnon, K.T., Li, L., Chu, Y., Janowski, B.A. & Corey, D.R. RNAi factors are present and active in human cell nuclei. *Cell Rep* **6**, 211-21 (2014).
195. Lee, N.S. et al. Functional and intracellular localization properties of U6 promoter-expressed siRNAs, shRNAs, and chimeric VA1 shRNAs in mammalian cells. *RNA* **14**, 1823-33 (2008).
196. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
197. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).
198. Prensner, J.R. et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* **29**, 742-9 (2011).
199. Ge, X. et al. Overexpression of long noncoding RNA PCAT-1 is a novel biomarker of poor prognosis in patients with colorectal cancer. *Med Oncol* **30**, 588 (2013).
200. Prensner, J.R. et al. PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. *Cancer Res* **74**, 1651-60 (2014).
201. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C.C. p53 mutations in human cancers. *Science* **253**, 49-53 (1991).

202. Weber, J.D., Taylor, L.J., Roussel, M.F., Sherr, C.J. & Bar-Sagi, D. Nucleolar Arf sequesters Mdm2 and activates p53. *Nat Cell Biol* **1**, 20-6 (1999).
203. Lohrum, M.A., Ashcroft, M., Kubbutat, M.H. & Vousden, K.H. Identification of a cryptic nucleolar-localization signal in MDM2. *Nat Cell Biol* **2**, 179-81 (2000).
204. Gajjar, M. et al. The p53 mRNA-Mdm2 interaction controls Mdm2 nuclear trafficking and is required for p53 activation following DNA damage. *Cancer Cell* **21**, 25-35 (2012).
205. Le, M.T. et al. MicroRNA-125b is a novel negative regulator of p53. *Genes Dev* **23**, 862-76 (2009).
206. Counter, C.M. et al. Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J* **11**, 1921-9 (1992).
207. Shay, J.W. & Bacchetti, S. A survey of telomerase activity in human cancer. *Eur J Cancer* **33**, 787-91 (1997).
208. Feng, J. et al. The RNA component of human telomerase. *Science* **269**, 1236-41 (1995).
209. Brabletz, T. To differentiate or not--routes towards metastasis. *Nat Rev Cancer* **12**, 425-36 (2012).
210. Cavallaro, U. & Christofori, G. Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. *Nat Rev Cancer* **4**, 118-32 (2004).
211. Kogo, R. et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res* **71**, 6320-6 (2011).
212. Geng, Y.J., Xie, S.L., Li, Q., Ma, J. & Wang, G.Y. Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. *J Int Med Res* **39**, 2119-28 (2011).
213. Liu, X.H. et al. Lnc RNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer. *Mol Cancer* **13**, 92 (2014).
214. Uchida, T. et al. Prolonged hypoxia differentially regulates hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha expression in lung epithelial cells: implication of natural antisense HIF-1alpha. *J Biol Chem* **279**, 14871-8 (2004).
215. Thrash-Bingham, C.A. & Tartof, K.D. aHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia. *J Natl Cancer Inst* **91**, 143-51 (1999).
216. Rossignol, F., Vache, C. & Clottes, E. Natural antisense transcripts of hypoxia-inducible factor 1alpha are detected in different normal and tumour human tissues. *Gene* **299**, 135-40 (2002).
217. Khaitan, D. et al. The melanoma-upregulated long noncoding RNA SPRY4-IT1 modulates apoptosis and invasion. *Cancer Res* **71**, 3852-62 (2011).
218. Leeksa, O.C. et al. Human sprouty 4, a new ras antagonist on 5q31, interacts with the dual specificity kinase TESK1. *Eur J Biochem* **269**, 2546-56 (2002).
219. Tennis, M.A. et al. Sprouty-4 inhibits transformed cell growth, migration and invasion, and epithelial-mesenchymal transition, and is regulated by Wnt7A through PPARgamma in non-small cell lung cancer. *Mol Cancer Res* **8**, 833-43 (2010).
220. Yan, X.J. et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet* **43**, 309-15 (2011).
221. Yamashita, Y. et al. Array-based genomic resequencing of human leukemia. *Oncogene* **29**, 3723-31 (2010).
222. Ley, T.J. et al. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424-33 (2010).

223. Varambally, S. et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624-9 (2002).
224. Kleer, C.G. et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* **100**, 11606-11 (2003).
225. Morin, R.D. et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* **42**, 181-5 (2010).
226. Sneeringer, C.J. et al. Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc Natl Acad Sci U S A* **107**, 20980-5 (2010).
227. Shah, N. & Sukumar, S. The Hox genes and their roles in oncogenesis. *Nat Rev Cancer* **10**, 361-71 (2010).
228. Qu, Y. et al. Differential methylation in CN-AML preferentially targets non-CGI regions and is dictated by DNMT3A mutational status and associated with predominant hypomethylation of HOX genes. *Epigenetics* **9**(2014).
229. Simon, C. et al. A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. *Genes Dev* **26**, 651-6 (2012).
230. Lee, R.C., Feinbaum, R.L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843-54 (1993).
231. Reinhart, B.J. et al. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901-6 (2000).
232. Pasquinelli, A.E. et al. Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature* **408**, 86-89 (2000).
233. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20 (2005).
234. Friedman, R.C., Farh, K.K., Burge, C.B. & Bartel, D.P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92-105 (2009).
235. Lee, Y. et al. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051-60 (2004).
236. Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L. & Bradley, A. Identification of mammalian microRNA host genes and transcription units. *Genome Res* **14**, 1902-10 (2004).
237. Lee, Y. et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-9 (2003).
238. Han, J. et al. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**, 3016-27 (2004).
239. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95-8 (2004).
240. Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-6 (2001).
241. Doi, N. et al. Short-interfering-RNA-mediated gene silencing in mammalian cells requires Dicer and eIF2C translation initiation factors. *Curr Biol* **13**, 41-6 (2003).
242. Elbashir, S.M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* **15**, 188-200 (2001).
243. Kim, D.H., Saetrom, P., Snove, O., Jr. & Rossi, J.J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci U S A* **105**, 16230-5 (2008).

244. Tay, Y., Zhang, J., Thomson, A.M., Lim, B. & Rigoutsos, I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**, 1124-8 (2008).
245. Orom, U.A., Nielsen, F.C. & Lund, A.H. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* **30**, 460-71 (2008).
246. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. & Burge, C.B. Prediction of mammalian microRNA targets. *Cell* **115**, 787-98 (2003).
247. Guo, H., Ingolia, N.T., Weissman, J.S. & Bartel, D.P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835-40 (2010).
248. Baek, D. et al. The impact of microRNAs on protein output. *Nature* **455**, 64-71 (2008).
249. Yekta, S., Shih, I.H. & Bartel, D.P. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594-6 (2004).
250. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-33 (2009).
251. Ladewig, E., Okamura, K., Flynt, A.S., Westholm, J.O. & Lai, E.C. Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res* **22**, 1634-45 (2012).
252. Berezikov, E., Chung, W.J., Willis, J., Cuppen, E. & Lai, E.C. Mammalian mirtron genes. *Mol Cell* **28**, 328-36 (2007).
253. Xie, M. et al. Mammalian 5'-capped microRNA precursors that generate a single microRNA. *Cell* **155**, 1568-80 (2013).
254. Ahlenstiel, C.L. et al. Direct evidence of nuclear Argonaute distribution during transcriptional silencing links the actin cytoskeleton to nuclear RNAi machinery in human cells. *Nucleic Acids Res* **40**, 1579-95 (2012).
255. Ameyar-Zazoua, M. et al. Argonaute proteins couple chromatin silencing to alternative splicing. *Nat Struct Mol Biol* **19**, 998-1004 (2012).
256. Kim, D.H., Villeneuve, L.M., Morris, K.V. & Rossi, J.J. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol* **13**, 793-7 (2006).
257. Rudel, S., Flatley, A., Weinmann, L., Kremmer, E. & Meister, G. A multifunctional human Argonaute2-specific monoclonal antibody. *RNA* **14**, 1244-53 (2008).
258. Francia, S. et al. Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature* **488**, 231-5 (2012).
259. Elbashir, S.M. et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494-8 (2001).
260. Liu, Y. et al. Cloning of two candidate tumor suppressor genes within a 10 kb region on chromosome 13q14, frequently deleted in chronic lymphocytic leukemia. *Oncogene* **15**, 2463-73 (1997).
261. Cimmino, A. et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A* **102**, 13944-9 (2005).
262. Ota, A. et al. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. *Cancer Res* **64**, 3087-95 (2004).
263. Mendell, J.T. miRiad roles for the miR-17-92 cluster in development and disease. *Cell* **133**, 217-22 (2008).
264. He, L. et al. A microRNA polycistron as a potential human oncogene. *Nature* **435**, 828-33 (2005).
265. Steck, P.A. et al. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nat Genet* **15**, 356-62 (1997).

266. Li, D.M. & Sun, H. TEP1, encoded by a candidate tumor suppressor locus, is a novel protein tyrosine phosphatase regulated by transforming growth factor beta. *Cancer Res* **57**, 2124-9 (1997).
267. Ali, I.U., Schriml, L.M. & Dean, M. Mutational spectra of PTEN/MMAC1 gene: a tumor suppressor with lipid phosphatase activity. *J Natl Cancer Inst* **91**, 1922-32 (1999).
268. Saal, L.H. et al. Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair. *Nat Genet* **40**, 102-7 (2008).
269. Mirmohammadsadeh, A. et al. Epigenetic silencing of the PTEN gene in melanoma. *Cancer Res* **66**, 6546-52 (2006).
270. Garcia, J.M. et al. Promoter methylation of the PTEN gene is a common molecular change in breast cancer. *Genes Chromosomes Cancer* **41**, 117-24 (2004).
271. Alimonti, A. et al. Subtle variations in Pten dose determine cancer susceptibility. *Nature Genetics* **42**, 454-U136 (2010).
272. Chen, Z. et al. Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436**, 725-30 (2005).
273. Manning, B.D. & Cantley, L.C. AKT/PKB signaling: navigating downstream. *Cell* **129**, 1261-74 (2007).
274. Mayo, L.D. & Donner, D.B. A phosphatidylinositol 3-kinase/Akt pathway promotes translocation of Mdm2 from the cytoplasm to the nucleus. *Proc Natl Acad Sci U S A* **98**, 11598-603 (2001).
275. Zhou, B.P. et al. HER-2/neu induces p53 ubiquitination via Akt-mediated MDM2 phosphorylation. *Nat Cell Biol* **3**, 973-82 (2001).
276. Zhou, B.P. et al. Cytoplasmic localization of p21Cip1/WAF1 by Akt-induced phosphorylation in HER-2/neu-overexpressing cells. *Nat Cell Biol* **3**, 245-52 (2001).
277. Viglietto, G. et al. Cytoplasmic relocalization and inhibition of the cyclin-dependent kinase inhibitor p27(Kip1) by PKB/Akt-mediated phosphorylation in breast cancer. *Nat Med* **8**, 1136-44 (2002).
278. Shin, I. et al. PKB/Akt mediates cell-cycle progression by phosphorylation of p27(Kip1) at threonine 157 and modulation of its cellular localization. *Nat Med* **8**, 1145-52 (2002).
279. Liang, J. et al. PKB/Akt phosphorylates p27, impairs nuclear import of p27 and opposes p27-mediated G1 arrest. *Nat Med* **8**, 1153-60 (2002).
280. Machama, T. & Dixon, J.E. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J Biol Chem* **273**, 13375-8 (1998).
281. Stambolic, V. et al. Regulation of PTEN transcription by p53. *Mol Cell* **8**, 317-25 (2001).
282. Meng, F. et al. MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology* **133**, 647-58 (2007).
283. Davies, H. et al. Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-54 (2002).
284. Pollock, P.M. & Meltzer, P.S. A genome-based strategy uncovers frequent BRAF mutations in melanoma. *Cancer Cell* **2**, 5-7 (2002).
285. Pollock, P.M. et al. High frequency of BRAF mutations in nevi. *Nat Genet* **33**, 19-20 (2003).
286. Michaloglou, C. et al. BRAFE600-associated senescence-like cell cycle arrest of human naevi. *Nature* **436**, 720-4 (2005).
287. Dankort, D. et al. Braf(V600E) cooperates with Pten loss to induce metastatic melanoma. *Nat Genet* **41**, 544-52 (2009).
288. Bollag, G. et al. Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat Rev Drug Discov* **11**, 873-86 (2012).

289. Flaherty, K.T. et al. Inhibition of mutated, activated BRAF in metastatic melanoma. *N Engl J Med* **363**, 809-19 (2010).
290. Bollag, G. et al. Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **467**, 596-9 (2010).
291. Sosman, J.A. et al. Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *N Engl J Med* **366**, 707-14 (2012).
292. Paraiso, K.H. et al. PTEN loss confers BRAF inhibitor resistance to melanoma cells through the suppression of BIM expression. *Cancer Res* **71**, 2750-60 (2011).
293. Schubbert, S., Shannon, K. & Bollag, G. Hyperactive Ras in developmental disorders and cancer. *Nat Rev Cancer* **7**, 295-308 (2007).
294. Warne, P.H., Viciana, P.R. & Downward, J. Direct interaction of Ras and the amino-terminal region of Raf-1 in vitro. *Nature* **364**, 352-5 (1993).
295. Leever, S.J., Paterson, H.F. & Marshall, C.J. Requirement for Ras in Raf activation is overcome by targeting Raf to the plasma membrane. *Nature* **369**, 411-4 (1994).
296. Howe, L.R. et al. Activation of the MAP kinase pathway by the protein kinase raf. *Cell* **71**, 335-42 (1992).
297. Flockhart, R.J. et al. BRAFV600E remodels the melanocyte transcriptome and induces BANCR to regulate melanoma cell migration. *Genome Res* **22**, 1006-14 (2012).
298. Yu, H. & Jove, R. The STATs of cancer--new molecular targets come of age. *Nat Rev Cancer* **4**, 97-105 (2004).
299. Takeda, K. & Akira, S. STAT family of transcription factors in cytokine-mediated biological responses. *Cytokine Growth Factor Rev* **11**, 199-207 (2000).
300. Schindler, C., Levy, D.E. & Decker, T. JAK-STAT signaling: from interferons to cytokines. *J Biol Chem* **282**, 20059-63 (2007).
301. Heinrich, P.C. et al. Principles of interleukin (IL)-6-type cytokine signalling and its regulation. *Biochem J* **374**, 1-20 (2003).
302. Darnell, J.E., Jr., Kerr, I.M. & Stark, G.R. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* **264**, 1415-21 (1994).
303. Darnell, J.E., Jr. STATs and gene regulation. *Science* **277**, 1630-5 (1997).
304. Bollrath, J. et al. gp130-mediated Stat3 activation in enterocytes regulates cell survival and cell-cycle progression during colitis-associated tumorigenesis. *Cancer Cell* **15**, 91-102 (2009).
305. Bromberg, J.F. et al. Stat3 as an oncogene. *Cell* **98**, 295-303 (1999).
306. Bromberg, J. Stat proteins and oncogenesis. *J Clin Invest* **109**, 1139-42 (2002).
307. Zhang, Q. et al. STAT3- and DNA methyltransferase 1-mediated epigenetic silencing of SHP-1 tyrosine phosphatase tumor suppressor gene in malignant T lymphocytes. *Proc Natl Acad Sci U S A* **102**, 6948-53 (2005).
308. Oka, T. et al. Gene silencing of the tyrosine phosphatase SHP1 gene by aberrant methylation in leukemias/lymphomas. *Cancer Res* **62**, 6390-4 (2002).
309. Chim, C.S., Fung, T.K., Cheung, W.C., Liang, R. & Kwong, Y.L. SOCS1 and SHP1 hypermethylation in multiple myeloma: implications for epigenetic activation of the Jak/STAT pathway. *Blood* **103**, 4630-5 (2004).
310. Lee, H. et al. Acetylated STAT3 is crucial for methylation of tumor-suppressor gene promoters and inhibition by resveratrol results in demethylation. *Proc Natl Acad Sci U S A* **109**, 7765-9 (2012).
311. Iliopoulos, D., Jaeger, S.A., Hirsch, H.A., Bulyk, M.L. & Struhl, K. STAT3 activation of miR-21 and miR-181b-1 via PTEN and CYLD are part of the epigenetic switch linking inflammation to cancer. *Mol Cell* **39**, 493-506 (2010).

312. Loffler, D. et al. Interleukin-6 dependent survival of multiple myeloma cells involves the Stat3-mediated induction of microRNA-21 through a highly conserved enhancer. *Blood* **110**, 1330-3 (2007).
313. Brock, M. et al. Interleukin-6 modulates the expression of the bone morphogenic protein receptor type II through a novel STAT3-microRNA cluster 17/92 pathway. *Circ Res* **104**, 1184-91 (2009).
314. Diosdado, B. et al. MiR-17-92 cluster is associated with 13q gain and c-myc expression during colorectal adenoma to adenocarcinoma progression. *Br J Cancer* **101**, 707-14 (2009).
315. Carraro, G. et al. miR-17 family of microRNAs controls FGF10-mediated embryonic lung epithelial branching morphogenesis through MAPK14 and STAT3 regulation of E-Cadherin distribution. *Dev Biol* **333**, 238-50 (2009).
316. Wang, Y. et al. Lethal-7 is down-regulated by the hepatitis B virus x protein and targets signal transducer and activator of transcription 3. *J Hepatol* **53**, 57-66 (2010).
317. Wong, H.H. & Wang, J. Epstein-Barr virus positive diffuse large B-cell lymphoma of the elderly. *Leuk Lymphoma* **50**, 335-40 (2009).
318. Morales, D. et al. Epstein-Barr virus as a prognostic factor in de novo nodal diffuse large B-cell lymphoma. *Leuk Lymphoma* **51**, 66-72 (2010).
319. Mele, A. et al. Hepatitis C virus and B-cell non-Hodgkin lymphomas: an Italian multicenter case-control study. *Blood* **102**, 996-9 (2003).
320. Fisher, S.G. & Fisher, R.I. The epidemiology of non-Hodgkin's lymphoma. *Oncogene* **23**, 6524-34 (2004).
321. A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. *Blood* **89**, 3909-18 (1997).
322. Alizadeh, A.A. et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-11 (2000).
323. Wright, G. et al. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci U S A* **100**, 9991-6 (2003).
324. Rosenwald, A. et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* **346**, 1937-47 (2002).
325. Naresh, K.N. et al. Report of an International Network of Cancer Treatment and Research workshop on non-Hodgkin's lymphoma in developing countries. *Blood Cells Mol Dis* **33**, 330-7 (2004).
326. Hogfeldt, T. et al. Patients with activated B-cell like diffuse large B-cell lymphoma in high and low infectious disease areas have different inflammatory gene signatures. *Leuk Lymphoma* **54**, 996-1003 (2013).
327. Liu, Y. et al. S1PR1 is an effective target to block STAT3 signaling in activated B cell-like diffuse large B-cell lymphoma. *Blood* **120**, 1458-65 (2012).
328. Lam, L.T. et al. Cooperative signaling through the signal transducer and activator of transcription 3 and nuclear factor- κ B pathways in subtypes of diffuse large B-cell lymphoma. *Blood* **111**, 3701-13 (2008).
329. Sado, T., Hoki, Y. & Sasaki, H. Tsix defective in splicing is competent to establish Xist silencing. *Development* **133**, 4925-31 (2006).
330. Yue, X. et al. Transcriptional regulation by small RNAs at sequences downstream from 3' gene termini. *Nat Chem Biol* **6**, 621-9 (2010).
331. Nishioka, C., Ikezoe, T., Yang, J., Udaka, K. & Yokoyama, A. Imatinib causes epigenetic alterations of PTEN gene via upregulation of DNA methyltransferases and polycomb group proteins. *Blood Cancer J* **1**, e48 (2011).

332. Saayman, S. et al. An HIV-encoded antisense long noncoding RNA epigenetically regulates viral transcription. *Mol Ther* **22**, 1164-75 (2014).
333. Salmena, L., Carracedo, A. & Pandolfi, P.P. Tenets of PTEN tumor suppression. *Cell* **133**, 403-14 (2008).
334. Deng, T. et al. An essential role for DNA methyltransferase 3a in melanoma tumorigenesis. *Biochem Biophys Res Commun* **387**, 611-6 (2009).
335. Zhang, Y. et al. DNMT3a plays a role in switches between doxorubicin-induced senescence and apoptosis of colorectal cancer cells. *Int J Cancer* **128**, 551-61 (2011).
336. Verma, S.K. et al. Identification of Potent, Selective, Cell-Active Inhibitors of the Histone Lysine Methyltransferase EZH2. *ACS Med Chem Lett* **3**, 1091-6 (2012).
337. Murthy, K.G. & Manley, J.L. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev* **9**, 2672-83 (1995).
338. Smith, A.L. et al. The miR-106b-25 cluster targets Smad7, activates TGF-beta signaling, and induces EMT and tumor initiating cell characteristics downstream of Six1 in human breast cancer. *Oncogene* **31**, 5162-71 (2012).
339. Yue, P. & Turkson, J. Targeting STAT3 in cancer: how successful are we? *Expert Opin Investig Drugs* **18**, 45-56 (2009).