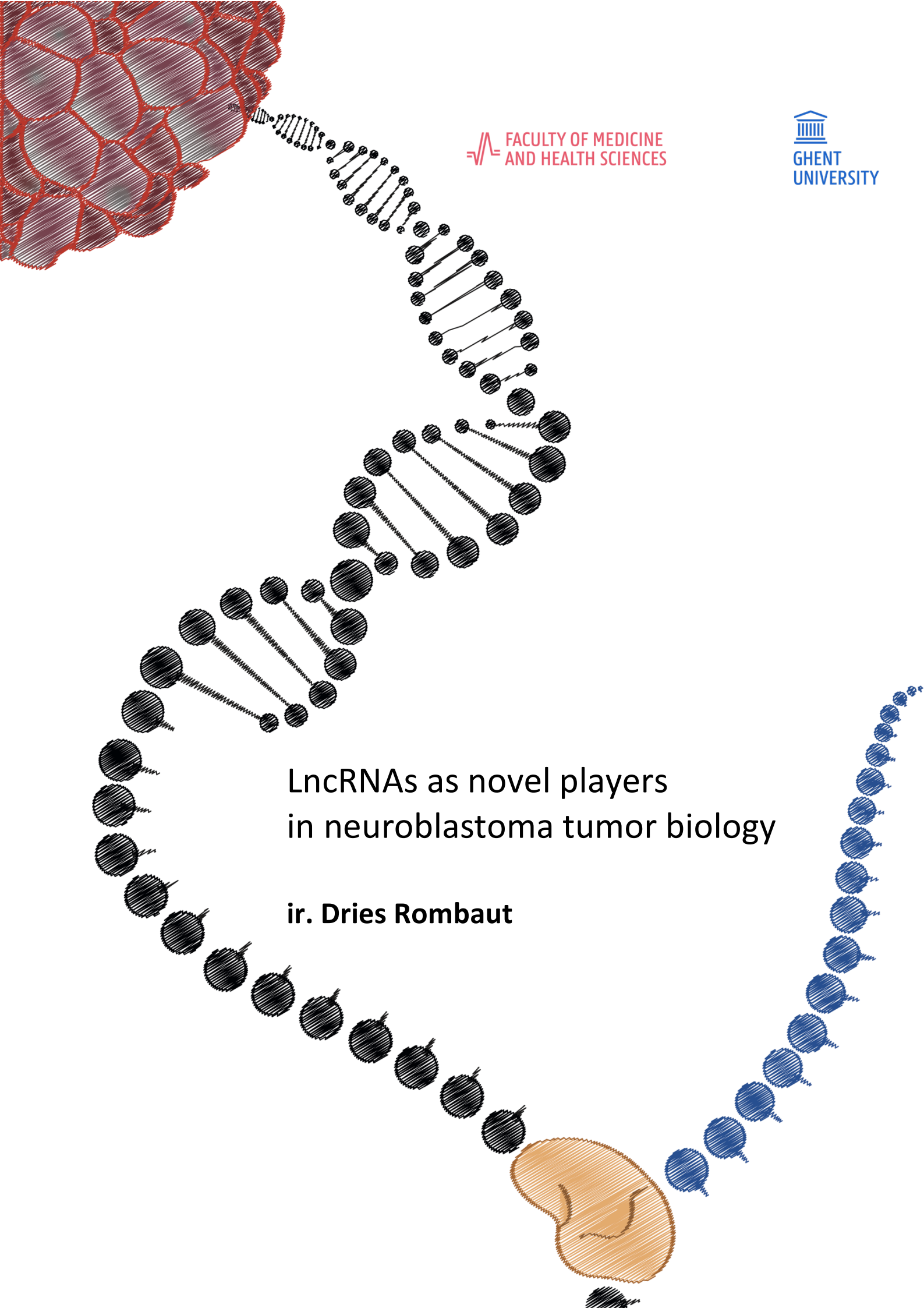


FACULTY OF MEDICINE AND HEALTH SCIENCES

GHENT UNIVERSITY

LncRNAs as novel players in neuroblastoma tumor biology

ir. Dries Rombaut

FACULTY OF MEDICINE AND HEALTH SCIENCES

Ghent University, Faculty of Medicine and Health Sciences

LncRNAs as novel players in neuroblastoma tumor biology

Thesis submitted to fulfill the requirements for the degree of Doctor of Health Sciences
2019

Dries Rombaut

Promoter: Prof. dr. ir. Pieter Mestdagh
Co-promoter: dr. Steve Lefever

Center for Medical Genetics
Cancer Research Institute Ghent
Ghent University Hospital, Medical Research Building 1
Corneel Heymanslaan 10, 9000, Ghent, Belgium
+32-9-3321951
Dries.Rombaut@UGent.be

GHENT UNIVERSITY

Thesis submitted to fulfill the requirements for the degree of Doctor of Health Sciences

Promoter

Prof. dr. ir. Pieter Mestdagh

Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

Co-promoter

dr. Steve Lefever

Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

Members of the examination committee

Prof. dr. Jolanda van Hengel (chairman)

Department of Human Structure and Repair, Ghent University, Ghent, Belgium

dr. Kaat Durinck

Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

dr. Bram De Wilde, M.D., PhD

Department of Pediatric Hematology and Oncology, Ghent University Hospital, Ghent, Belgium

dr. Filip Matthijssens

Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

Prof. dr. Katleen De Preter

Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

Prof. dr. Claude Libert

Department of biomedical molecular biology, Ghent University, Ghent, Belgium

Center for inflammation Research, VIB/Ghent University, Ghent, Belgium

Prof. dr. Wim Vanden Berghe

Departement of Biomedical Sciences, Antwerp University, Antwerp, Belgium

The research described in this thesis was conducted at the Center for Medical Genetics, Ghent University, Ghent, Belgium. This work and Dries Rombaut was supported by the FWO and the Emmanuel Van der Schueren research grant (Kom Op Tegen Kanker).

Contents

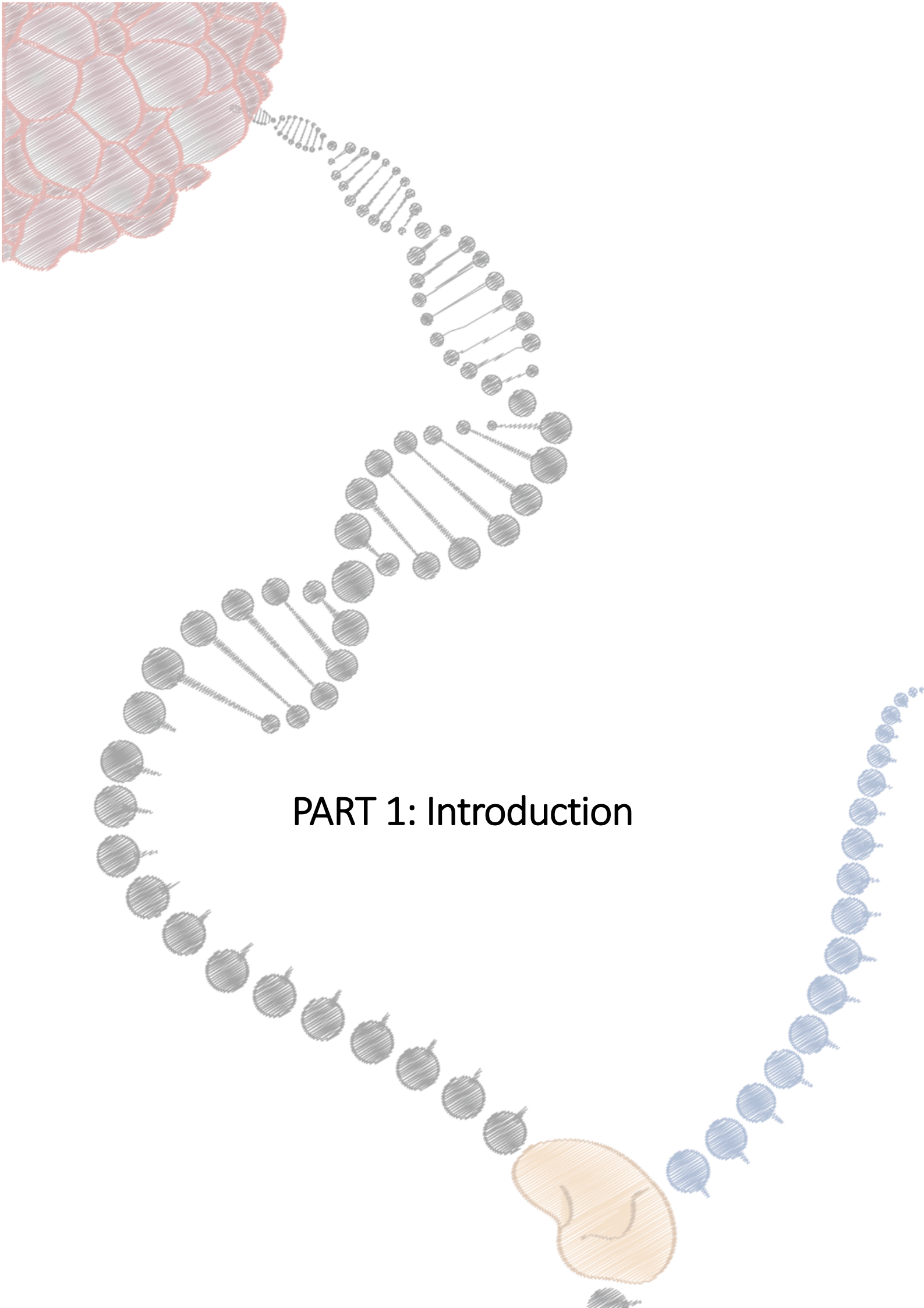
List of abbreviations

2’OMe	2’-O-Methyl
2’MOE	2’-O-Methoxyethyl
4C-seq	Chromatin conformation capture-on-chip sequencing
Ago2	Argonaute 2
ALK	Anaplastic Lymphoma Kinase Receptor Tyrosine Kinase
ALT	Alternative Lengthening of Telomeres
AML	Acute Myeloid Leukemia
Amp	Amplified
AR	Androgen Receptor
ASO	Antisense Oligonucleotide
BALL	B-Cell Acute Lymphoblastic Leukemia
BLCA	Bladder Carcinoma
BNA	Bridged Nucleic Acid
BRCA	Breast Carcinoma
BrU	Bromouridine
CAGE-seq	Cap Analysis Gene Expression sequencing
CCLE	Cancer Cell Line Encyclopedia
ChIP-seq	Chromatin Immunoprecipitation-sequencing
ChIRP-seq	Chromatin Isolation by RNA Purification-sequencing
CLIP	Crosslinking and Immunoprecipitation
CML	Chronic Myeloid Leukemia
CMV	Cytomegalovirus
CNA	Copy Number Alteration
CNV	Copy Number Variation
CRC	Core Regulatory Circuit
CRCA	Colon Adenocarcinoma
CRISPRi	Clustered Regularly Interspaced Short Palindromic Repeats Interference
dCas9	dead Cas9
dCor	distance Correlation
DNMT	DNA methyltransferase
dsRNA	double stranded RNA
ecCEBPA	extra-coding CEBPA
eCLIP	enhanced Crosslinking and Immunoprecipitation
EMT	Epithelial-to-Mesenchymal Transition
ENCODE	Encyclopedia of DNA Elements
ER	Estrogen Receptor
eRNA	enhancer RNA
ESCC	Esophageal Squamous Cell Carcinoma

FANTOM	Functional Annotation of the Mammalian genome
FBS	Fetal Bovine Serum
FDA	Food and Drug Administration
GBM	Glioblastoma
GCA	Gastric Carcinoma
GM-CSF	Granulocyte macrophage colony-stimulating factor
GN	Ganglioneuroma
GNB	Ganglioneuroblastoma
GRO-seq	Global Nuclear Run-On sequencing
GSEA	Gene Set Enrichment Analysis
HCC	Hepatocellular Carcinoma
HL	Hodgkin Lymphoma
HOTTIP	HOXA transcript at the distal tip
HSNCC	Head and Neck Squamous Cell Carcinoma
IDRF	Image Defined Risk Factor
INRG	International Neuroblastoma Risk Group
INRGSS	International Neuroblastoma Risk Group Staging System
INSS	International Neuroblastoma Staging System
JS	Jensen-Shannon
KO	knock-out
LGG	Brain Lower Grade Glioma
LINES	Low and Intermediate risk Neuroblastoma European Study
LNA	Locked Nucleic Acid
lncRNA	long non-coding RNA
lincRNA	long intergenic non-coding RNA
logFC	log Fold Change
mRNA	messenger RNA
miRNA	microRNA
MB	Medulloblastoma
MEL	Melanoma
MESO	Mesothelioma
MM	Plasma cell myeloma
MSigDB	Molecular Signatures Database
MVB	Multivesicular Body
MYCNa	MYCN amplified
MYCNsc	MYCN single copy
NA	Non-amplified
NAT	Natural Antisense Transcript
NB	Neuroblastoma
NCA	Numerical copy alteration
NCC	Neural Crest Cell

NESPR	NEuroblastoma Specific PHOX2B Regulatory RNA
NHL	Non-Hodgkin Lymphoma
NOR	Noradrenergic
ncRNA	non-coding RNA
PACA	Pancreatic Adenocarcinoma
PAM	Protospacer Adjacent Motif
PBS	Phosphate-Buffered Saline
PCAT-1	Prostate Cancer Associated Transcript 1
pcRNA	positionally conserved lncRNAs
piRNA	piwi interacting RNA
PR	Progesterone Receptor
PRC2	Polycomb Repressive Complex 2
PRCA	Prostate carcinoma
PS	Phosphorothioate
PSA	Prostate-specific antigen
PWM	Position Weight Matrix
RB	Retinoblastoma
RBP	RNA-binding proteins
RCC	Renal Cell Carcinoma
RIN	RNA integrity number
RIP	RNA Immunoprecipitation
RISC	RNA-induced silencing complex
RNA	Ribonucleic Acid
RNA-seq	RNA-sequencing
RNP	Ribonucleoprotein
RT	Radiation therapy
RT-qPCR	quantative Polymerase Chain Reaction with Reverse Transcription
SCLC	Small Cell Lung Carcinoma
SCP	Schwann Cell Precursor
sgRNA	single guide RNA
shControl	short hairpin Control
shDNA	short hairpin DNA
shRNA	short hairpin RNA
SIOPEN	European SIOP Neuroblastoma group
siPOOL	siRNA Pool
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
SNP	Single Nucleotide Polymorphisms
SRA	Steroid Receptor RNA activator
TAD	Topologically Associated Domain
TALL	T-cell Acute Lymphoblastic Leukemia

tapRNA	Topological Anchor Point RNA
TCGA	The Cancer Genome Atlas
TERC	Telomerase RNA Component
TF	Transcription Factor
THCA	Thyroid gland carcinoma
TPM	Transcripts Per Million
TSP-1	Thrombospondin-1
TSS	Transcription Start Site
VEGF-A	Vascular Endothelial Growth Factor-A
Xi	Inactivated X chromosome
XIST	X-inactive-specific transcript



PART 1: Introduction

1.1 Cancer

1.1.1 Cancer incidence and mortality

In this day and age, nearly everyone knows someone who was diagnosed with some form of cancer. This assumption is substantiated by the fact that in Belgium approximately one in three males and one in four females will develop cancer before the age of 75[1]. Globally, the prevalence of cancer has more than doubled since 1990, reaching a total of 42 million affected people in 2016. This development can partly be ascribed to the growing and aging world population and increased and improved screenings[2]. Furthermore, risk factors increase the incidence for one third of the cancer types. These risk factors include smoking, alcohol, UV exposure, HPV infection and chronic diseases such as obesity and Type 2 diabetes[3]. In 2018, 18.1 million people were diagnosed with cancer worldwide, resulting in 9.6 million cancer related deaths. In 91 of 172 investigated countries, cancer is the first or second leading cause of death before the age of 70[4]. In Belgium, the number of cancer diagnoses is predicted to rise from 67,820 in 2014 to 79,135 in 2025[5].

1.1.2 Cancer is a genetic disease

Cancer is a genetic disease, caused by alterations in the genome that affect the expression or activity of tumor suppressor genes or proto-oncogenes. Both gene classes play an important role in cell growth and proliferation[6]. Tumor suppressors normally inhibit cancer initiation or progression, whereas oncogenes possess the ability to initiate these processes. For a normal cell to develop into a cancerous one, multiple genes typically need to be affected, either via mutations, copy number variations (CNV) or hyper- or hypomethylation.

Tumor suppressor genes commonly encode proteins inhibiting cell proliferation[6]. This inhibition is accomplished by regulation of cell cycle progression, checkpoint control leading to cell cycle arrest, activation of the apoptotic circuitry, production of hormone receptors linked with cell proliferation, and DNA repair. Promotion of tumor formation requires inactivation or loss of both alleles of a tumor suppressor gene, since in most cases a single copy of a tumor suppressor gene is sufficient to control cell proliferation[6,7]. This inactivation

can arise through hypermethylation in the promoter region or mutations throughout the entire length of the tumor suppressor gene in a protein-altering manner[7]. For instance, a mutation in the retinoblastoma gene, a regulator of cell cycle progression and the first discovered tumor suppressor gene to harbor a mutation, can initiate tumorigenesis after acquiring additional mutations[6,8–10]. Deletions of the tumor suppressor locus or chromosomal arms containing tumor suppressor genes also lead to a loss of tumor suppressive activity. The first reported common CNV (deletion) in a protein coding gene with a familial breast cancer risk association was a locus containing *MTUS1*, implicated in regulation of apoptosis and proliferation[11]**.**

In contrast, oncogenes are mostly involved in cell growth and differentiation and typically get activated by dominant mutations, primarily in their functional domains. Among the first discovered oncogenic mutations were *RAS* proteins, members of signaling transduction pathways, migration, adhesion and differentiation[6,12]. Alterations in their genetic code results in constitutively active proteins, promoting uncontrolled growth signals leading to tumor formation[6,12]. Estimations of an activating mutation in one of the *RAS* genes go up to as high as 20% of all tumors[12]. Other activating alterations are hypomethylation of the oncogene promoter, as well as amplifications of proto-oncogenes or copy number gains of chromosomal regions containing genes with an oncogenic potential, all leading to a higher expression of these genes. For example, hypomethylation of *PLAU* results in overexpression, which in turn leads to tumor progression in breast and prostate cancers[13]. *MYC,* one of the best characterized oncogenes, regulates important aspects of transformation and is found to be amplified in a variety of tumor cells[14–16]. Gains of larger chromosomal regions also lead to an excessive activity of oncogenes, as a recent study[17] identified eight oncogenes located in copy number gains in at least 250 tumor samples, among which *INTS8* and *DDHD2*. Structural rearrangements such as fusion proteins and enhancer hijacking can also mediate oncogene activation. For these chromosomal translocations to occur, two critical hurdles need to be overcome. First, DNA double strand breaks have to be present in two genomic loci at the same time. Second, an erroneous joining of the double strand breaks leads to the fusion of two chromosomal parts that should not unite. This illegitimate joining can lead to fusion genes, which arise from multiple, previously separate genes that will be translated to fusion proteins, with a functionality derived from the genes that make up the fusion protein[18]. Another way of

activating proto-oncogenes through chromosomal translocation is by bringing enhancer elements into their vicinity, called enhancer hijacking[18,19]. Through activation by these cis-regulatory elements, the neighboring genes will be overexpressed[18].

1.1.3 Hallmarks of cancer

Alterations in oncogenes and tumor suppressor genes can transform a normal cell into a malignant cell and ultimately a full-blown tumor. These alterations help circumvent the inherent cellular defense mechanisms against malignant transformation, and provide the cancer cell with a number of hallmarks that are essential for tumor progression[20]. A concise overview of these hallmarks is provided below (figure 1).

1.1.3.1 Sustaining proliferative signaling

The most critical ability a cancer cell has to obtain, is sustained proliferation[20,21]. Cell cycle progression and the production and release of growth-promoting signals are tightly controlled in normal cells, but commonly deregulated in cancer cells. Cancer cells can obtain the capacity to sustain proliferation in multiple ways. They can produce their own growth ligands, send signals to neighboring cells which in turn reciprocate with growth factors, structurally alter growth receptor molecules leading to ligand independent activation or deregulate growth receptor signaling through increased receptor expression on their cellular surface. Moreover, constitutive activation of members of the signaling pathways downstream of growth factors may also result in an independence of exogenous activation[20,21].

1.1.3.2 Evading growth suppressors

Complementary to the first hallmark, bypassing growth suppression is integral to cancer cell formation[20,21]. Negative regulation of proliferation is largely dependent on the activity of tumor suppressor genes. The two most studied tumor suppressors are the RB (retinoblastoma) and TP53 proteins. These key regulators of cell proliferation circuits control both senescence and apoptotic processes. Defects in the RB pathway will leave the cell unresponsive to extracellular and intracellular signals, enabling uncontrolled cell proliferation. In contrast to RB, TP53 reacts on intracellular stress and DNA damage signals. Inactivation of this pathway inhibits the apoptotic programs, permitting continued cell division. A second

method to circumvent growth suppression is by evading contact inhibition[20,21]. Contact inhibition ensures the proper formation of tissue layers and is abolished during oncogenesis.

1.1.3.3 Resisting cell death

The third hallmark relates to the capability of cancer cells to resist the apoptotic programs[20,21]. Cancer cells experience various types of physiologic stress, triggering apoptosis in response to those signals. This stress ranges from increased abundance of oncogene signaling, over hypoxia to DNA damage induced by uncontrolled proliferation. Apoptosis is activated by an imbalance between pro- and anti-apoptotic members of the Bcl-2 family. To limit the induction of the apoptotic circuitry, cancer cells often lose TP53 tumor suppressor function, by the occurrence of inactivating *TP53* mutations or TP53 inhibition by *MDM2* overexpression. Other mechanisms include autophagy, overexpression of anti-apoptotic regulators or survival signals or downregulation of pro-apoptotic factors, creating a permanent imbalance and blockage of apoptosis[20,21].

1.1.3.4 Enabling replicative immortality

Most normal cell lineages in the human body divide a limited number of times successively[20,21]. To develop a cancerous mass, cells need to replicate continuously. To achieve this, two obstacles must be tackled, namely senescence and the crisis phase/apoptosis. Circumvention of the apoptotic pathway was already discussed above. Telomerase, a DNA polymerase adding telomere repeat sequences to the end of telomeric DNA, is a crucial gene to inhibit senescence and apotosis[20,21]. Telomere lengthening to counteract their degradation is viewed as the driving factor in cell immortalization and determines the replicative capacity. Around 90% of immortalized cells express telomerase, while it is almost undetectable in normal tissues. Alternative lengthening of telomeres (ALT) is a telomerase independent mechanism to circumvent progressive telomere erosion, potentially based on homology-directed telomere synthesis[22].

1.1.3.5 Inducing angiogenesis

Due to the increased cellular turnover associated with tumor formation and growth, cancer cells require large amounts of nutrients and oxygen, as well as a way to discard metabolic waste and carbon dioxide[20,21]. The development of a vasculature network is essential in this

regard. In contrast to normal tissues, where the vasculature becomes quiescent with only brief, transient activations of angiogenesis in case of healing, the angiogenic program remains active during tumor growth. Well-known examples of angiogenesis inducers and inhibitors are the vascular endothelial growth factor-A (VEGF-A) and thrombospondin-1 (TSP-1), respectively. Oncogene signaling can upregulate VEGF gene expression, resulting in higher ligand levels binding to stimulatory receptors on vascular endothelial cells, bringing about new blood vessel growth. Chronically activated angiogenesis leads to aberrant blood vessels, characterized by excessive vessel branching, enlarged and deformed vessels, erratic blood flow and leakiness[20,21].

1.1.3.6 Activating invasion and metastasis

Metastasis is a series of discrete steps, bundled in the invasion-metastasis cascade. It starts with local invasion of cancer cells in the surrounding tissue, followed by intravasation of malignant cells into blood and lymphatic vessels through which they are transported to distant parts of the body[20,21]. Eventually, the cells will escape the blood and lymphatic systems, form micrometastases and ultimately grow into macroscopic tumors. Cells first develop phenotypic alterations, while detaching from other cells and the extracellular matrix. Loss of E-cadherin, a cell-to-cell adhesion molecule, is one of the best characterized alterations leading to metastasis. Epithelial cells normally express E-cadherin, which forms adherent junctions and maintains their composure[23]. However, in human carcinoma cells, downregulation or mutational inactivation is often observed. Genes with similar functions, such as N-cadherin, can also harbor alterations in some highly aggressive carcinomas. Another developmental regulatory program, the epithelial-mesenchymal transition (EMT), has been suggested as a driver of metastasis. Here, transformed epithelial cells with the ability to invade trigger the process, resist programmed cell death and propagate. EMT-inducing transcription factors appear to be able to orchestrate most steps of the invasion-metastasis cascade[20,21].

1.1.3.7 Enabling characteristics and emerging hallmarks

The previously described hallmarks that allow cancer cells to grow, proliferate and metastasize are made possible by two enabling characteristics, the development of genomic instability and tumor-promoting inflammation[20,21]. The first characteristic alludes to the mutations and chromosomal rearrangements that occur in the genome of a precursor cancer

cell, by chance leading to acquirement of hallmark capabilities. The latter hints at the involvement of the immune system in the development of malignant lesions, which can promote tumor progression.

Two other attributes functionally important for cancer development involve reprogramming of the cellular energy metabolism – to support the continuous cell growth and proliferation – and active evasion of elimination by immune cells. These four features allow and enable cancer cells to form and produce tumor masses, in concert with the already established hallmarks[20,21].

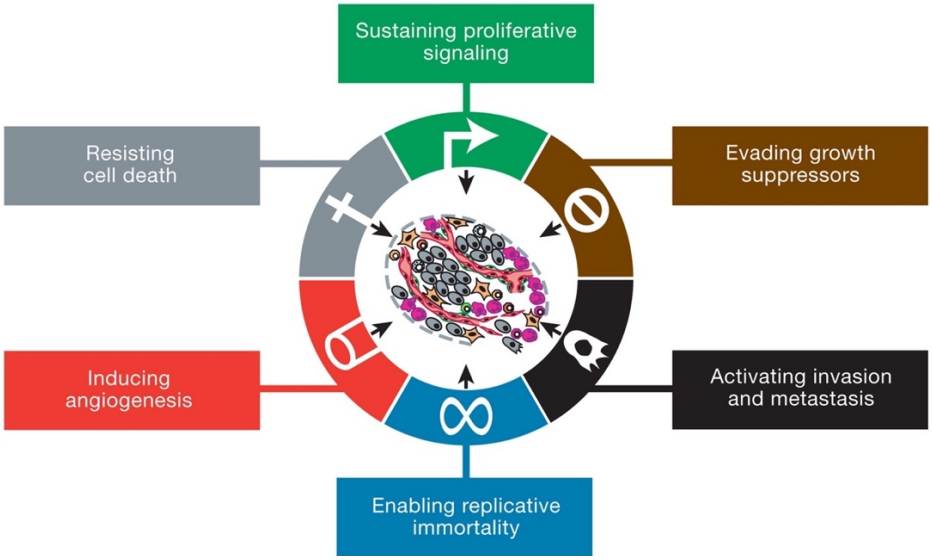


Figure 1: The hallmarks of cancer. These are the six original hallmarks suggested by Hanahan and Weinberg in 2000. Each hallmark is a capability that needs to be acquired to develop tumors. (Source: Hallmarks of Cancer: The Next Generation, Hanahan and Weinberg, 2011[21])

1.2 Neuroblastoma

Neuroblastoma is the most common extracranial tumor in children, diagnosed in around 10.2 per million children younger than 15 years old, accounting for 7-10% of all pediatric cancers[24,25]. Although it is a rather rare malignancy, neuroblastoma causes 15% of all pediatric cancer deaths[24]. Despite advances in understanding the underlying biology, and intense multimodal therapy, the survival rate for high-risk patients is only around 50%.

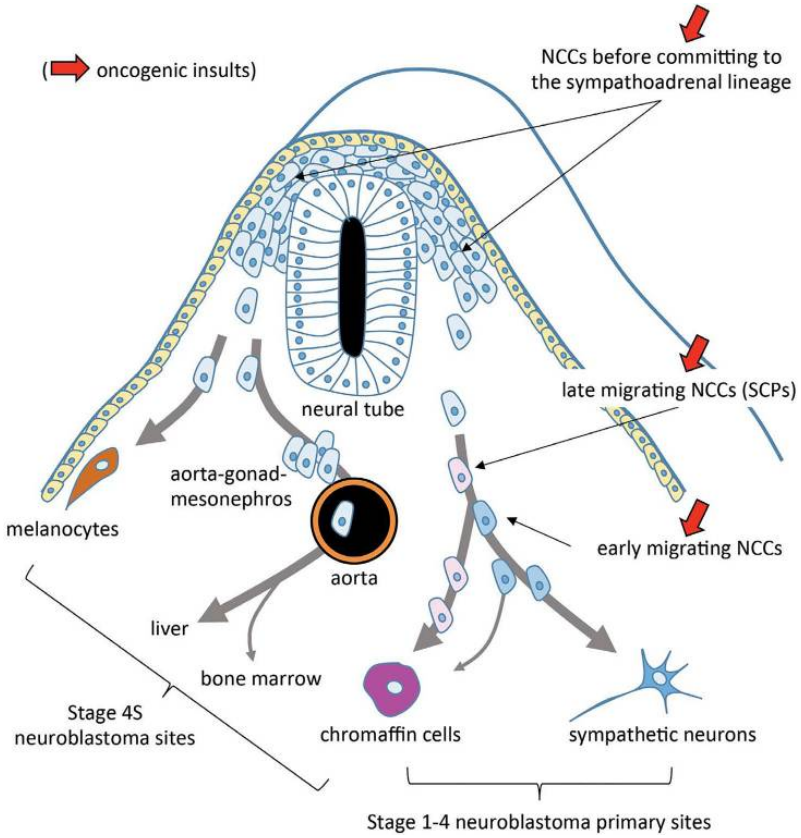


Figure 2: Origin of neuroblastoma. Neuroblastoma arises due to mistakes in the normal embryonal development. During the third and fourth week of normal embryonal development the neural tube closes, neural crest cells (NCCs) emanate, migrate and generate various tissues. These NCCs differentiate into neurons and glia, giving rise to the sympathetic ganglia, paraspinal ganglia and adrenal medulla. Late-migrating NCCs can also differentiate into Schwann cell precursors (SCPs), which will make up the greater part of the adrenal medulla. SCPs can differentiate further into chromaffin cells, providing the majority of chromaffin cells in the adrenal medulla. (source: Origin and mechanism of neuroblastoma[26])

As neuroblastoma tumors arise as a consequence of genetic abnormalities in precursor cells of the sympathoadrenal lineage, they can develop anywhere in the sympathetic nervous

system[27]. However, the exact origin of neuroblastoma remains uncertain. During the third and fourth week of normal embryonal development, round the time the neural tube closes, neural crest cells (NCCs) emanate, migrate and generate various tissues. Through differentiation initiated by multiple transcription factors, they transform into neurons and glia, giving rise to the sympathetic ganglia, paraspinal ganglia and adrenal medulla[26–29]. Recently, Furlan et al. published new findings pertaining neuroblastoma initiation[30]. They showed that peripheral glial stem cells, called Schwann cell precursors (SCPs), make up the greater part of the adrenal medulla. These SCPs differentiate from late-migrating NCCs and can differentiate into chromaffin cells, providing the majority of chromaffin cells in the adrenal medulla. The divergence of the sympathetic and the adrenal lineage as proposed by Furlan et al. provides extra pieces of the puzzle of neuroblastoma onset (Figure 2).

Because neuroblastoma tumors develop alongside the sympathetic nervous system, they can occur from the neck down to the pelvis, with at least half of all tumors populating the abdomen[31]. Most of these abdominal masses are located on the adrenal gland, of which the adrenal medulla is a part. The primary location of the tumor, the degree of metastasis and potential lymph node involvement are used to classify neuroblastoma, conform the International Neuroblastoma Staging System (INSS)[32,33]. In stage 1 patients, the tumor has not metastasized and complete resection of the mass is possible. In case complete removal is not attainable, the tumor will be classified as stage 2A or stage 2B, depending on lymph node involvement. Stage 3 tumors are inoperable unilateral tumors, metastasizing to nearby lymph nodes or spreading to the other side of the body. Stage 4 comprises tumors that have metastasized to distant locations, including remote lymph nodes, the skeleton, bone marrow, the liver and other organs. Stage 4S is a special neuroblastoma stage with metastases in liver, skin and bone marrow, but patients show high regression rates and complete remissions. Another important variable to predict outcome is the patient's age at diagnosis. Patients older than 18 months have a worse prognosis than younger children[34].

Although some patients around the world are still classified based on INSS stages, a considerable disadvantage of this system is the reliance on the extent of surgical tumor resection and metastasis to the lymph nodes to stratify patients into the different stages. These criteria can vary between medical centers, as they depend on the competence of the

surgeon and the treating physician. To cater to this drawback, the International Neuroblastoma Risk Group (INRG) classification system was introduced in 2009, containing the International Neuroblastoma Risk Group Staging System (INRGSS), to allow the diagnosis of patients in a uniform matter, regardless of the patient's origin or treatment center[35] (Table 1). This staging system is based on radiological imaging, considering several risk factors associated with an increased risk of complications during surgery (Image Defined Risk Factors or IDRF). The INRGSS stratifies patients into four different categories: L1, L2, M and MS (Table 2). These classifications can be translated into the INSS, where L1 is equivalent with INSS stage 1, L2 with stage 2 and 3, M with stage 4 and MS equals INSS stage 4S. Using IDRFs and the spread of tumors at diagnosis as a method of staging, before the patient receives any form of treatment, allows for a more uniform (pre-treatment) risk classification of patients and the ability of hospitals to share and compare data[35,36].

Table 1: The INRG classification system for neuroblastoma patient risk stratification. (Amp: amplified; GN: ganglioneuroma; GNB: ganglioneuroblastoma; NA: non-amplified)

INRG Stage	Age (months)	Histologic Category	Grade of Tumor Differentiation	*MYCN*	11q Aberration	Ploidy	Pretreatment Risk Group
L1/L2		GN maturing; GNB intermixed					A Very low
L1		Any, except GN maturing or GNB intermixed		NA			B Very low
				Amp			K High
L2	< 18	Any, except GN maturing or GNB intermixed		NA	No		D Low
					Yes		G Intermediate
	≥ 18	GNB nodular; neuroblastoma	Differentiating	NA	No		E Low
					Yes		H Intermediate
			Poorly differentiated or undifferentiated	NA			
				Amp			N High
M	< 18			NA		Hyperdiploid	F Low
	< 12			NA		Diploid	I Intermediate
	12 to < 18			NA		Diploid	J Intermediate
	< 18			Amp			O High
	≥ 18						P High
MS				NA	No		C Very low
	< 18				Yes		Q High
				Amp			R High

Table 2: International Neuroblastoma Risk Group Staging System

Stage	Description
L1	Localized tumor not involving vital structures as defined by the list of image-defined risk factors and confined to one body compartment
L2	Locoregional tumor with presence of one or more image-defined risk factors
M	Distant metastatic disease (except stage MS)
MS	Metastatic disease in children younger than 18 months with metastases confined to skin, liver, and/or bone marrow

The primary tumor location, age of the patient, disease stage, genomic alterations and other parameters allow clinicians to segregate patients into three different groups: low-, intermediate- and high-risk patients. Currently there are 2 different treatment protocols used by the European SIOP Neuroblastoma group (SIOPEN) based on these risk groups. The low- and intermediate risk patients are grouped and treated according to the Low and Intermediate risk Neuroblastoma European Study (LINES) protocol, whereas the treatment of high-risk neuroblastoma patients follows the HR-NBL protocol[36]. The LINES protocol allows for further subdivisions into more than 16 different categories, based on stages, genetic alterations and life threatening symptoms among others, with every group receiving different treatments. Patients categorized as stage MS only require minimal intervention, whereas low- and intermediate risk patients need surgery to resect the tumor, chemotherapy, and in rare cases radiotherapy[37,38].

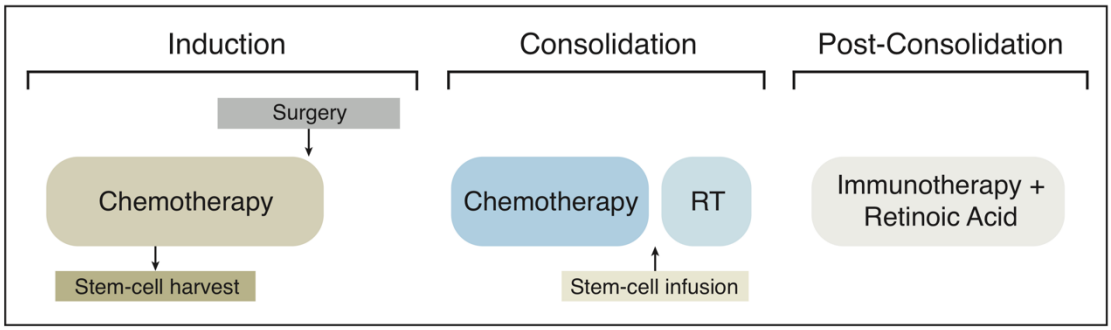


Figure 3: Treatment protocol for high-risk neuroblastoma patients. Treatment consists of three sections. In the induction phase, the patient is treated with chemotherapy to shrink the tumor, commonly followed by removal of the mass through surgical intervention. Harvesting of stem cells is also included in this phase. After the induction block, a consolidation phase ensures the elimination of remaining tumor cells through an intensive regiment of chemotherapy with (autologous) stem cell rescue, followed by radiation (RT). The post-consolidation phase focuses on minimizing the chance of relapse through immunotherapy and administration of retinoic acid. (Adapted from: Advances in risk classification and treatment strategies for neuroblastoma[36])

However, high-risk neuroblastoma patients need to follow a more intensive treatment protocol. Despite therapy involving chemotherapy, tumor resection, radiation, (autologous) stem cell rescue, immunotherapy and treatment with retinoic acid, survival rates of high-risk neuroblastoma patients after the first 5 years remains low, with only around 50% of patients surviving. As around half of this group of patients die, even with an intensive treatment

schedule, a lot of progress in therapy protocols can be made. A more detailed representation of the high-risk treatment protocol can be found in figure 3[36].

1.2.1 The genetic background of neuroblastoma tumors

Because of the extremely heterogenous clinical course of neuroblastoma tumors, neuroblastoma is quite unique among pediatric tumors[39]. The genetic heterogeneity that characterizes these tumors can explain the diversity in disease progression and classification.

1.2.1.1 MYCN amplification

Structural alterations dominate the genomic landscape of neuroblastoma tumors, including amplifications and chromosomal gains and losses. One of the most frequent genetic aberrations is the amplification of the *MYCN* oncogene[40,41]. The MYCN amplicon is located on chromosome 2p24 and more than 20% of all primary tumors show amplification of this locus. This amplification generally consists of 50 to 100 copies, leading to an increase in MYCN oncogene expression and uncontrolled cell proliferation[42]. *MYCN* amplification strongly correlates with advanced disease stage and was the first prognostic genetic marker identified[43].

MYCN belongs to the *MYC* proto-oncogene family and plays a role in embryonic development, with high expression levels in the developing brain[44–46]. In healthy individuals, MYCN is downregulated after embryonic development and virtually absent in adult tissues. Mutation or deletion of the *MYCN* gene results in birth defects or mortality in mice, underscoring its importance in embryogenesis[46]. *MYCN* functions as a transcription factor, with many direct targets that are regulated by binding of MYCN to E-box sequences in a heterodimeric complex with Max[47–49]. Through regulation of these genes, essential pathways including proliferation, growth, apoptosis and differentiation are under *MYCN* control.

Recently it was shown that not only *MYCN*, but also its family member *MYC*, plays an important role in the clinical outcome of high-risk neuroblastoma patients[50]. However, amplification of distal enhancer elements and translocations of enhancers from other

genomic locations to the *MYC* locus cause the upregulation and not amplification of the gene itself, as is the case with *MYCN*.

1.2.1.2 Chromosomal gains and losses

1.2.1.2.1 Segmental aberrations

In addition to amplifications, various larger chromosomal regions are gained and lost in neuroblastoma. These structural abnormalities are predominantly found in advanced stage tumors and are associated with higher relapse rates. As the presence of segmental CNVs associates with worse progression-free survival, it is used for risk determination and therapy stratification. The most frequent CNVs are deletions of chromosome arm 1p and 11q and gains of the 2p and 17q arm.

Deletion of the small arm of chromosome 1 was one of the first chromosomal aberrations to be discovered, with the smallest region of overlap located on 1p36[51,52]. It occurs in 25%-35% of all neuroblastoma patients and is associated with *MYCN*-amplification, in addition to other clinical and genetic variables (age older than 12 months and INSS stage 4). Tumor suppressor genes located in that genomic region include *CHD5*[53,54], a chromatin remodeling protein that regulates transcription of several genes, and *KIF1Bß*[55], a pro-apoptotic factor in sympathetic neurons and mediator of EGLN3 function. Patients with a 1p36 deletion receive a poor prognosis and show higher chances of relapse[56]. Loss of heterozygosity of chromosome arm 11q (11q23) occurs in 35-40% of neuroblastoma tumors[56]. This deletion is mutually exclusive with *MYCN* amplification and 1p deletion and is mostly found in high-risk neuroblastoma tumors. Since several tumor suppressor genes, including *ATM*[57], a tumor suppressor activated by double stranded DNA breaks, lay in this chromosomal region it correlates with INSS stage 4 and unfavorable outcome. Notably, loss of 11q often coincides with deletion of 3p and gain of 17q[56,58]. Gain of 17q is the most frequent observed genetic alteration in neuroblastoma, affecting more than half of neuroblastoma patients. Similar to the other CNVs, strong associations between 17q gain and INSS stage 4, the age at diagnosis over 12 months, 1p deletion and *MYCN* amplification have been identified. Correlation with these clinical and genetic variables results in a more aggressive phenotype[59]. The frequency and association with survival suggest that the gained regions contain one or multiple genes with oncogenic potential, contributing to neuroblastoma tumorigenesis through a dosage effect[60].

1.2.1.2.2 Numerical aberrations

Whereas the incidence of segmental chromosomal variations is higher in high-risk metastatic tumors, a higher presence of numerical copy alterations (NCA), observed as gains and losses of whole chromosomes, is associated with low-risk patients and localized tumors[61]. Diagnosis of NCA tumors occurs often at an age younger than 12 months and have a good prognosis[62,63].

1.2.1.3 Mutations

As neuroblastoma is primarily a copy number disease, only a few recurrent pathogenic mutations have been identified. The protein coding gene with the highest mutation frequency is anaplastic lymphoma kinase tyrosine kinase receptor (*ALK*). 8-12% of all cases harbor *ALK* mutations[64,65]. The majority of the base substitutions are present in the kinase domain, with two mutation hotspots at F1174 and R1275, resulting in a constitutively active protein[66]. This consistent phosphorylation – together with additional mutations – is sufficient for oncogenic transformation[67]. These ALK mutations can be targeted by *ALK* inhibitors, providing new therapeutic opportunities for a subset of patients[68,69]. *ALK* mutations are not only found in sporadic cases, it is also the major driver in neuroblastoma predisposition in familial cases[70].

Around 1-2% of the observed neuroblastoma tumors are familial, inherited in an autosomal dominant way, with the remaining cases occurring in a sporadic fashion[71,72]. Diagnosis of hereditary neuroblastoma usually happens at an earlier age, with primary tumors located at multiple sites in the body. The first mutated gene described to predispose patients to neuroblastoma in a hereditary context was *PHOX2B*[73,74], a regulator in development and differentiation of the nervous system. Only 5-10% of familial neuroblastoma tumors will harbor a *PHOX2B* mutation, suggesting that, next to *ALK* and *PHOX2B,* other predisposition genes are yet to be identified[75,76].

Several independent studies have reported other recurrent somatic mutations, in *ARID1A*[77], *ARID1B*[77], *ATRX*[78] and *PTPN11*[65,79], among others. Loss-of-function mutations or focal deletions of *ARID1A* and *ARID1B* were uncovered through low coverage whole genome sequencing in 11% of cases[65,77]. Interestingly, *ATRX* mutations and deletions are identified in older children (>5 years) and young adults[78]. These inactivating *ATRX* mutations correlate with activation of the ALT pathways, as ATRX represses telomere lengthening through ALT. Gain-

of-function mutations in *PTPN11* (2.9-3.4%) stimulate tumorigenesis and the proliferation of hyperplastic neuroblasts, and promote penetration of the tumor in surrounding tissues[80].

1.2.1.4 Genomic rearrangements

Recurrent genomic rearrangements of a chromosomal locus at 5p15 in the vicinity of *TERT* were also discovered to play a role in the outcome of neuroblastoma[81]. The chromosomal rearrangements place potent enhancer elements in the neighborhood of *TERT*, leading to chromatin remodeling and altered DNA methylation patterns of that genomic region. Although the structural alterations vary between tumors, they all result in higher expression levels of *TERT*, a protein coding gene instrumental in telomere lengthening, preventing deterioration of chromosomal ends allowing them to evade apoptosis. *TERT* rearrangements are almost exclusively identified in high-risk neuroblastoma tumors without MYCN amplification and are associated with a worse prognosis[81].

Another form of genomic rearrangements described in neuroblastoma is chromothripsis, numerous chromosomal rearrangements of a confined region in one or a few chromosomes through fragmentation of chromosomal regions followed by erroneous reassembly[82]. These chromosomal rearrangements affect genes implicated in neuronal growth cone stabilization, the stabilization of the extension of a neurite, including *ODZ3*, *PTPRD* and *CSMD1.* Around 18% of high-risk tumors show signs of chromothripsis, which are associated with poor survival[82].

1.2.2 Towards an integrated molecular-based risk classification of neuroblastoma tumors

Although several copy number variations and mutations correlate with survival in patients, recent findings suggest that the clinical phenotypes can be classified based on alterations affecting telomere maintenance and RAS or p53 pathways[83]. Low-risk neuroblastoma tumors can be identified through the lack of telomerase or ALT activation. Since telomere maintenance is crucial for the cell to obtain immortality, the lack thereof can lead to spontaneous regression and differentiation, regardless of mutations in the RAS or p53 pathways.

However, in high-risk patients, *MYCN* amplification, TERT rearrangements or ALT pathway induction activate telomerase maintenance[83]. Additional mutations in the RAS or p53 pathway reinforce tumor aggressiveness, resulting in a lower chance of survival. As telomerase maintenance and mutations in those two pathways result in a synergistic effect, new combination therapies can be evaluated, targeting both oncogenic pathways.

1.2.3 Neuroblastoma cell identity is heterogenous

Based on gene expression profiles of multiple transcription factors, neuroblastomas can be divided into two main types of tumor cells, (nor)adrenergic and mesenchymal cells (neural crest-like cells)[84,85]. A third state is a mixed type, showing characteristics of both groups. Most neuroblastoma tumors are a mixture of both cell types, often with a larger fraction of adrenergic cells. This heterogeneity is controlled by regulatory programs that monitor specific and divergent expression of super-enhancer associated transcription factors in the two cell types. The difference in expression patterns of these transcription factors determines the cell fate and differentiated state. These core regulatory circuits (CRC) contain well-known neuroblastoma transcription factors, such as *PHOX2B*, *GATA3* and *HAND2* defining the noradrenergic identity, and *AP-1* transcription factors defining the mesenchymal identity[84,85]. The partitioning into these two (or three) classes can have an effect on the treatment course. Mesenchymal tumor cells are more chemoresistant *in vitro* and show a higher prevalence in relapsed and post-therapy patients. This result suggests that therapy exerts selective pressure, enriching for mesenchymal tumor cells[85].

These CRCs mostly consist of lineage dependence genes, which are master regulatory genes during development. Normal functions of these genes are critical within the progenitor cells of the tissue and ensure lineage survival and cell differentiation. Chromatin remodeling together with time-dependent activity of transcription factors regulates these lineage-associated expression programs, distinctive for axis formation, body segmentation and subsequent lineage differentiation. Logically, each lineage is governed by other transcription factors, dependent on their progenitor cells subtypes[86].

A perfect example in neuroblastoma is *PHOX2B*, a transcription factor essential in neuronal lineage development, proliferation and differentiation[87,88]. Although only a small fraction of neuroblastoma tumors harbors *PHOX2B* mutations, its expression persists in, and shows a rather specific expression profile for neuroblastoma cells, whereas other – differentiated cells – lose their *PHOX2B* expression[89]. Furthermore, the presence of PHOX2B is necessary for neuroblastoma tumor survival, as loss of *PHOX2B* expression inhibits neuroblastoma cell proliferation and self-renewal, promoted neuronal differentiation and constrains tumorigenicity[88]. This persistent expression and the subsequent deregulation of survival mechanisms leads to lineage addiction on which the tumor cells depend to survive and proliferate.

1.3 Non-coding RNAs, new players in an old game

RNA-sequencing based exploration of the transcriptome demonstrated that the majority of the genome is being actively transcribed[90,91]. This results in the production of thousands of so-called non-coding RNAs (ncRNAs). As the name suggests, ncRNAs are not translated into proteins and are expected to have a function at the RNA level itself. This type of RNA transcripts has been divided into two major classes, according to their size and function, small non-coding RNAs and long non-coding RNAs. Known subtypes of small ncRNAs include microRNAs (miRNAs), piwi interacting RNAs (piRNAs) and small nucleolar RNAs (snoRNAs)[92]. The long non-coding RNAs (lncRNAs) represent the largest fraction of the non-coding RNA class and are arbitrarily defined as being at least 200 nucleotides in length[92–94].

1.3.1 The history of lncRNA discovery

The discovery of the first eukaryotic lncRNA, namely *H19*[95], was made in murine fetal liver cells. Researchers concluded that *H19* did not encode for a protein, although several small open reading frames were present in the gene. The transcript possessed features similar to mRNAs, such as high sequence conservation across mammals, RNA polymerase II transcription, splicing, polyadenylation and cytoplasmic localization. Induced expression in transgenic mice was lethal, revealing the necessity of tight regulation of *H19* expression and its critical function in embryogenesis[96–98]. While the critical importance of *H19* in embryonic development was established, the role of the RNA transcript itself was only later discovered, revealing its regulation over *IGF2*, a growth factor promoting fetal and placental growth[99]. The fundamental research on *H19* and other key lncRNAs further underscored their biological relevance, sparking the interest in lncRNA identification and characterization in various biological systems and processes[98].

In 2001, a worldwide sequencing effort called the Human Genome Project, generated the first full draft of the human reference genome[100,101]. Interestingly, the number of protein coding genes identified in this project appeared to be lower than predictions based on CpG islands

made in earlier studies[102]. It wasn't until later that large-scale consortium efforts, such as the ENCODE (Encyclopedia of DNA Elements) project[90] and the FANTOM (Functional ANnoTation Of the Mammalian genome) consortium[91], revealed that the human genome is pervasively transcribed and producing thousands of non-coding RNAs. These discoveries would not have been achieved without the availability of high-throughput RNA sequencing (RNA-seq) techniques.

1.3.2 RNA-sequencing

In non-coding RNA research, groups and consortia use different high-throughput sequencing methods. Initial studies, such as research by the RIKEN's FANTOM consortium[103], used cDNA cloning, which resulted in the publication of 34,030 murine polyadenylated lncRNAs. Because of the determination of a growing number of lncRNA sequences, researchers could us microarrays to determine expression values in cellular model systems and primary tissue samples. However, these techniques are limited in the information they provide about the mature RNA structure and, in case of microarray expression analysis, prior knowledge of the mature transcript sequence is necessary. In that aspect, RNA-seq provides a major leap forward as reconstruction of the transcriptome through algorithmic approaches facilitates the discovery of exon structures from RNA-seq reads, without previous knowledge of the RNA sequence or their possible isoforms[104]. The information that RNA-seq supplies, ranges from the transcripts' position, to length measurements and exon-intron structures. RNA-seq can be used to study all (coding and non-coding) or only a subset of RNA transcripts, splice junctions and single nucleotide polymorphisms (SNPs)[105]. This technique quantifies expression levels of transcribed genomic regions in a more precise and more sensitive manner in comparison to microarrays, as the lack of potential cross hybridization and the lower degree of background noise associated with RNA-seq enables easier detection of low abundant genes[105,106]. Several types of RNA-seq exist, such as total RNA-seq, polyA+ RNA-seq and small RNA-seq, enabling us to detect non-polyadenylated RNAs, polyadenylated RNAs and microRNAs. The first lncRNA discoveries were limited to the detection of long intergenic non-coding RNAs (lincRNAs), as the distinction between antisense and sense overlapping transcripts was challenging using unstranded sequencing technologies. Through the invention of stranded library preparation methods, the strand of origin of RNA molecules and the structural organization of overlapping

transcripts could more easily be determined, together with a more precise quantification. However, information obtained about lncRNAs overlapping in sense with protein coding genes was still not accurate. The invention of long-read sequencing addressed this problem, and resulted in a higher effectiveness in (*de novo*) assembly and more accurate identification of overlapping transcripts[107]. The relative ease of these high-throughput RNA-seq techniques allowed many different research groups to profile transcriptomes on large scales in a plethora of different tissues and cell lines, making it the current standard. Other techniques such as global nuclear run-on sequencing[108] (GRO-seq) and cap analysis of gene expression sequencing (CAGE-seq)[109] were developed around the same time as the advent of RNA-seq. GRO-seq was designed to define transcriptional units and discover new transcripts[108]. This technique provides information about genome-wide transcriptional activities at a certain time point, by labelling nascent transcripts with bromouridine (BrU) through RNA polymerases that are already attached to the DNA[108]. Through CAGE-seq, researchers determined the exact transcription start sites of all RNA transcripts, including novel RNA molecules, by capturing the 5'-cap of complete cDNA sequences, adding a polyA-tail and sequencing the cDNA strand[109]. Together, all these techniques provide the research community with powerful, complementary tools, to better the understanding of our transcriptome.

1.3.3 lncRNA classification

lncRNAs can be classified into five different categories based on their genomic location[110,111]. lncRNAs transcribed from regions that do not overlap with protein coding genes are called intergenic lncRNAs (lincRNAs). A subdivision of this category are the enhancer RNAs (eRNAs), which are relatively short (500 to 2000 nucleotides) and transcribed from enhancer elements. In contrast, if transcription of a lncRNA is contained within an intron of a protein coding gene, it is classified as an intronic lncRNA. Sense lncRNAs contain part of or the whole sequence of the protein coding gene, residing on the same strand. Similarly, antisense lncRNAs overlap partially or completely with the protein coding gene, but are transcribed from the opposite strand. The last class, bidirectional lncRNAs, encompass those genes that are transcribed from the opposite strand as the neighboring protein coding gene, but in a diverging manner. Transcription start sites of both genes are typically located within a few hundred base pairs of each other.

1.3.4 lncRNA categories and mechanisms of action

Numerous studies have been published over the last years describing lncRNA functions and the mechanisms by which they operate. Nevertheless, the majority of lncRNAs have not yet been functionally characterized. Although the number of characterized lncRNAs is rather low, the biological processes in which lncRNAs are implicated are plentiful. lncRNAs can act as transcriptional and post-transcriptional regulators in various processes in the cell. These mechanisms are categorized in direct (recruitment and inhibition of transcription factors) and indirect regulation, the latter through chromatin remodeling.

1.3.4.1 lncRNAs as guides or scaffolds in chromatin remodeling

In 1975, it was discovered that purified heterochromatin contained twice as much RNA as DNA, raising the idea that RNA has a functional role in regulating chromatin structure and ultimately, gene expression[112]. The exact class of RNA molecules responsible remained elusive, until multiple lncRNAs involved in controlling access of regulatory proteins to chromatin were identified[113]. LncRNAs are crucial components in the regulation of chromatin structure. Through interaction with chromatin binding proteins and complexes, they can modify specific amino-acids in the N-terminal tails of histones. Histone modification impacts the accessibility of the chromatin, allowing (euchromatin) or restricting (heterochromatin) transcription[114–116]. These enzymes are ubiquitously expressed in human cells, regulating a wide range of genes. However, epigenetic profiles are rather cell type specific, implying that tissue-specific lncRNAs are essential for the genomic binding specificity of chromatin modifying proteins and complexes[92,117].

XIST is the best described lncRNA with chromatin modifying functionality. This lncRNA is involved in dosage compensation through inactivation of the X chromosome. In female cells, only a single X chromosome remains active to ensure protection from information overload. This transcriptional inactivation appears to be initiated during differentiation and coincides with structural modifications resulting in heterochromatin[118]. The regulatory region driving this process contains a 17 kb lncRNA named *XIST* (X-inactive-specific transcript), which in turn controls other regulatory genes in *cis*[119]. *XIST* transcription is restricted to the inactivated X chromosome (Xi) and silences it by coating it and recruiting the chromatin modifying complex

Polycomb Repressive Complex 2 (PRC2)[120]. PRC2 functions as a histone methyltransferase involved in transcriptional silencing by adding three methyl groups to lysine 27 of histone H3 (H3K27me3)[121,122120] (Figure 4a). On the active chromosome, *XIST* expression is controlled by lncRNA *TISX,* resulting in extremely low expression *XIST* levels thereby preventing X chromosome inactivation.

Several lncRNAs interact with PRC2, such as *HOTAIR*[123], *XIST*[124], *Kcnq1ot1*[125] and *ANRIL*[126]. *HOTAIR*, a lncRNA on chromosome 12, is required for PRC2 occupancy at the HOXD locus on chromosome 2 and subsequent H3K27me3 mediated epigenetic repression[123]. This was the first report of lncRNA-based gene regulation in *trans*. *Kcnq1ot1*, functional in imprinting, binds with PRC2 and recruits the complex to chromatin[125]. This PRC2-lncRNA complex is associated with repression of protein coding genes over a 1-Mb region. The PRC2-induced H3K27me3 repressive mark causes the silencing in this locus[127]. *Kcnq1ot1* is transcribed from this 1-Mb region, regulating transcription of neighboring genes as well as distal genes, suggesting *cis*- and *trans*-regulation as the working mechanisms[121].

Aside from PRC2, other lncRNA-chromatin modifying complexes have been identified. HOXA transcript at the distal tip (*HOTTIP*) regulates chromatin interactions in the *HOX* cluster[128]. It coordinates gene activation through binding of WDR5, a protein essential for embryonic stem cell self-renewal. The latter adaptor protein binds with an MLL H3K4 methylase complex through recruitment by *HOTTIP*, catalyzing and preserving H3K4 methylation. The presence of *HOTTIP* is necessary in the formation of the methylation complex and thus for maintenance of the methylation pattern.

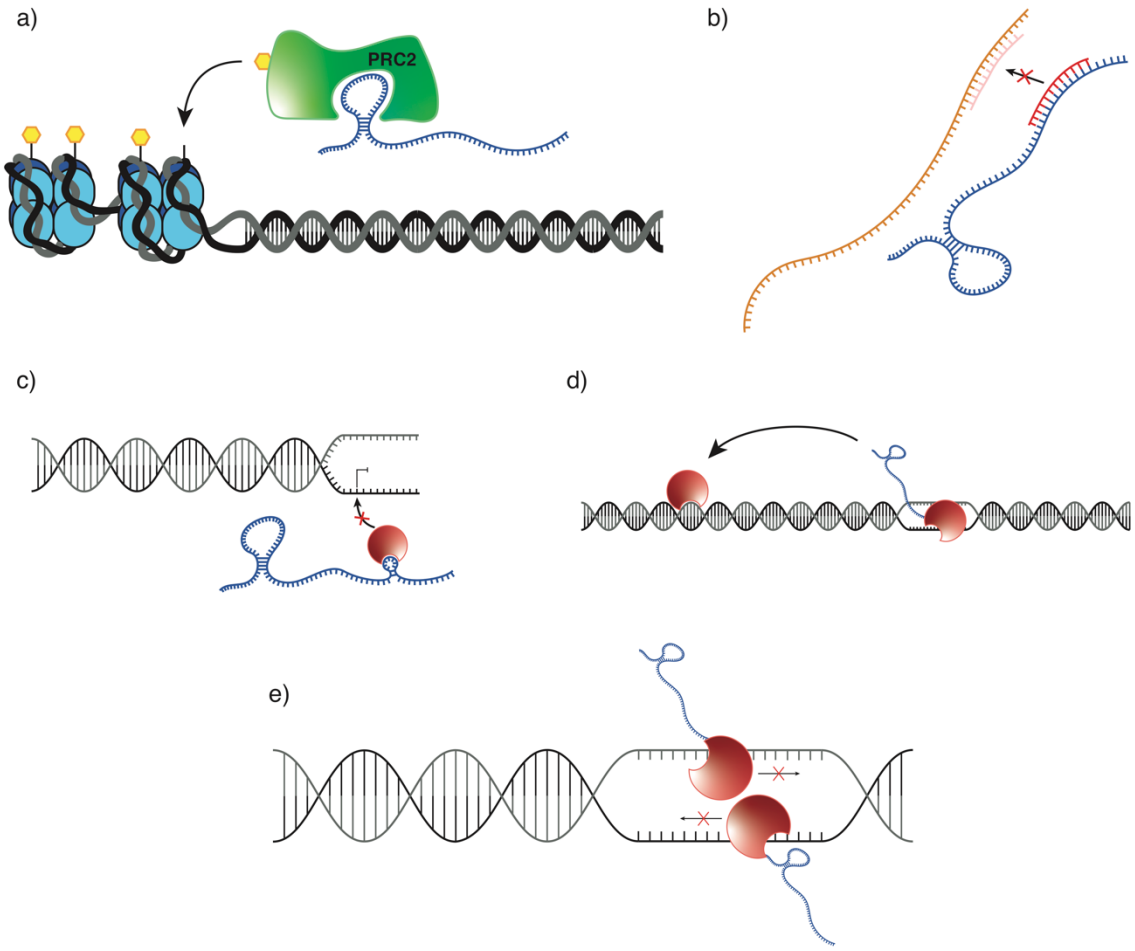


Figure 4: The different functions of lncRNAs. a) LncRNAs can act as guides or scaffolds for histone-modifying complexes, such as PRC2. Through transfer of methyl groups to the histones by PRC2, the target genes are silenced. b) miRNAs are regulators of translation through degradation by binding complementary RNA sequences. LncRNAs can inhibit their function by binding and titrating them away, preventing downregulation. c) LncRNAs can bind proteins by containing their specific binding sequences or mimicking secondary structures and act as decoy, regulating transcription of its target genes. d) Enhancer RNAs can induce transcription of neighboring genes in a cis-regulatory manner. e) Natural antisense transcripts can inhibit transcription by steric hindrance of the RNA polymerase II that moves in opposite directions and collide. A NAT can also bind the complementary RNA strand and inhibit or promote translation.

eRNAs are lncRNAs transcribed from enhancer regions, impacting chromatin interactions in the region[92]. These eRNAs regulate transcription in *cis* through chromatin looping by acting as tethers attached to the DNA, attracting interacting proteins to the region (Figure 4d). Loss of these eRNAs leads to a decrease of insulated neighborhoods, reducing gene expression in that region. *KLK3e*, an eRNA upstream of *KLK3*, an androgen receptor (AR) regulated gene which codes for prostate-specific antigen (PSA), structures the chromatin to allow interaction between the promoter of *KLK2* and the *KLK3* enhancer elements to activate transcription of *KLK2*. The AR-associated protein complex, using the eRNA, that modulates the interaction between the two genomic regions, as a scaffold, will be able to regulate specific transcription of AR-dependent genes[129].

1.3.4.2 Decoy lncRNAs

Another mechanism lncRNAs employ to regulate transcription is to act as a molecular decoy. By binding to a protein, lncRNAs can titrate it away from its interaction partners, thus rendering the protein incapable of exerting its function[117] (Figure 4c). lncRNAs – such as *PANDA*, a p53-dependent RNA molecule – categorized as decoys do not have any other function than sequestering their targets. Upon DNA damage, *PANDA* expression is induced and inhibits the activation of apoptotic genes through binding to NF-YA, a transcription factor in control of the apoptotic program[130]. Through depletion of NF-YA, *PANDA* promotes cell survival. Decoys inhibit other transcription factors by as well. Repression of the glucocorticoid receptor occurs through binding to a specific portion of the *GAS5* lncRNA[131]. One of the 6 formed hairpin structures is composed of glucocorticoid response elements-like sequences, binds the glucocorticoid receptor and inhibits its normal function as a transcription factor. The competition between *GAS5* and the DNA glucocorticoid response elements for binding to the DNA binding domain of the glucocorticoid receptor restricts access of the latter to the chromatin.

In addition to sequestering transcription factors, lncRNAs can interact with DNA methyltransferases (DNMTs)[132]. The three major DNMTs – DNMT1, DNMT3A and DNMT3B – interact with lncRNAs, resulting in altered DNA methylation patterns and gene expression levels. These lncRNAs regulate genes in *cis* or *trans* by acting as decoys for DNMTs, preventing methylation and gene silencing. *DALI*[133], a lncRNA essential for neuronal differentiation,

interacts with the DNMT1 machinery. Interestingly, DNMT1 binds secondary RNA structures with a higher affinity than DNA through its catalytic domain. The folded *DALI* transcript competes for binding to co-factors or the DNMT1 DNA substrate. The competitive binding inhibits DNA methylation at a subset of bound and regulated regions, preventing gene silencing in *trans*. Gene silencing of *CEBPA* happens in a cis-regulatory manner [134]. A lncRNA spanning the entire *CEBPA* region, extra-coding CEBPA (*ecCEBPA*), is transcribed together with *CEBPA*. Structural analysis revealed its secondary stem-loop structure to be instrumental for DNMT1 interaction, preventing the genomic region to be methylated through sequestration of DNMT1.

lncRNAs can also interact with other non-coding RNA classes, such as miRNAs (Figure 4b). This large, well-studied type of small non-coding RNAs has an important role in gene regulation[135]. They bind to RNA molecules by incomplete complementary base pairing with seed sites in the coding sequence or 3' UTR. The miRNAs recruit protein complexes that affect translation or lead to RNA degradation, influencing the target protein abundance. *PTENP1*, the pseudogene of tumor suppressor *PTEN*, has recently been identified as a miRNA interacting lncRNA. The 3' UTR of this pseudogene and the actual tumor suppressor gene *PTEN* exhibit multiple similar miRNA seed sequences. The homology of the 3' UTRs allows *PTENP1* to act as a miRNA decoy, thereby restricting miRNA binding to and reducing the degradation of *PTEN* mRNA[136]. Such lncRNAs are called miRNA sponges or competing endogenous RNAs.

1.3.4.3 lncRNAs as naturally antisense transcripts (NATs)

NATs are transcripts that partly overlap with protein-coding or non-coding genes, but are transcribed from the opposite strand. These transcripts are pervasive in the human genome, as over 60% of genomic regions are bidirectionally transcribed[137]. NATs can have either an activating or repressing effect on their sense transcripts through a range of mechanisms[138]. A first mechanism to impact the neighboring gene's expression, is through transcriptional collision[138] (Figure 4e). Transcription by RNA Polymerase II on both strands will lead to transcriptional stalling, as the complexes are not able to pass each other because of steric hindrance, thus inhibiting RNA production from the sense strand.

Another mechanism through which NATs regulate gene expression is duplex formation between sense and antisense transcripts**.** The duplex RNA modulates sense RNA expression through degradation, alternative splicing and alterations in localization, transport or stability[138]. Double stranded RNA resembles foreign, exogenic structures (e.g. viral particles), which can trigger degradation of the RNA molecules. Binding of the antisense to the sense strand can also result in changes in isoform abundance through alternative splicing. *UXT-AS1* is a NAT complementary to the 5' end of *UXT*, a protein coding gene with two isoforms (*UXT1* and *UXT2)* that are involved in transcriptional regulation[139]. In the presence of *UXT-AS1*, a decrease in *UXT1* expression is accompanied by an increase in the expression of *UXT2*, leading to a reduction in cell apoptosis and promotion of cell proliferation. Due to the complementarity of NAT with the 5' end of *UXT*, it is suggested that *UXT-AS1* promotes a switch in alternative splicing, resulting in preferential translation of the *UXT2* transcript from the second start codon.

1.3.4.4 Topological anchor point RNAs (tapRNA)

TapRNAs are a subtype of lncRNAs that are positionally conserved and associated with developmental transcription factor regions, showing co-expression in a tissue-specific manner[140]. The majority of these lncRNAs alter chromatin structure and associate with CTCF binding loci, chromatin loop anchor points and borders of topologically associated domains (TADs). Through loop formation, tapRNAs are likely to come into contact with enhancer sequences at the other end of the loop.

Although tapRNA sequence conservation is less pronounced compared to their protein coding counterparts, human tapRNAs show on average a 31% similarity with syntenic mouse genes. In around 73% of tapRNAs, highly conserved stretches of sequences have been found between human and mouse. These conserved non-coding genes all show high enrichment for developmental pathways. Similar relations to development have been observed for associated protein coding genes in the same insulated neighborhoods. Generally, a set of specific factors regulates these protein coding genes, in an identical fashion to the tapRNAs. Moreover, tapRNAs are essential for the expression of the associated coding genes, since knockdown of these lncRNAs leads to reduction or loss of the protein coding RNA transcript

and vice versa. Notably, tapRNA expression levels are conserved across mouse and human tissue[140].

Several tapRNAs – for example *FOXA2-DS-S* – have already been identified. *FOXA2-DS-S* is associated with the *FOXA2* protein coding gene, a master regulator in the liver[141]. The same transcription factors occupy promoter regions of both genes, indicating a partial explanation for the observed co-expression profile[140]. Interestingly, knockdown of the tapRNA results in *FOXA2* downregulation and loss of *FOXA2* expression results in a reduction of *FOXA2-DS-S* in liver and lung cancer cells. Other tapRNAs and their associated coding genes have been shown to be deregulated in cancer cells, stressing their importance in disease and their potential as a therapeutic target[140].

1.3.5 lncRNAs in cancer

Given their pivotal role in many critical cellular pathways, it's not surprising that lncRNA dysregulation is frequently linked with disease. Multiple diseases, such as Alzheimer's disease[142], cardiovascular disease[143], asthma[144] and COPD[145] show involvement of lncRNAs in their origin and pathogenesis. However, the involvement of lncRNAs in cancer has been studied most extensively. As mentioned before, lncRNAs have a multitude of functions, impacting diverse regulatory pathways. Today, multiple examples of lncRNAs can be categorized into the six hallmarks of cancer, introduced by Hanahan and Weinberg in 2000[146] (Figure 5).

1.3.5.1 lncRNAs involved in sustaining proliferative signaling

Several cancers pertaining our reproductive organs show aberrant expression or function of hormone receptors, such as the estrogen (ER), progesterone (PR) and androgen (AR) receptors[146,147]. Co-factors often affect functionality of these receptors, with the possibility of lncRNAs acting as such co-factors. *SRA* (steroid receptor RNA activator) is a coactivator of the steroid receptors and activates them through their AF-1 domain[148]. Increased *SRA* abundance may lead to aberrant ER/PR activity during breast cancer progression[149]. Another example of a lncRNA with a role in cell proliferation is *PCAT-1* (prostate cancer associated transcript 1)[150]. A subset of metastatic and localized prostate cancers shows high expression of this lncRNA. If

overexpressed, cells display a higher rate in cell proliferation, whereas loss of the lncRNA results in a decreased proliferative rate. Other examples influencing cell proliferation are *PANDA*[130], *lincRNA-p21*[151] and *MEG3*[152].

1.3.5.2 lncRNAs involved in evading growth suppressors

In addition to mutation or indel induced inactivation of growth and proliferation inhibitors, such as *TP53* or *PTEN*, cancer cells can acquire additional ways to impede tumor suppressor functions through the assistance of lncRNAs. *ANRIL* blocks the activity of tumor suppressor genes by interacting with SUZ12, a component of the PRC2[126]. The complex represses p15 expression through methylation, leading to sustained growth. Moreover, a subunit of the PRC1, namely CBX7, is an interaction partner of *ANRIL* as well, meaning *ANRIL* expression also results in recruitment of PRC1 to the p16/p14 locus, followed by methylation and silencing of the region[153].

In addition to helping protein coding genes exert their tumor suppressive function, lncRNAs can also act as tumor suppressors themselves. As mentioned before, *GAS5* binds to the DNA binding domain of glucocorticoid receptors, competing with the response elements – binding sequences of specific transcription factors in a gene's promoter region – in the genome[131]. This competitive interaction suppresses activation of multiple targets, eventually sensitizing the cells to apoptosis[131,154]. Furthermore, *GAS5* has a regulatory role in the mTOR pathway – which regulates cellular growth, protein synthesis and proliferation – via the Rapamycin immunosuppressant[155]. Low expression levels of *GAS5*, as observed in breast cancers compared to normal breast tissue[154], protect the tumor cells against the antiproliferative effect of Rapamycin. Moreover, genetic alterations in the *GAS5* locus have been identified in several cancer types, among which melanoma, breast and prostate cancer[146,156,157].

1.3.5.3 lncRNAs involved in replicative immortality

90% of all cancers acquire the ability to circumvent the shortening of the telomeres by expressing telomerase. The activity of telomerase critically relies on the lncRNA *TERC* (Telomerase RNA Component)[146]. TERT, a telomerase reverse transcriptase, reverse transcribes a small part of *TERC*, essential for the telomere synthesis. Interestingly, amplification of the *TERC* gene has been detected in several human cancers[158,159].

RNA transcripts derived from telomeric and subtelomeric regions, jointly coined *TERRA*, are another group of lncRNAs involved in telomere lengthening[160]. A reduction in *TERRA* expression levels is essential for telomerase-mediated telomere lengthening[161]. The method of *TERRA* inactivation, either through dysregulation, silencing or mutation, remains elusive[162]. However, because of the connection with telomere maintenance, a link with cancer initiation and sustainment is possible.

1.3.5.4 lncRNAs involved in invasion and metastasis

Before a cancer cell can metastasize, deregulation of many complex interactions and regulatory mechanisms is necessary[20,21]. Recently, research by Ping et al. established a role for lncRNAs in this process. *MALAT1*[163], an extremely abundant lncRNA in many cancers, is a key player in the pre-mRNA processing pathway in the nuclear speckles[164] and regulates alternative splicing by modulating the levels of active serine/arginine splicing factors through phosphorylation[165]. In lung cancer**,** *MALAT1* has been proposed as a potential driver of the metastatic process by regulating motility related genes in a transcriptional or post-transcriptional manner[166].

HOTAIR is also involved in cancer metastasis, with overexpression of the lncRNA and increased interaction with PRC2 in primary tumors and metastases[167]. A higher expression level of *HOTAIR* alters gene expression by PRC2 mediated H3K27 methylation, leading to increased cancer invasiveness and metastases. On the other hand, depletion of *HOTAIR* reduces cell invasion in multiple cancer types, including breast and liver cancer[167,168].

1.3.5.5 lncRNAs involved in inducing angiogenesis

Expression of *HIF1A* – a critical regulator of angiogenesis – is inhibited by *aHIF*, a NAT complementary to the 3' UTR of *HIF1A*. Elevated *aHIF* expression levels induce HIF1a mRNA degradation. Notably, several cancers express *aHIF*, and in breast cancer it is even used as a marker for poor prognosis[169–171].

sONE is a second NAT associated with angiogenesis. Hypoxic conditions induce the lncRNA and in turn it negatively regulates the expression of *eNOS* in a post-transcriptional manner. Hypoxia in endothelial cells decreases the *eNOS* mRNA stability, whereas the half-life of the *sONE* transcript increases. In normal conditions, a multiprotein complex stabilizes the *eNOS* RNA, whereas stress conditions can lead to interactions of the *sONE* and *eNOS* transcripts.

This interaction can impede formation of the stabilizing complex, resulting in the destabilization of the *eNOS* mRNA. Competition of *sONE* and *eNOS* RNA molecules for the stabilizing complexes is another possible manner of post-transcriptional regulation[172,173].

1.3.5.6 lncRNAs involved in resisting cell death

Since lncRNAs influence cell death decisions, they might provide insights in cell death evasion. *PCGEM1*, a prostate-specific and prostate cancer associated lncRNA, is an apoptosis inhibitor. It is hypothesized that overexpression of *PCGEM1* results in a delayed p53 and p21 response, creating an anti-apoptotic effect[174].

CUDR is another lncRNA with anti-apoptotic functions. When squamous carcinoma cells are treated with doxorubicin or etoposide, two chemotherapeutic compounds, expression of *CUDR* confers resistance to the treatment as well as the drug-induced apoptosis. Observed downregulation of caspase 3 after *CUDR* overexpression might explain the underlying mechanism[175].

DNA damage induces *PANDA*, a lncRNA from the *CDKN1A* promoter, limiting the transcription of pro-apoptotic genes through interaction with transcription factor NF-YA in *trans*[130]. *PANDA* and *CDKNA1* share a p53 binding site and are both regulated by p53. Knockdown of *PANDA* activates a series of genes associated with apoptosis, including *APAF1*, *BIK*, *FAS* and *LRDD.* However, depletion of *PANDA* has no effect on p53 or *CDKNA1* expression, implying an independent activity of CDKN1A and a lack of a regulatory role over p53. Interestingly, in metastatic ductal carcinomas selective induction of *PANDA* could be observed without simultaneous expression of p21, whereas normal breast tissue does not exhibit this partial regulation.

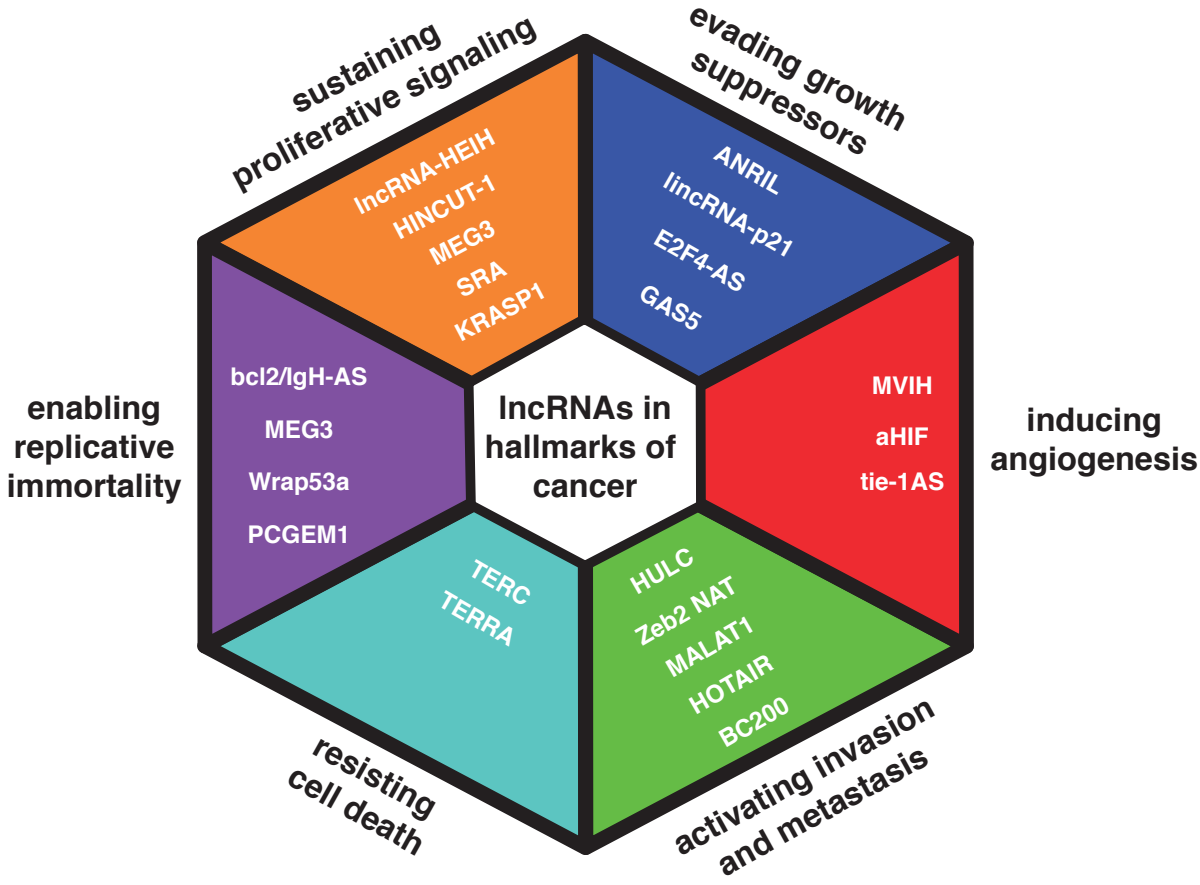


Figure 5: LncRNAs implicated in the hallmarks of cancer. These lncRNAs are examples of non-coding genes implicated in the procurement of their associated hallmarks in the development of cancerous malignancies. (adapted from: A pathophysiological view of the long non-coding RNA world.)

1.3.6 lncRNAs in neuroblastoma

While plenty of lncRNAs have been described in context of cancer[146], only a few seem to be involved in neuroblastoma[176], including but not limited to *GAS5*[177], *CASC15-S*[178], *lncUSMycN*[179], *linc00467*[180], *NBAT1*[181] and *SNHG1*[182]. Description of several interesting ones follow below.

1.3.6.1 NBAT1

NBAT1, located on chromosome 6p22, was discovered as one of the most significantly downregulated lncRNAs in high-risk versus low-risk tumors[181]. In addition, it harbors a SNP associated with high-risk disease in its intronic region. Low *NBAT1* expression correlates with an unfavorable outcome in patients, revealing its tumor suppressive potential. Several factors

contribute to NBAT1 inactivation, including promoter hypermethylation. In addition, the intronic SNP overlaps a region with enhancer-like traits and mediates enhancer activity. The presence of the wild type A/A or heterozygous A/G genotype shows a lower rate of proliferation and invasiveness in comparison to the risk-associated G/G genotype. NBAT1 expression is substantially lower in neuroblastoma cells harboring the G/G genotype.

NBAT1 inhibits cell proliferation and invasion through downregulation of oncogenes such as *SOX9, VCAN and OSMR*, among others[181]. The inhibition of *NBAT1* target genes is epigenetically controlled by EZH2, with *NBAT1* acting as a scaffold for EZH2 recruitment. Loss of EZH2 or *NBAT1* results in a decrease in H3K27me3 in the promoter regions of the protein target genes of *NBAT1*, thereby upregulating their expression. In addition to its role in apoptosis, *NBAT1* is also involved in differentiation.

1.3.6.2 lncUSMycN

LncUSMycN maps to a region on chromosome arm 2p, upstream of *MYCN*, and often co-amplified with *MYCN*[179]. Expression levels are high in *MYCN* amplified neuroblastoma cells and tumors as *lncUSMycN* is frequently co-amplified with *MYCN*. In contrast, neuroblastoma tumors without MYCN amplification do not express *lncUSMycN*. *lncUSMycN* promotes *MYCN* protein expression by binding to the NonO protein, which increases *MYCN* levels in a post-transcriptional manner. It also induces neuroblastoma cell proliferation and associates with poor patient prognosis. Knockdown of *lncUSMycN* reduces *MYCN* expression in *MYCN* amplified tumors and impedes tumor progression in mice.

1.3.6.3 Linc00467

MYCN is a transcription factor and a key driver gene in neuroblastoma. It regulates a multitude of protein target genes. However, recent studies elucidated lncRNAs to be under the control of MYCN as well[183–185]. One example in neuroblastoma is *linc00467*, an inhibitor of *DKK1*. DKK1 is known to activate apoptosis and acts as a tumor suppressor[180]. Reduction of *linc00467* expression impacts the number of viable neuroblastoma cells and induces apoptosis through the loss of inhibition on *DKK1* expression and functionality, suggesting a role for the lncRNA in tumorigenesis.

1.4 RNA therapeutics

The myriad of lncRNAs identified in cancer and other diseases sparked the interest in exploiting them therapeutically. At present, several technologies have been successfully used to knock down lncRNAs, including antisense oligonucleotides (ASO) and RNA interference (siRNA/shRNA). These technologies are not only excellent research tools, they also have potential for application in patients. As of January 2017, the Food and Drug Administration (FDA) approved six ASOs as a clinical therapy. Moreover, in 2018 the FDA approved the first RNAi based therapeutic.

1.4.1 Antisense oligonucleotides

ASOs are small, synthetic DNA sequences that can bind endogenous RNA targets through sequence complementarity. The endogenous RNAseH enzyme recognizes the resulting DNA-RNA hybrid and will degrade the target RNA[186]. The ubiquitously expressed endonuclease RNAseH1 will only cleave the RNA strand of the RNA/DNA hybrid[187], leaving the ASO intact for successive hybridizations (figure 7a). As the phosphodiester backbone of the ASO is sensitive to degradation, modifications are implemented to improve the stability and affinity[186]. Because of the improved affinity, similar knockdown levels can be achieved with lower doses, reducing possible adverse side-effects. The modifications are classified in three categories: modifications of the backbone, ribose modifications and base modifications[188] (figure 6).

The most widely used modification of the backbone is the replacement of the phosphodiester bond with a phosphorothioate (PS) link, by substituting a non-bridging oxygen atom with a sulfur atom[186,188,189]. The stability increases through increased nuclease resistance. In addition, cellular uptake and protein binding are enhanced, and the capacity to induce RNAseH mediated degradation is preserved[188]. Second generation ASOs also frequently use modifications to the ribose, including replacement of the 2'-hydroxyl by 2'-O-Methyl (2'OMe), 2'-O-Methoxyethyl (2'MOE) or 2'-fluoro. ASOs that use ribose modifications are often constructed in a chimeric gapmer configuration**,** comprised of a PS DNA sequence core flanked by modified sugars at both ends[188]. Bridging of the 2'-oxygen with the 4'-carbon is another alteration to form bridged nucleic acids (BNA)[190,191]. The most commonly used modification in

this class are locked nucleic acids (LNA), using a methylene group linker[192]. This type of modification increases the strength of the hybridization, as the ribose is locked in a conformation optimal for complementary base pair binding. The higher affinity allows the ASO sequences to be shorter. Similar to the ribose modifications, LNAs are often used in a chimeric gapmer conformation[191] (figure 6). Although the modifications increase half-life and binding efficacy, they can lead to the obstruction of RNAseH cleavage if used without a chimeric gapmer configuration[191]. Hybrids formed between 2'-modified sequences and endogenous RNA will not be cleaved by RNAseH, as the 2'-modified nucleotides resemble more to RNA than they do DNA. However, the steric hindrance due to ASO binding can be used to prevent interaction with proteins and modulate alternative splicing.

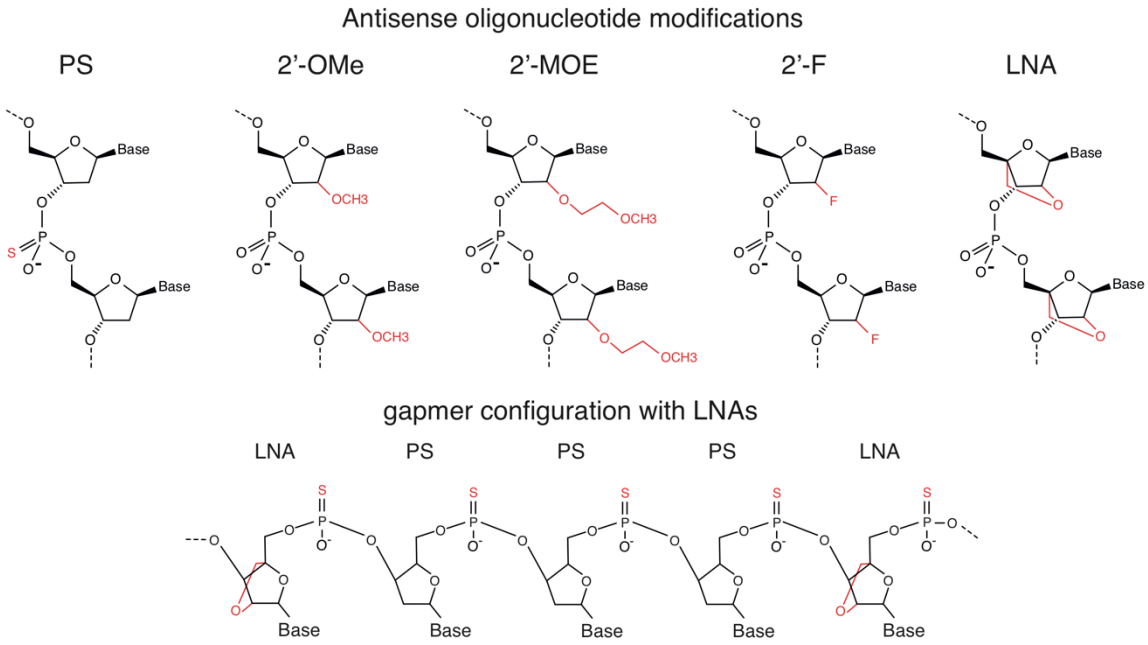


Figure 6: The different types of chemical modifications used in ASO technology.

The sequence complementarity should ensure specific binding and degradation of the target RNA[191]. However, off-target effects still occur and should not be dismissed[191,193]. Oligonucleotide-based therapies are susceptible to unwanted interactions as a consequence of their small size, negative charge and ability to bind both RNA and DNA. ASOs also have the potential to bind RNA- or DNA-interacting proteins. Although the interaction will be less strong in comparison to their native binding partners, they can still occur. Moreover, off-target binding with partial sequence complementarity is conceivable and might alter expression of

these genes. The elimination and understanding of these off-target effects is extremely challenging and is essential when designing clinically applicable ASOs with tolerable side-effects[191].

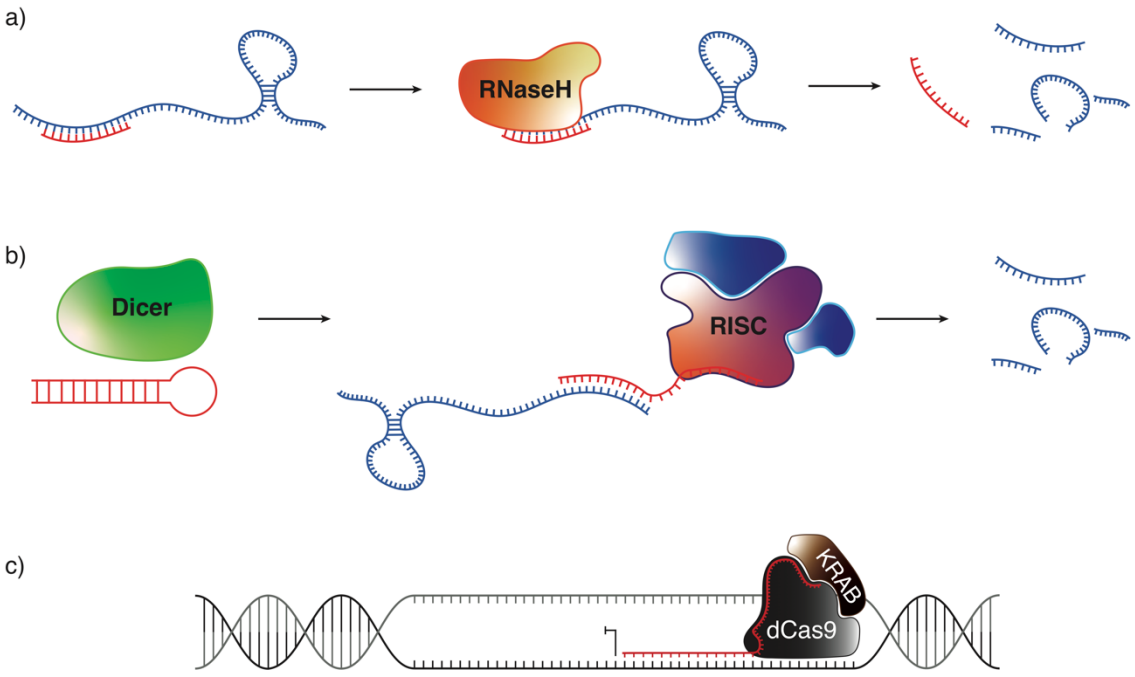


Figure 7: Methods of lncRNA inactivation. a) Antisense oligonucleotides are modified DNA sequences with a complementary sequence to the target of interest. Through the interaction of the complementary sequences, the DNA:RNA hybrid will be recognized by RNaseH. RNaseH will cleave the RNA transcript, rendering it inert. b) siRNAs enter the cell as double stranded RNA (dsRNA) molecules, which are processed into dsRNAs of 21 nucleotides by Dicer. The dsRNA will associate with the RISC complex. Sequence complementarity of the guide strand will lead the RISC complex to the target, after which the target RNA transcript will be degraded. c) Specific recognition of the DNA element of interest is provided by a sgRNA that has bonded with a dCas9 protein linked with a KRAB repressor domain. The dCas9-KRAB complex will bind to that genomic locus and will impede transcription through steric hindrance and repressive activity of the KRAB domain.

However, one of the biggest challenges remains the efficient delivery of the ASOs to their target cells. In case of a systemic administration of the ASOs, the first barrier that needs to be overcome is the localization to the tissue of interest, without losing part of the therapeutic concentration in surrounding irrelevant tissues[194]. This is the biggest challenge, as several barriers block efficient delivery at the therapeutic site of interest, including the vascular endothelial barrier, the reticuloendothelial system and the blood brain barrier. With systemic

administration, the rate of excretion of the ASOs, which will influence pharmacokinetic effects and biodistribution of the therapeutic molecules, should also be taken into account. Several approaches can be taken to evade the pitfalls of delivery, such as sugar and backbone modifications of the ASOs resulting in higher retainment rates and lower percentage of degradation or cell specific targeting through conjugation of a ligand to the ASOs that will bind a specific receptor. Other approaches include encasing the ASOs in lipid nanoparticles or polymeric nanocarrier[194].

The following barrier is transporting the ASO into the correct cellular compartment and avoiding degradation[194]. All free oligonucleotides enter the cell through endocytosis in the same manner. Afterwards, the ASOs will be transported through an elaborate network of intracellular compartments. Eventually, the majority of ASOs internalize in multivesicular bodies (MVB), which will ultimately fuse with lysosomes leading to degradation. However, a small part of the enveloped ASOs will be able to spontaneously leave the MVB and roam around in the cytosol and nucleus, in search of their intended target[194].

The first FDA-approved ASO was fomivirsen[191,195], a 21 nucleotide PS-modified DNA sequence targeting and blocking translation of mRNA transcribed from *UL123*, a gene critical for the cytomegalovirus (CMV). Fomivirsen is used as treatment for CMV retinitis – an inflammation of the retina – in immunocompromised patients, i.e. AIDS-patients. Delivery of the ASO occured directly in the eye and was beneficial for the patients[191,196]. However, this method was not used very long as anti-retroviral therapies emerged and became the primary method of treatment. The most recently approved RNA therapeutic was inotersen[197], an ASO shown to reduce mutant TTR production in transthyretin amyloidosis, where misfolding of the TTR protein leads to heart dysfunction and failure[191,195]. The other FDA approved ASO therapeutics are mipomersen[198], eteplirsen[199,200], nusinersen[201] and volanesorsen[202].

1.4.2 RNA interference

Another method to achieve RNA degradation is by means of RNAi, using small interfering RNAs (siRNAs) or short hairpin RNAs (shRNAs). SiRNAs are double stranded sequences, processed into ~21 nucleotide RNA molecules by DICER[203,204]. The dsRNA associates with the argonaute

2 (Ago2) containing RNA-induced silencing complex (RISC), releasing the passenger strand from the dsRNA while protecting the guide strand[203]. This results in the activation of the complex. Target specificity is provided by the guide strand, which recognizes the RNA target sequence through complementary base pairing. Ago2 cleaves the target sequence if both sequences are fully complementary, causing RNA degradation and gene silencing[203,205]. As the RISC complex generally localizes to and functions in the cytoplasm, this strategy is best applied on RNA transcripts that are primarily confined to the cytoplasm (figure 7b). shRNAs rely on the same method of action as do siRNAs, however, they are expressed by the cell's machinery through transcription of inserted plasmids. These plasmids can be delivered through bacterial or viral vectors, after which they can be integrated in the host's genome. As lentiviral vectors incorporate the shDNA into actively transcribed chromatin, expression of the shRNA is mediated by RNA polymerase II or III[206].

Limitations of the siRNA technology are similar to those found for ASOs. Off-target effects due to partial sequence complementarity can occur, leading to erroneous gene silencing[207]. Moreover, the delivery reagents might cause cytotoxic effects resulting in global degradation and inhibition of mRNA translation[208]. Interference with miRNA pathways, by competing with endogenous RNAs for the protein machinery necessary for miRNA functionality, might cause saturation of the normal pathways and their function[209]. Another limiting factor in the case of siRNAs is the delivery to the intracellular target considering the negative charge, molecular weight and size[210]. Once inside the cell, lysosomes will degrade the siRNAs, greatly reducing siRNA activity.

Presently, there is only one FDA-approved siRNA, called patisiran[211,212]. The molecule is delivered encased in a lipid nanoparticle and produces a robust, continuous knockdown of TTR expression in patients with transthyretin amyloidosis[191].

1.4.3 CRISPR interference

CRISPR interference (CRISPRi) is a recent, complementary approach to RNAi and ASO based gene expression modulation[213]. It is an RNA-based method for targeted repression of transcription[214], by binding a catalytically inactive Cas9 protein (dCas9) to a specific locus in

the DNA. A customizable single guide RNA (sgRNA) recognizes the specific recognition of the DNA element, complementary to the target sequence. As the dCas9 lacks endonucleolytic activity, it will not cleave the DNA strand, but will induce steric hindrance that interrupts transcription, leading to gene silencing. Because of the lack of DNA cleavage, this approach is fully reversible[214]. To further improve transcriptional silencing, the dCAS9 is often linked to a KRAB repressor domain (figure 7c).

While this technology can achieve potent silencing of gene expression, there are several limitations. For the dCas9 protein to be able to bind the DNA, a PAM sequence is required[214]. This requirement limits the number of potential genomic binding sites. Nevertheless, there is risk of off-target effects[214]. Furthermore, the degree of repression differs between genes, as chromatin conformation will affect the ability of dCas9 to bind the DNA. While CRISPR-based technologies for gene expression silencing are increasingly used in research, they are not introduced in the clinic yet. However, there are *in vivo* clinical trials up and running in which the CRISPR methodology – regarding genome editing – is used to provide treatment. In one of these cases, a point mutation in *CEP290*, a gene implicated in a structural role in the cilia of light-sensing photoreceptor cells in the retina, is corrected through CRISPR-based genome editing, with two sgRNAs leading the CRISPR protein to the correct location. Another company received FDA approval for an ex vivo CRISPR-mediated therapy in ß-thalassemia and sickle cell disease[215]. Other examples are clinical trials conducted in the treatment of several cancers, including melanoma[216].

1.5 References

1. Registery, B. C. *Cancer incidence in Belgium 2008*. (2011).
2. Ritchie, M. R. and H. Cancer. (2019). Available at: https://ourworldindata.org/cancer.
3. Oke, J. L., O'Sullivan, J. W., Perera, R. & Nicholson, B. D. The mapping of cancer incidence and mortality trends in the UK from 1980–2013 reveals a potential for overdiagnosis. *Sci. Rep.* **8**, 14663 (2018).
4. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
5. Registery, B. C. *Cancer Incidence Projections in Belgium, 2015 to 2025*. (2017).
6. Lodish, H. *et al.* Proto-Oncogenes and Tumor Suppressor Genes. in *Molecular Cell Biology* (ed. Freeman, W. H.) (2000).
7. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
8. Friend, S. H. *et al.* A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**, 643 (1986).
9. Eischen, C. M. Genome Stability Requires p53. *Cold Spring Harb. Perspect. Med.* **6**, a026096
10. Eischen, C. M. & Lozano, G. The Mdm network and its regulation of p53 activities: a rheostat of cancer risk. *Hum. Mutat.* **35**, 728–737 (2014).
11. Frank, B. *et al.* Copy number variant in the candidate tumor suppressor gene MTUS1 and familial breast cancer risk. *Carcinogenesis* **28**, 1442–1445 (2007).
12. Rajalingam, K., Schreck, R., Rapp, U. R. & Albert, Š. Ras oncogenes and their downstream targets. *Biochim. Biophys. Acta - Mol. Cell Res.* **1773**, 1177–1195 (2007).
13. Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1**, 239–259 (2009).
14. Matsui, A., Tatsuya, I., Hiraku, S., Hirofumi, M. & Kentaro, S. Gene amplification: mechanisms and involvement in cancer. *BioMolecular Concepts* **4**, 567 (2013).
15. Klein, G. & Klein, E. Conditioned Tumorigenicity of Activated Oncogenes. *Cancer Res.* **46**, 3211 LP – 3224 (1986).
16. Miller, D. M., Thomas, S. D., Islam, A., Muench, D. & Sedoris, K. c-Myc and cancer metabolism. *Clin. Cancer Res.* **18**, 5546–5553 (2012).
17. Wee, Y., Wang, T., Liu, Y., Li, X. & Zhao, M. A pan-cancer study of copy number gain and up-regulation in human oncogenes. *Life Sci.* **211**, 206–214 (2018).
18. Zheng, J. Oncogenic chromosomal translocations and human cancer. *Oncol. Rep.* 2011–2019 (2013). doi:https://doi.org/10.3892/or.2013.2677
19. Haller, F. *et al.* Enhancer hijacking activates oncogenic transcription factor NR4A3 in acinic cell carcinomas of the salivary glands. *Nat. Commun.* **10**, 368 (2019).
20. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
21. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
22. Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A. & Reddel, R. R. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat. Med.* **3**, 1271–1274 (1997).
23. Pećina-Slaus, N. Tumor suppressor gene E-cadherin and its role in normal and malignant cells. *Cancer Cell Int.* **3**, 17 (2003).
24. Maris, J. M. Recent Advances in Neuroblastoma. *N. Engl. J. Med.* **362**, 2202–2211

(2010).
25. Maris, J. M., Hogarty, M. D., Bagatell, R. & Cohn, S. L. Neuroblastoma. *Lancet* **369**, 2106–2120 (2007).
26. Tsubota, S. & Kadomatsu, K. Origin and mechanism of neuroblastoma. *Oncoscience* **4**, 70–72 (2017).
27. Jiang, M., Stanke, J. & Lahti, J. M. The connections between neural crest development and neuroblastoma. *Curr. Top. Dev. Biol.* **94**, 77–127 (2011).
28. Marshall, G. M. *et al.* The prenatal origins of cancer. *Nat. Rev. Cancer* **14**, 277–289 (2014).
29. Louis, C. U. & Shohet, J. M. Neuroblastoma: molecular pathogenesis and therapy. *Annu. Rev. Med.* **66**, 49–63 (2015).
30. Furlan, A. *et al.* Multipotent peripheral glial cells generate neuroendocrine cells of the adrenal medulla. *Science* **357**, eaal3753 (2017).
31. Papaioannou, G. & McHugh, K. Neuroblastoma in childhood: review and radiological findings. *Cancer Imaging* **5**, 116–127 (2005).
32. Brodeur, G. M. *et al.* International criteria for diagnosis, staging, and response to treatment in patients with neuroblastoma. *J. Clin. Oncol.* **6**, 1874–1881 (1988).
33. Brodeur, G. M. *et al.* Revisions of the international criteria for neuroblastoma diagnosis, staging, and response to treatment. *J. Clin. Oncol.* **11**, 1466–1477 (1993).
34. Sharp, S. E., Gelfand, M. J. & Shulkin, B. L. Pediatrics: Diagnosis of Neuroblastoma. *Semin. Nucl. Med.* **41**, 345–353 (2011).
35. Monclair, T. *et al.* The International Neuroblastoma Risk Group (INRG) staging system: an INRG Task Force report. *J. Clin. Oncol.* **27**, 298–303 (2009).
36. Pinto, N. R. *et al.* Advances in Risk Classification and Treatment Strategies for Neuroblastoma. *J. Clin. Oncol.* **33**, 3008–3017 (2015).
37. Rubie, H. *et al.* Excellent Outcome With Reduced Treatment in Infants With Nonmetastatic and Unresectable Neuroblastoma Without MYCN Amplification: Results of the Prospective INES 99.1. *J. Clin. Oncol.* **29**, 449–455 (2010).
38. De Bernardi, B. *et al.* Excellent Outcome With Reduced Treatment for Infants With Disseminated Neuroblastoma Without MYCN Gene Amplification. *J. Clin. Oncol.* **27**, 1034–1040 (2009).
39. Brossard, J., Bernstein, M. L. & Lemieux, B. Neuroblastoma: an enigmatic disease. *Br. Med. Bull.* **52**, 787–801 (1996).
40. Kohl, N. E. *et al.* Transposition and amplification of oncogene-related sequences in human neuroblastomas. *Cell* **35**, 359–367 (1983).
41. Kohl, N. E., Gee, C. E. & Alt, F. W. Activated expression of the N-myc gene in human neuroblastomas and related tumors. *Science (80-.).* **226**, 1335 LP – 1337 (1984).
42. Schwab, M., Westermann, F., Hero, B. & Berthold, F. Neuroblastoma: biology and molecular and chromosomal pathology. *Lancet Oncol.* **4**, 472–480 (2003).
43. Huang, M. & Weiss, W. A. Neuroblastoma and MYCN. *Cold Spring Harb. Perspect. Med.* **3**, a014415–a014415 (2013).
44. Hurlin, P. J. N-myc functions in transcription and development. *Birth Defects Res. Part C Embryo Today Rev.* **75**, 340–352 (2005).
45. Ruiz-Pérez, M. V., Henley, A. B. & Arsenian-Henriksson, M. The MYCN Protein in Health and Disease. *Genes (Basel).* **8**, 113 (2017).
46. Beltran, H. The N-myc Oncogene: Maximizing its Targets, Regulation, and Therapeutic Potential. *Mol. Cancer Res.* **12**, 815 LP – 822 (2014).

47. Amati, B. *et al.* Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. *Nature* **359**, 423–426 (1992).
48. Eilers, M. & Eisenman, R. N. Myc's broad reach. *Genes Dev.* **22**, 2755–2766 (2008).
49. Grandori, C., Cowley, S. M., James, L. P. & Eisenman, R. N. The Myc/Max/Mad Network and the Transcriptional Control of Cell Behavior. *Annu. Rev. Cell Dev. Biol.* **16**, 653–699 (2000).
50. Zimmerman, M. W. *et al.* MYC Drives a Subset of High-Risk Pediatric Neuroblastomas and Is Activated through Mechanisms Including Enhancer Hijacking and Focal Enhancer Amplification. *Cancer Discov.* **8**, 320 LP – 335 (2018).
51. Brodeur, G. M. *et al.* Cytogenetic Features of Human Neuroblastomas and Cell Lines. *Cancer Res.* **41**, 4678 LP – 4686 (1981).
52. White, P. S. *et al.* Detailed molecular analysis of 1p36 in neuroblastoma. *Med. Pediatr. Oncol.* **36**, 37–41 (2001).
53. Kolla, V., Zhuang, T., Higashi, M., Naraparaju, K. & Brodeur, G. M. Role of CHD5 in Human Cancers: 10 Years Later. *Cancer Res.* **74**, 652 LP – 658 (2014).
54. Thompson, P. M., Gotoh, T., Kok, M., White, P. S. & Brodeur, G. M. CHD5, a new member of the chromodomain gene family, is preferentially expressed in the nervous system. *Oncogene* **22**, 1002 (2003).
55. Yeh, I.-T. *et al.* A germline mutation of the KIF1Bβ gene on 1p36 in a family with neural and nonneural tumors. *Hum. Genet.* **124**, 279–285 (2008).
56. Attiyeh, E. F. *et al.* Chromosome 1p and 11q Deletions and Outcome in Neuroblastoma. *N. Engl. J. Med.* **353**, 2243–2253 (2005).
57. Mandriota, S. J. *et al.* Ataxia-telangiectasia mutated (ATM) silencing promotes neuroblastoma progression through a MYCN independent mechanism. *Oncotarget* **6**, 18558–18576 (2015).
58. Plantaz, D. *et al.* Comparative genomic hybridization (CGH) analysis of stage 4 neuroblastoma reveals high frequency of 11q deletion in tumors lacking MYCN amplification. *Int. J. Cancer* **91**, 680–686 (2001).
59. Bown, N. *et al.* Gain of Chromosome Arm 17q and Adverse Outcome in Patients with Neuroblastoma. *N. Engl. J. Med.* **340**, 1954–1961 (1999).
60. Janoueix-Lerosey, I. *et al.* Molecular analysis of chromosome arm 17q gain in neuroblastoma. *Genes, Chromosom. Cancer* **28**, 276–284 (2000).
61. Tonini, G. P. Growth, progression and chromosome instability of Neuroblastoma: a new scenario of tumorigenesis? *BMC Cancer* **17**, 20 (2017).
62. Brodeur, G. M. NEUROBLASTOMA : BIOLOGICAL INSIGHTS INTO A CLINICAL ENIGMA. *Nat. Rev. Cancer* **3**, 203–216 (2003).
63. Kaneko, Y. & Knudson, A. G. Mechanism and relevance of ploidy in neuroblastoma. *Genes, Chromosom. Cancer* **29**, 89–95 (2000).
64. George, R. E. *et al.* Activating mutations in ALK provide a therapeutic target in neuroblastoma. *Nature* **455**, 975–978 (2008).
65. Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nat. Genet.* **45**, 279 (2013).
66. Hallberg, B. & Palmer, R. H. The role of the ALK receptor in cancer biology. *Ann. Oncol.* **27**, iii4–iii15 (2016).
67. Ogawa, S., Takita, J., Sanada, M. & Hayashi, Y. Oncogenic mutations of ALK in neuroblastoma. *Cancer Sci.* **102**, 302–308 (2011).
68. Carpenter, E. L. & Mossé, Y. P. Targeting ALK in neuroblastoma--preclinical and clinical

advancements. *Nat. Rev. Clin. Oncol.* **9**, 391–399 (2012).
69. Barone, G., Anderson, J., Pearson, A. D. J., Petrie, K. & Chesler, L. New strategies in neuroblastoma: Therapeutic targeting of MYCN and ALK. *Clin. Cancer Res.* **19**, 5814–5821 (2013).
70. Mossé, Y. P. *et al.* Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature* **455**, 930–935 (2008).
71. Kushner, B. H., Gilbert, F. & Helson, L. Familial neuroblastoma. Case reports, literature review, and etiologic considerations. *Cancer* **57**, 1887–1893 (1986).
72. Knudson Jr, A. G. & Strong, L. C. Mutation and cancer: neuroblastoma and pheochromocytoma. *Am. J. Hum. Genet.* **24**, 514–532 (1972).
73. Trochet, D. *et al.* Germline Mutations of the Paired–Like Homeobox 2B (PHOX2B) Gene in Neuroblastoma. *Am. J. Hum. Genet.* **74**, 761–764 (2004).
74. Mosse, Y. P. *et al.* Germline PHOX2B Mutation in Hereditary Neuroblastoma. *Am. J. Hum. Genet.* **75**, 727–730 (2004).
75. Raabe, E. H. *et al.* Prevalence and functional consequence of PHOX2B mutations in neuroblastoma. *Oncogene* **27**, 469 (2007).
76. Barr, E. K. & Applebaum, M. A. Genetic Predisposition to Neuroblastoma. *Child. (Basel, Switzerland)* **5**, 119 (2018).
77. Sausen, M. *et al.* Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. *Nat. Genet.* **45**, 12–17 (2013).
78. Cheung, N.-K. V *et al.* Association of age at diagnosis and genetic mutations in patients with neuroblastoma. *JAMA* **307**, 1062–1071 (2012).
79. Bentires-Alj, M. *et al.* Activating Mutations of the Noonan Syndrome-Associated SHP2/PTPN11 Gene in Human Solid Tumors and Adult Acute Myelogenous Leukemia. *Cancer Res.* **64**, 8816 LP – 8820 (2004).
80. Zhang, X. *et al.* Critical Role for GAB2 in Neuroblastoma Pathogenesis through the Promotion of SHP2/MYCN Cooperation. *Cell Rep.* **18**, 2932–2942 (2017).
81. Peifer, M. *et al.* Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **526**, 700 (2015).
82. Molenaar, J. J. *et al.* Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature* **483**, 589 (2012).
83. Ackermann, S. *et al.* A mechanistic classification of clinical phenotypes in neuroblastoma. *Science (80-.).* **362**, 1165 LP – 1170 (2018).
84. Boeva, V. *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* **49**, 1408 (2017).
85. van Groningen, T. *et al.* Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat. Genet.* **49**, 1261 (2017).
86. Garraway, L. A. & Sellers, W. R. Lineage dependency and lineage-survival oncogenes in human cancer. *Nat. Rev. Cancer* **6**, 593–602 (2006).
87. Pei, D. *et al.* Distinct neuroblastoma-associated alterations of PHOX2B impair sympathetic neuronal differentiation in zebrafish models. *PLoS Genet.* **9**, e1003533–e1003533 (2013).
88. Yang, L. *et al.* PHOX2B Is Associated with Neuroblastoma Cell Differentiation. *Cancer Biother. Radiopharm.* **31**, 44–51 (2016).
89. Stutterheim, J. *et al.* PHOX2B Is a Novel and Specific Marker for Minimal Residual

Disease Testing in Neuroblastoma. *J. Clin. Oncol.* **26**, 5443–5449 (2008).
90. The ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science (80-.).* **306**, 636 LP – 640 (2004).
91. (DGT), T. F. C. and the R. P. M. I. and C. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462 (2014).
92. Fang, Y. & Fullwood, M. J. Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics. Proteomics Bioinformatics* **14**, 42–54 (2016).
93. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
94. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
95. Pachnis, V., Belayew, A. & Tilghman, S. M. Locus unlinked to alpha-fetoprotein under the control of the murine raf and Rif genes. *Proc. Natl. Acad. Sci.* **81**, 5523 LP – 5527 (1984).
96. Bartolomei, M. S., Zemel, S. & Tilghman, S. M. Parental imprinting of the mouse H19 gene. *Nature* **351**, 153–155 (1991).
97. Barlow, D. P., Stöger, R., Herrmann, B. G., Saito, K. & Schweifer, N. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature* **349**, 84–87 (1991).
98. Jarroux, J., Morillon, A. & Pinskaya, M. History, Discovery, and Classification of lncRNAs BT - Long Non Coding RNA Biology. in (ed. Rao, M. R. S.) 1–46 (Springer Singapore, 2017). doi:10.1007/978-981-10-5203-3_1
99. Nordin, M., Bergman, D., Halje, M., Engström, W. & Ward, A. Epigenetic regulation of the Igf2/H19 gene cluster. *Cell Prolif.* **47**, 189–199 (2014).
100. Venter, J. C. *et al.* The Sequence of the Human Genome. *Science (80-.).* **291**, 1304 LP – 1351 (2001).
101. Consortium, I. H. G. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860 (2001).
102. Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11995–11999 (1993).
103. Carninci, P. *et al.* The Transcriptional Landscape of the Mammalian Genome. *Science (80-.).* **309**, 1559 LP – 1563 (2005).
104. Sun, M. & Kraus, W. L. From discovery to function: the expanding roles of long noncoding RNAs in physiology and disease. *Endocr. Rev.* **36**, 25–64 (2015).
105. Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS One* **9**, e78644 (2014).
106. Rai, M. F., Tycksen, E. D., Sandell, L. J. & Brophy, R. H. Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling of anterior cruciate ligament tears. *J. Orthop. Res.* **36**, 484–497 (2018).
107. Tombácz, D., Balázs, Z., Csabai, Z., Snyder, M. & Boldogkői, Z. Long-Read Sequencing Revealed an Extensive Transcript Complexity in Herpesviruses . *Frontiers in Genetics* **9**, 259 (2018).
108. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
109. Kanamori-Katayama, M. *et al.* Unamplified cap analysis of gene expression on a single-molecule sequencer. *Genome Res.* **21**, 1150–1159 (2011).
110. Ma, L., Bajic, V. B. & Zhang, Z. On the classification of long non-coding RNAs. *RNA Biol.*

10, 924–933 (2013).
111. Mattick, J. S. & Rinn, J. L. Discovery and annotation of long noncoding RNAs. *Nat. Struct. &Amp; Mol. Biol.* **22**, 5 (2015).
112. Paul, I. J. & Duerksen, J. D. Chromatin-associated RNA content of heterochromatin and euchromatin. *Mol. Cell. Biochem.* **9**, 9–16 (1975).
113. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
114. Kornberg, R. D. & Lorch, Y. Chromatin-modifying and -remodeling complexes. *Curr. Opin. Genet. Dev.* **9**, 148–151 (1999).
115. Lodish, H. *et al. Molecular cell biology*. (Freeman, W. H., 2000).
116. Aagaard-Tillery, K. M., Suter, M. A., Harris, A., Abramovici, A. J. C. & Cantu, J. Epigenetics and reproduction and the developmental origins of health and disease. *Anim Reprod* **7**, 103–116 (2010).
117. Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43**, 904–914 (2011).
118. Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A. & Panning, B. Xist RNA and the Mechanism of X Chromosome Inactivation. *Annu. Rev. Genet.* **36**, 233–278 (2002).
119. 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**, 527–542 (1992).
120. 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).
121. Davidovich, C. & Cech, T. R. The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. *RNA* **21**, 2007–2022 (2015).
122. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
123. 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).
124. Plath, K. *et al.* Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science (80-.).* **300**, 131 LP – 135 (2003).
125. Pandey, R. R. *et al.* Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. *Mol. Cell* **32**, 232–246 (2008).
126. 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**, 1956–1962 (2011).
127. Mancini-Dinardo, D., Steele, S. J. S., Levorse, J. M., Ingram, R. S. & Tilghman, S. M. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev.* **20**, 1268–1282 (2006).
128. Yang, Y. W. *et al.* Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *Elife* **3**, e02046–e02046 (2014).
129. Hsieh, C.-L. *et al.* Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7319–7324 (2014).
130. Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* **43**, 621–629 (2011).
131. Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci.*

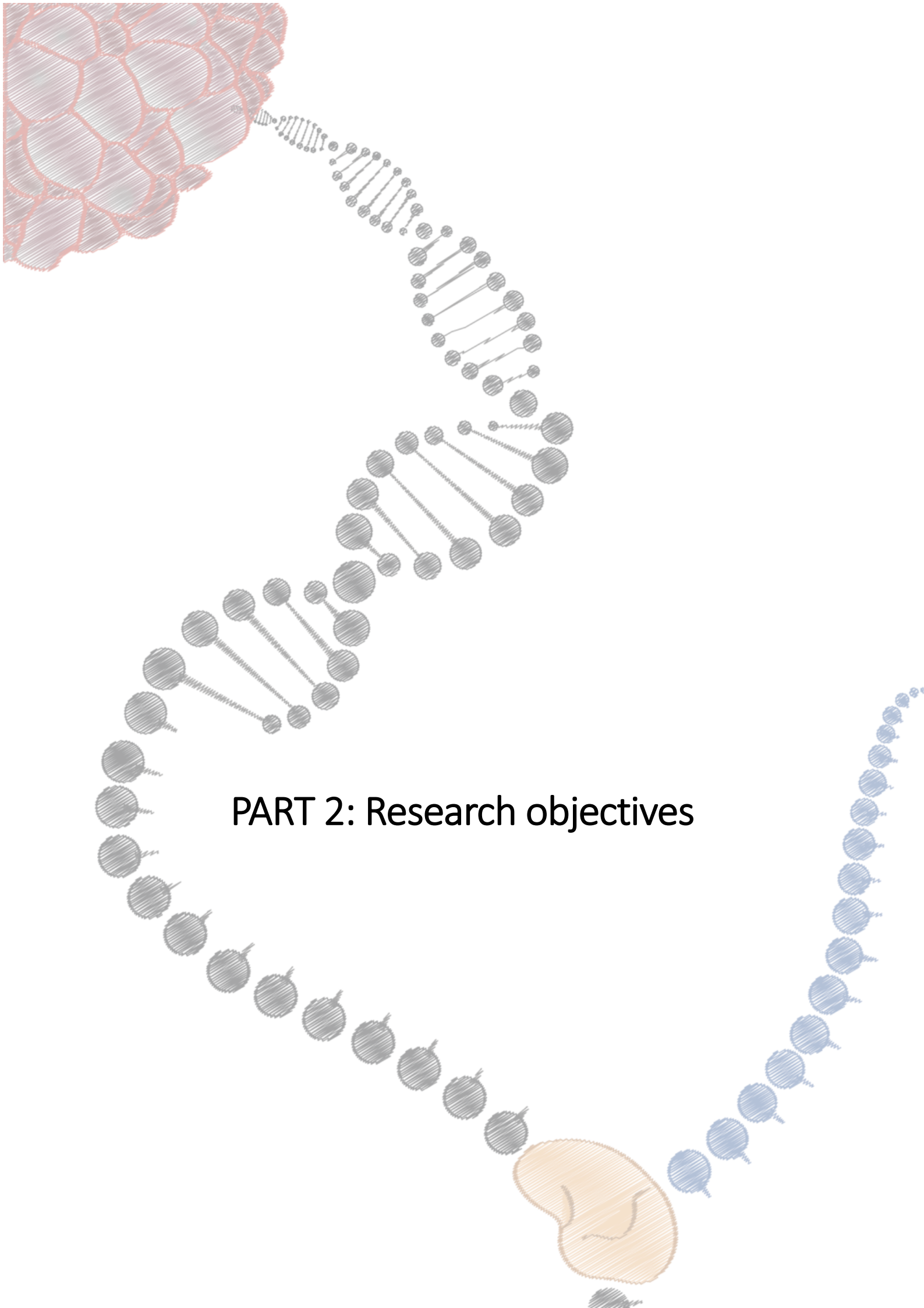
Signal. **3**, ra8–ra8 (2010).
132. Long, Y., Wang, X., Youmans, D. T. & Cech, T. R. How do lncRNAs regulate transcription? *Sci. Adv.* **3**, eaao2110–eaao2110 (2017).
133. Chalei, V. *et al.* The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. *Elife* **3**, e04530–e04530 (2014).
134. Di Ruscio, A. *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371–376 (2013).
135. He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5**, 522 (2004).
136. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
137. Khorkova, O., Myers, A. J., Hsiao, J. & Wahlestedt, C. Natural antisense transcripts. *Hum. Mol. Genet.* **23**, R54–R63 (2014).
138. Faghihi, M. A. & Wahlestedt, C. Regulatory roles of natural antisense transcripts. *Nat. Rev. Mol. Cell Biol.* **10**, 637–643 (2009).
139. Yin, J. *et al.* UXT-AS1-induced alternative splicing of UXT is associated with tumor progression in colorectal cancer. *Am. J. Cancer Res.* **7**, 462–472 (2017).
140. Amaral, P. P. *et al.* Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. *Genome Biol.* **19**, 32 (2018).
141. Costa, R. H., Kalinichenko, V. V, Holterman, A.-X. L. & Wang, X. Transcription factors in liver development, differentiation, and regeneration. *Hepatology* **38**, 1331–1347 (2003).
142. Tan, L., Yu, J.-T., Hu, N. & Tan, L. Non-coding RNAs in Alzheimer's Disease. *Mol. Neurobiol.* **47**, 382–393 (2013).
143. Shizuka, U. & Stefanie, D. Long Noncoding RNAs in Cardiovascular Diseases. *Circ. Res.* **116**, 737–750 (2015).
144. Austin, P. J. *et al.* Transcriptional profiling identifies the long noncoding RNA plasmacytoma variant translocation (PVT1) as a novel regulator of the asthmatic phenotype in human airway smooth muscle. *J. Allergy Clin. Immunol.* **139**, 780–789 (2017).
145. De Smet, E. G., Mestdagh, P., Vandesompele, J., Brusselle, G. G. & Bracke, K. R. Non-coding RNAs in the pathogenesis of COPD. *Thorax* **70**, 782 LP – 791 (2015).
146. Gutschner, T. & Diederichs, S. The hallmarks of cancer. *RNA Biol.* **9**, 703–719 (2012).
147. McKenna, N. J., Lanz, R. B. & O'Malley, B. W. Nuclear Receptor Coregulators: Cellular and Molecular Biology*. *Endocr. Rev.* **20**, 321–344 (1999).
148. Lanz, R. B. *et al.* A Steroid Receptor Coactivator, SRA, Functions as an RNA and Is Present in an SRC-1 Complex. *Cell* **97**, 17–27 (1999).
149. Leygue, E., Dotzlaw, H., Watson, P. H. & Murphy, L. C. Expression of the Steroid Receptor RNA Activator in Human Breast Tumors. *Cancer Res.* **59**, 4190 LP – 4193 (1999).
150. 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**, 742–749 (2011).
151. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).
152. Balik, V. *et al.* MEG3: a novel long noncoding potentially tumour-suppressing RNA in meningiomas. *J. Neurooncol.* **112**, 1–8 (2013).

153. 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**, 662–674 (2010).
154. Mourtada-Maarabouni, M., Pickard, M. R., Hedge, V. L., Farzaneh, F. & Williams, G. T. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* **28**, 195 (2008).
155. Mourtada-Maarabouni, M., Hasan, A. M., Farzaneh, F. & Williams, G. T. Inhibition of human T-cell proliferation by mammalian target of rapamycin (mTOR) antagonists requires noncoding RNA growth-arrest-specific transcript 5 (GAS5). *Mol. Pharmacol.* **78**, 19–28 (2010).
156. Morrison, L. E. *et al.* Effects of ERBB2 amplicon size and genomic alterations of chromosomes 1, 3, and 10 on patient response to trastuzumab in metastatic breast cancer. *Genes, Chromosom. Cancer* **46**, 397–405 (2007).
157. Smedley, D. *et al.* Characterization of chromosome 1 abnormalities in malignant melanomas. *Genes, Chromosom. Cancer* **28**, 121–125 (2000).
158. Phatak, P. & Burger, A. M. Telomerase and its potential for therapeutic intervention. *Br. J. Pharmacol.* **152**, 1003–1011 (2007).
159. Andersson, S. *et al.* Frequent gain of the human telomerase gene TERC at 3q26 in cervical adenocarcinomas. *Br. J. Cancer* **95**, 331–338 (2006).
160. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. Telomeric Repeat–Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. *Science (80-.).* **318**, 798 LP – 801 (2007).
161. Redon, S., Reichenbach, P. & Lingner, J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res.* **38**, 5797–5806 (2010).
162. Caslini, C. Transcriptional regulation of telomeric non-coding RNA: implications on telomere biology, replicative senescence and cancer. *RNA Biol.* **7**, 18–22 (2010).
163. Ji, P. *et al.* MALAT-1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **22**, 8031 (2003).
164. Hutchinson, J. N. *et al.* A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 39 (2007).
165. Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39**, 925–938 (2010).
166. Tano, K. *et al.* MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. *FEBS Lett.* **584**, 4575–4580 (2010).
167. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
168. 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**, 2119–2128 (2011).
169. Rossignol, F., Vaché, C. & Clottes, E. Natural antisense transcripts of hypoxia-inducible factor 1alpha are detected in different normal and tumour human tissues. *Gene* **299**, 135–140 (2002).
170. Uchida, T. *et al.* Prolonged Hypoxia Differentially Regulates Hypoxia-inducible Factor (HIF)-1α and HIF-2α Expression in Lung Epithelial Cells: IMPLICATION OF NATURAL ANTISENSE HIF-1α . *J. Biol. Chem.* **279**, 14871–14878 (2004).

171. 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**, 143–151 (1999).
172. Robb, G. B. *et al.* Post-transcriptional Regulation of Endothelial Nitric-oxide Synthase by an Overlapping Antisense mRNA Transcript. *J. Biol. Chem.* **279**, 37982–37996 (2004).
173. Fish, J. E. *et al.* Hypoxia-inducible Expression of a Natural cis-Antisense Transcript Inhibits Endothelial Nitric-oxide Synthase. *J. Biol. Chem.* **282**, 15652–15666 (2007).
174. Srikantan, V. *et al.* PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 12216–12221 (2000).
175. Tsang, W. P., Wong, T. W. L., Cheung, A. H. H., Co, C. N. N. & Kwok, T. T. Induction of drug resistance and transformation in human cancer cells by the noncoding RNA CUDR. *RNA* **13**, 890–898 (2007).
176. Pandey, G. K. & Kanduri, C. Long noncoding RNAs and neuroblastoma. *Oncotarget* **6**, 18265–18275 (2015).
177. Mazar, J., Rosado, A., Shelley, J., Marchica, J. & Westmoreland, T. J. The long non-coding RNA GAS5 differentially regulates cell cycle arrest and apoptosis through activation of BRCA1 and p53 in human neuroblastoma. *Oncotarget* **8**, 6589–6607 (2016).
178. Russell, M. R. *et al.* CASC15-S Is a Tumor Suppressor lncRNA at the 6p22 Neuroblastoma Susceptibility Locus. *Cancer Res.* **75**, 3155–3166 (2015).
179. Liu, P. Y. *et al.* Effects of a Novel Long Noncoding RNA, lncUSMycN, on N-Myc Expression and Neuroblastoma Progression. *JNCI J. Natl. Cancer Inst.* **106**, dju113–dju113 (2014).
180. Atmadibrata, B. *et al.* The Novel Long Noncoding RNA linc00467 Promotes Cell Survival but Is Down-Regulated by N-Myc. *PLoS One* **9**, e88112 (2014).
181. Pandey, G. K. *et al.* The Risk-Associated Long Noncoding RNA NBAT-1 Controls Neuroblastoma Progression by Regulating Cell Proliferation and Neuronal Differentiation. *Cancer Cell* **26**, 722–737 (2014).
182. Yang, T.-W. *et al.* RNA-Binding Proteomics Reveals MATR3 Interacting with lncRNA SNHG1 To Enhance Neuroblastoma Progression. *J. Proteome Res.* **18**, 406–416 (2019).
183. Yu, M.; Ohira, M.; Li, Y.; Niizuma, H.; Oo, M. L.; Zhu, Y.; Ozaki, T.; Isogai, E.; Nakamura, Y.; Koda, T.; Oba, S.; Yu, B.; Nakagawara, A. High expression of ncRAN, a novel non-coding RNA mapped to chromosome 17q25.1, is associated with poor prognosis in neuroblastoma. *Int. J. Oncol.* **34**, 931–938 (2009).
184. Buechner, J. & Einvik, C. N-myc and Noncoding RNAs in Neuroblastoma. *Mol. Cancer Res.* **10**, 1243 LP – 1253 (2012).
185. Mestdagh, P. *et al.* An integrative genomics screen uncovers ncRNA T-UCR functions in neuroblastoma tumours. *Oncogene* **29**, 3583 (2010).
186. Miller, C. M. & Harris, E. N. Antisense Oligonucleotides: Treatment Strategies and Cellular Internalization. *RNA Dis. (Houston, Tex.)* **3**, e1393 (2016).
187. Wu, H., Lima, W. F. & Crooke, S. T. Properties of Cloned and Expressed Human RNase H1. *J. Biol. Chem.* **274**, 28270–28278 (1999).
188. Verma, A. Recent Advances in Antisense Oligonucleotide Therapy in Genetic Neuromuscular Diseases. *Ann. Indian Acad. Neurol.* **21**, 3–8 (2018).
189. Dias, N. & Stein, C. A. Antisense Oligonucleotides: Basic Concepts and Mechanisms. *Mol. Cancer Ther.* **1**, 347 LP – 355 (2002).

190. Prakash, T. P. *et al.* Antisense Oligonucleotides Containing Conformationally Constrained 2′,4′-(N-Methoxy)aminomethylene and 2′,4′-Aminooxymethylene and 2′-O,4′-C-Aminomethylene Bridged Nucleoside Analogues Show Improved Potency in Animal Models. *J. Med. Chem.* **53**, 1636–1650 (2010).
191. Shen, X. & Corey, D. R. Chemistry, mechanism and clinical status of antisense oligonucleotides and duplex RNAs. *Nucleic Acids Res.* **46**, 1584–1600 (2017).
192. Braasch, D. A. & Corey, D. R. Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chem. Biol.* **8**, 1–7 (2001).
193. Gooderham, N. J. *et al.* In silico and in vitro evaluation of exonic and intronic off-target effects form a critical element of therapeutic ASO gapmer optimization. *Nucleic Acids Res.* **43**, 8638–8650 (2015).
194. Juliano, R. L. The delivery of therapeutic oligonucleotides. *Nucleic Acids Res.* **44**, 6518–6548 (2016).
195. Stein, C. A. & Castanotto, D. FDA-Approved Oligonucleotide Therapies in 2017. *Mol. Ther.* **25**, 1069–1075 (2017).
196. Geary, R. S., Henry, S. P. & Grillone, L. R. Fomivirsen. *Clin. Pharmacokinet.* **41**, 255–260 (2002).
197. Guo, S. *et al.* Suppressing transthyretin production in mice, monkeys and humans using 2nd-Generation antisense oligonucleotides AU - Ackermann, Elizabeth J. *Amyloid* **23**, 148–157 (2016).
198. Crooke, R. M. *et al.* An apolipoprotein B antisense oligonucleotide lowers LDL cholesterol in hyperlipidemic mice without causing hepatic steatosis. *J. Lipid Res.* **46**, 872–884 (2005).
199. Mendell, J. R. *et al.* Longitudinal effect of eteplirsen versus historical control on ambulation in Duchenne muscular dystrophy. *Ann. Neurol.* **79**, 257–271 (2016).
200. Stein, C. A. Eteplirsen Approved for Duchenne Muscular Dystrophy: The FDA Faces a Difficult Choice. *Mol. Ther.* **24**, 1884–1885 (2016).
201. Hua, Y. *et al.* Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes Dev.* **24**, 1634–1644 (2010).
202. Gouni-Berthold, I. The role of antisense oligonucleotide therapy against apolipoprotein-CIII in hypertriglyceridemia. *Atheroscler. Suppl.* **30**, 19–27 (2017).
203. Dana, H. *et al.* Molecular Mechanisms and Biological Functions of siRNA. *Int. J. Biomed. Sci.* **13**, 48–57 (2017).
204. Leuschner, P. J. F., Ameres, S. L., Kueng, S. & Martinez, J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep.* **7**, 314–320 (2006).
205. Jinek, M. & Doudna, J. A. A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457**, 405 (2008).
206. Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
207. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21**, 635 (2003).
208. Jacobsen, L. B., Calvin, S. A. & Lobenhofer, E. K. Transcriptional effects of transfection: the potential for misinterpretation of gene expression data generated from transiently transfected cells. *Biotechniques* **47**, 617–624 (2009).
209. Grimm, D. *et al.* Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441**, 537 (2006).

210. Wang, J., Lu, Z., Wientjes, M. G. & Au, J. L.-S. Delivery of siRNA therapeutics: barriers and carriers. *AAPS J.* **12**, 492–503 (2010).
211. Hoy, S. M. Patisiran: First global approval. *Drugs* **78**, 1625–1631 (2018).
212. Adams, D. *et al.* Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis. *N. Engl. J. Med.* **379**, 11–21 (2018).
213. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).
214. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180–2196 (2013).
215. Sheridan, C. Go-ahead for first in-body CRISPR medicine testing. (2018). Available at: https://www.nature.com/articles/d41587-018-00003-2. (Accessed: 3rd June 2019)
216. Baylis, F. & McLeod, M. First-in-human Phase 1 CRISPR Gene Editing Cancer Trials: Are We Ready? *Curr. Gene Ther.* **17**, 309–319 (2017).



PART 2: Research objectives

2.1 Research objectives

Although advances have been made in the understanding of neuroblastoma, this malignancy still remains an enigma. Research into the genetic background progresses at a slow pace, mainly focusing on the protein coding part of the genome. Nevertheless, the search for diagnostic biomarkers and personalized therapeutic targets still continues to provide better prognoses and survival chances for patients categorized in the high-risk population.

Long non-coding RNAs could provide new and alternative insights into the phenotypical heterogeneity of neuroblastoma. Despite the fact that the advent of lncRNA research is rather recent, developments in the field are evolving at a fast pace and a growing number of lncRNAs implicated in multiple biological processes and diseases - such as cancer - are being reported. This new class of non-coding molecules might prove to be the missing link to further increase the survival chances of neuroblastoma patients.

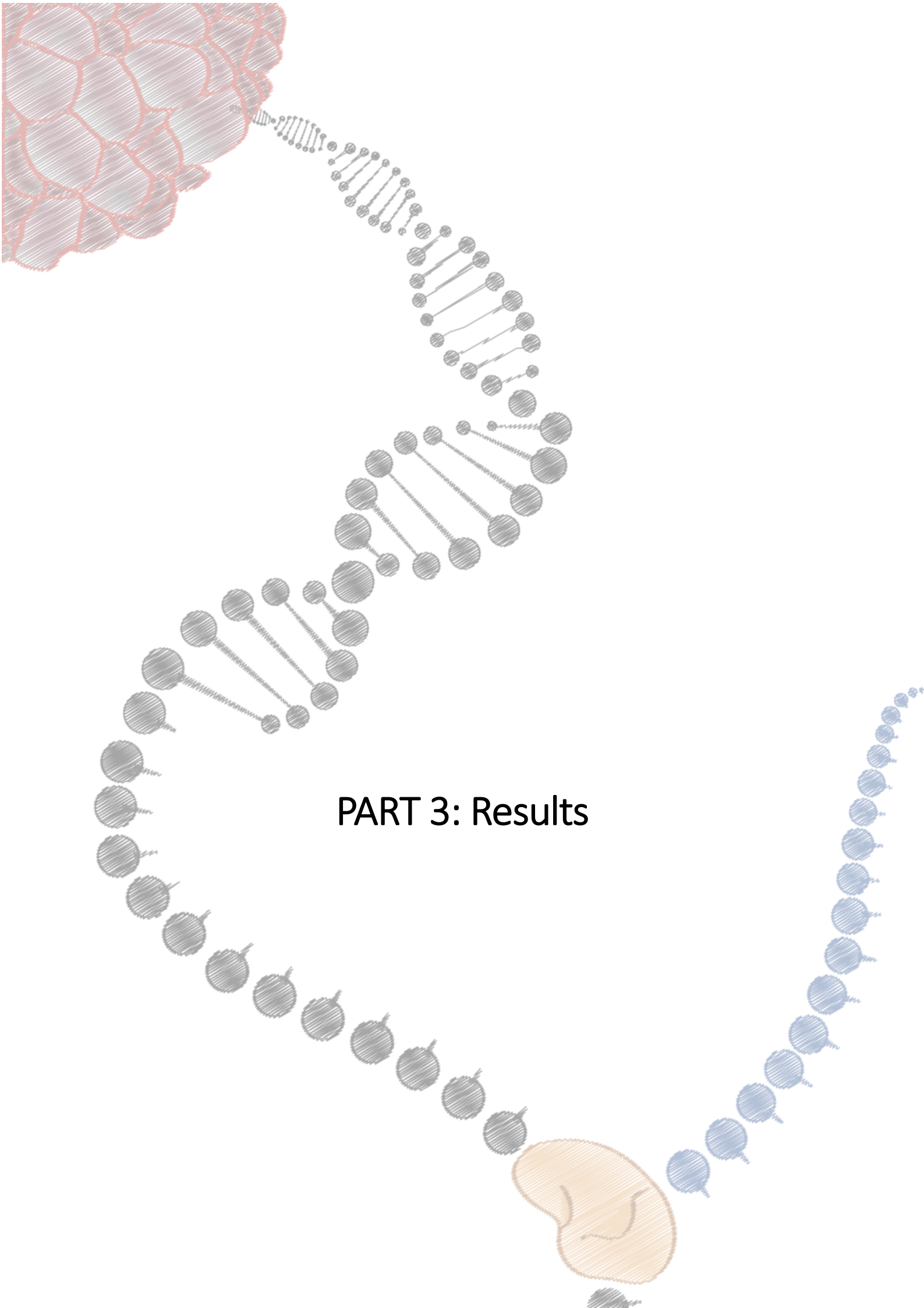
Aim 1: Unraveling the lincRNome of neuroblastoma in association with key neuroblastoma genes

To address this issue, we aimed to investigate the lincRNome in neuroblastoma and their associations with known key neuroblastoma genes (**research paper 1**). By comparing lincRNAs expressed in neuroblastoma primary tumors and neuroblasts - the presumed precursor cells of neuroblastoma - we intended to identify differences in transcriptional activity. Subsequently, a comprehensive correlation analysis between lincRNA expression profiles and genomic alterations important in neuroblastoma initiation and development, led to further insights in the role of lincRNAs in this malignancy. Based on the results of the differential expression analysis and the cellular perturbation models, we wished to generate a prioritized list of uncharacterized potentially oncogenic lincRNAs. Furthermore, we aimed to identify lincRNAs having a modulating or regulatory effect on these key driver oncogenes.

Aim 2: Functional validation of a candidate lncRNA implicated in neuroblastoma tumorigenesis

Here, we aimed to functionally validate the top candidate of the lincRNA list generated in aim 1 (**research paper 2**). To study the functionality of this lincRNA, coined *NESPR*, in

neuroblastoma tumorigenesis and its role in the network controlling neuroblastoma cell identity, knockdown experiments were performed in neuroblastoma cell lines. In addition, to unravel the functional mechanism of *NESPR* we focused on its associations with protein coding genes and its role in insulated neighborhood formation.



PART 3: Results

3.1 **Research paper 1:** Integrative analysis identifies lincRNAs up- and downstream of key neuroblastoma driver genes

Dries Rombaut, Hua-Sheng Chiu, Bieke Decaesteker, Celine Everaert, Nurten Yigit, Agathe Peltier, Isabelle Janoueix-Lerosey, Christoph Bartenhagen, Matthias Fischer, Stephen Roberts, Nicky D'Haene, Katleen De Preter, Frank Speleman, Geertrui Denecker, Pavel Sumazin, Jo Vandesompele, Steve Lefever, Pieter Mestdagh

Scientific Reports 9, Article number: 5685 (2019)
https://doi.org/10.1038/s41598-019-42107-y

Contributions: The candidate contributed in whole to design of the work, data acquisition and interpretation. Aside from the LongHorn analysis, the candidate performed the rest of the data analysis. For the matching paragraphs in the manuscript, the candidate contributed to the interpretation of the data and the revision of the text. The rest of the manuscript was drafted and revised by the candidate.

Integrative analysis identifies lincRNAs up- and downstream of neuroblastoma driver genes

Dries Rombaut[1,2], Hua-Sheng Chiu[3], Bieke Decaesteker[1,2], Celine Everaert[1,2], Nurten Yigit[1,2], Agathe Peltier[4,5], Isabelle Janoueix-Lerosey[4,5], Christoph Bartenhagen[7], Matthias Fischer[6,7], Stephen Roberts[8], Nicky D'Haene[9], Katleen De Preter[1,2], Frank Speleman[1,2], Geertrui Denecker[1,2], Pavel Sumazin[3], Jo Vandesompele[1,2], Steve Lefever[1,2], Pieter Mestdagh[*1,2]

1. Center for Medical Genetics, Ghent University, Ghent, 9000, Belgium
2. Cancer Research Institute Ghent (CRIG), Ghent, 9000, Belgium
3. Texas Children's Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA
4. Institut Curie, PSL Research University, Inserm U830, Equipe Labellisée contre le Cancer, F-75005, Paris, France.
5. SIREDO: Care, Innovation and Research for Children, Adolescents and Young Adults with Cancer, Institut Curie, F-75005, Paris, France.
6. Center for Molecular Medicine Cologne (CMMC), University of Cologne, 50931 Cologne, Germany
7. Department of Experimental Pediatric Oncology, University Children's Hospital of Cologne, Medical Faculty, University of Cologne, 50937 Cologne, Germany
8. Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY, United States of America
9. Hôpital Erasme, Cliniques Universitaires de Bruxelles, Bruxelles, 1070, Belgium

* Corresponding author: Pieter.Mestdagh@UGent.be

Abstract

Long intergenic non-coding RNAs (lincRNAs) are emerging as integral components of signaling pathways in various cancer types. In neuroblastoma, only a handful of lincRNAs are known as upstream regulators or downstream effectors of oncogenes. Here, we exploit RNA sequencing data of primary neuroblastoma tumors, neuroblast precursor cells, neuroblastoma cell lines and various cellular perturbation model systems to define the neuroblastoma lincRNome and map lincRNAs up- and downstream of neuroblastoma driver genes *MYCN*, *ALK* and *PHOX2B*. Each of these driver genes controls the expression of a particular subset of lincRNAs, several of which are associated with poor survival and are differentially expressed in neuroblastoma tumors compared to neuroblasts. By integrating RNA sequencing data from both primary tumor tissue and cancer cell lines, we demonstrate that several of these lincRNAs are expressed in stromal cells. Deconvolution of primary tumor gene expression data revealed a strong association between stromal cell composition and driver gene status, resulting in differential expression of these lincRNAs. We also explored lincRNAs that putatively act upstream of neuroblastoma driver genes, either as presumed modulators of driver gene

activity, or as modulators of effectors regulating driver gene expression. This analysis revealed strong associations between the neuroblastoma lincRNAs *MIAT* and *MEG3* and *MYCN* and *PHOX2B* activity or expression. Together, our results provide a comprehensive catalogue of the neuroblastoma lincRNome, highlighting lincRNAs up- and downstream of key neuroblastoma driver genes. This catalogue forms a solid basis for further functional validation of candidate neuroblastoma lincRNAs.

Introduction

During the past decade, detailed analysis of the human transcriptome revealed thousands of RNA molecules with no obvious coding potential[1–5]. These so-called long non-coding RNAs (lncRNAs) are poorly conserved at the sequence level and have a lower but more cell-type specific expression profile compared to protein coding mRNAs[1,6,7]. Based on functional studies of selected lncRNAs, it has become clear that they can act as important modulators of various processes in the cell, including chromatin conformation, transcription, splicing and post-transcriptional regulation[8–10]. Their capacity to interact with several bio-molecules in the cell (i.e. RNA, DNA and proteins) provides them with a plethora of mechanisms to exert their functions. Not surprisingly, deregulated expression of lncRNAs may cause human diseases including cancer[8,11–14]. At present, dozens of lncRNAs are known to function up- or downstream of cancer drivers or key signaling pathways. Notable examples are *TP53* pathway tumor suppressor lncRNAs *PANDAR* and *lincRNA-p21*[15,16] and oncogenic lncRNAs *CCAT2* and *MINCR* as respective activator or effector of *MYC*[17]. Systematic analysis of RNA-sequencing data across various adult tumor types further demonstrated that mutations in oncogenes and tumor suppressor genes can deregulate lncRNA expression[18]. Knockdown of lncRNAs driving oncogenic signaling can result in a therapeutic response *in vitro* and *in vivo*[17,19,20].

Neuroblastoma (NB) is one of the most enigmatic tumors, with clinical behavior ranging from spontaneous regression to metastatic disease refractory to aggressive multimodal therapy[21]. Tumors arise from neural crest-derived progenitor cells through deregulation of signaling pathways governing sympathetic nervous system development and differentiation. Despite the identification of predisposing mutations (*PHOX2B*[22]), somatic mutations (*ALK*[23,24]*, ATRX*[25]*)* amplifications (*MYCN*[26]) and translocations (TERT[27]) affecting several protein-coding genes, little progress was made in improving overall survival in the last decade. Unravelling the

lncRNA components involved in these signaling pathways could help us better understand how these networks are wired and may reveal novel regulatory mechanisms and, ultimately, therapeutic strategies. While only a handful of NB-associated lncRNAs have been identified today, for some, a putative therapeutic targeting potential has been demonstrated. *NBAT1*, a lncRNA on chromosome 6, possesses tumor suppressor functions, inhibiting NB proliferation and invasion and promoting differentiation[28]. Other examples of deregulated lncRNAs in NB are *ncRAN*, located on chromosome 17q and associated with poor prognosis, MYCN target gene *linc00467* that represses *DKK1* leading to increased NB cell survival and *lncUSMycN*, located on 2p and shown to regulate *MYCN* expression post-transcriptionally[29–31].

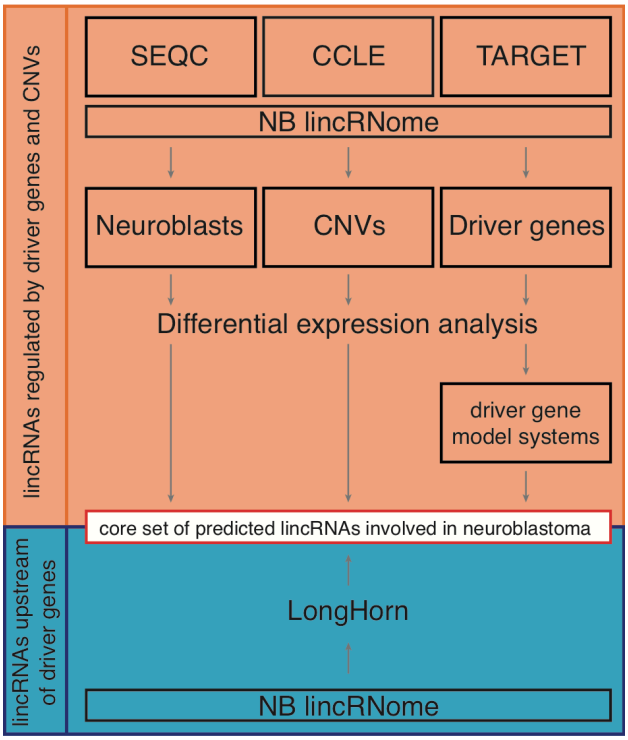


Figure 1: Included data sets and analyses in the study
The SEQC, CCLE and TARGET data sets were used to determine the NB lincRNome. The abundance of the robustly expressed lincRNAs in the lincRNome was used to compare expression between neuroblast and NB samples, primary tumor samples containing CNVs and samples without gains or deletions and NB tumors with and without mutations, amplifications or associations with NB driver genes. The regulation of lincRNAs correlated with NB driver genes was assessed in cellular perturbation models. To determine the involvement of lincRNAs in modulating the effect or regulation of these driver genes, we made use of a state-of-the-art algorithm called LongHorn. Combined, these analyses allow us to arrive at a core set of predicted NB associated lincRNAs.

In order to identify lncRNAs involved in key NB signaling pathways, we integrated RNA seq data from 497 primary NB tumor samples[32,33], human fetal neuroblasts and various model systems with perturbation of driver gene activity. We present a comprehensive view on the NB lincRNome and prioritize lncRNAs up- and downstream of well-established NB driver genes, such as *MYCN*, *PHOX2B* and *ALK* (Figure 1).

Results

The neuroblastoma lincRNome

To establish the NB lincRNome, we reanalyzed RNA seq data from 497 primary tumors, established by Zhang and colleagues as part of the SEQC study (further referred to as the SEQC dataset)[32,33]. Because the RNA sequencing data was unstranded (i.e. does not contain strand orientation information), we focused our analysis on lncRNAs classified as intergenic (lincRNAs) in Ensembl. This highly curated catalogue contains 7821 lincRNAs, of which 3295 were robustly expressed in NB tumors and hence define the NB lincRNome (Figure 2A). The NB lincRNome was further validated in an independent RNA seq dataset of 148 primary tumors, generated by the TARGET consortium[34], confirming expression for 3290 of the 3295 lincRNAs (Figure 2A). To asses independent transcription of the lncRNAs, we've integrated publicly available CAGE seq (Cap Analysis Gene Expression) data from NB cell lines. Through CAGE seq, the 5' end of a capped RNA molecule is sequenced, revealing the transcription start site of the transcript. By integrating this CAGE seq data using the Zipper plot[35], we uncovered an enrichment of CAGE peaks at lincRNA transcription start sites (TSS), with 255 lincRNAs having a CAGE peak within +/- 5 kb of their TSS (Figure 2B). CAGE TSS enrichment is less pronounced compared to protein coding mRNAs, likely due to the fact that lincRNAs are less abundant than mRNAs (Figure 2B) and public CAGE seq data is filtered based on a minimal expression cutoff[36]. Distribution of histone marks is more similar between lincRNAs and mRNAs, with 2933 and 2918 lincRNAs displaying a H3K4me3 or H3K27ac mark within +/- 5 kb of their TSS, respectively. Although the majority of lincRNAs is weakly expressed, several are highly abundant. Up to 20% of all lincRNA reads are consumed by only 5 highly abundant lincRNAs in each of the investigated datasets (Figure 2C). In contrast to adult cancers, where lincRNAs like *MALAT1*, *NEAT1* and *XIST* were most abundant[18], we identified *MEG3* and *MIAT*

as the most abundant lincRNAs in NB. While *MIAT* (also known as *GOMAFU*) has a dominant neuronal expression pattern[37,38], it promotes growth and proliferation of multiple cancer types[39–41]. Interestingly, several other uncharacterized lincRNAs were found among the most abundant in NB tumors, including *lnc-MEP1B-2* and *lnc-INAFM2-2*. Moreover, we found *MEG3* and *MIAT* to be quite specific for NB cells (Supplemental Figure 1). In fact, lincRNA expression patterns are known to exhibit remarkable cell-type specificity[1,6,7]. By comparing lincRNA expression in 917 cell lines representing 29 tumor types, we found NB among the tumor types with the highest number of specifically expressed lincRNAs (Figure 2D, E). As several tumor-type specific lincRNAs have been shown to play important roles in tumor biology[42–45], this information may thus help prioritize lincRNAs relevant for NB.

As tumor tissue biopsies to a variable degree are composed of stromal cells, we integrated the NB lincRNome with RNA seq data from NB cell lines to evaluate which lincRNAs are more likely to be derived from tumor or stromal cells. We found 2984 of 3295 lincRNAs from the NB lincRNome expressed in NB cell lines and 311 that were only detected in tumor biopsies (Figure 2F). To evaluate whether the latter fraction is indeed more likely to be stromal cell derived, we used lincRNA tissue specificity data from the FANTOM consortium to look for cell or tissue ontology enrichments[3]. Various stromal cell types, including CD8+ alpha/beta T-cells were enriched among lincRNAs with expression restricted to tumor biopsies (Figure 2G). These data suggest that a fraction of lincRNAs expressed in tumor tissue biopsies are indeed likely derived from stromal cells in these biopsies.

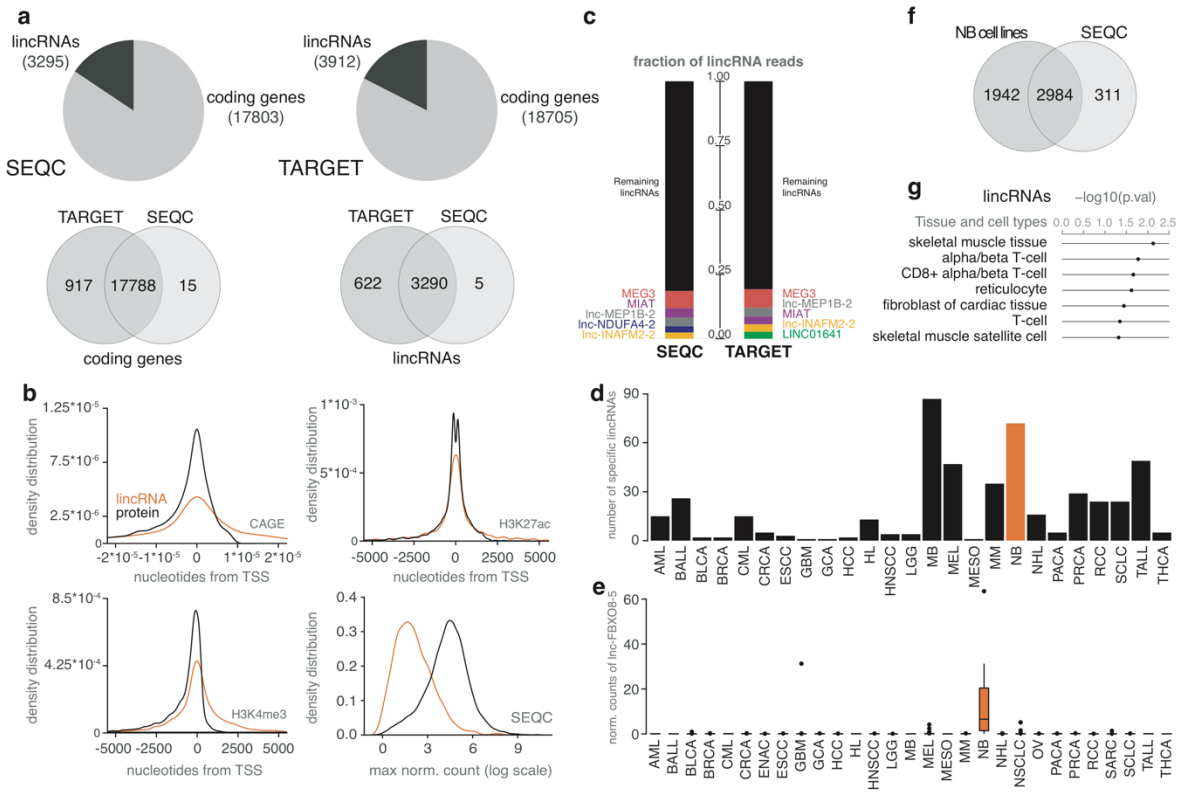


Figure 2: Establishing the NB lincRNome

a) Based on Ensembl annotation, a set of 3295 lincRNAs was robustly expressed in the SEQC dataset, whereas 3912 lincRNAs were found in the TARGET dataset. A total of 3290 lincRNAs were expressed in both groups. b) Density distribution plots showing distances of chromatin marks and CAGE peaks to the TSS and expression levels, for protein coding genes and lincRNAs. c) Percentages of the lincRNA derived read counts for the top 5 expressed lincRNAs for the SEQC and TARGET datasets. d) Bar plot showing the number of specific lincRNAs per cancer type. e) Expression pattern of a randomly selected neuroblastoma specific lincRNA lnc-FBX08-5 across the different cancer types. f) Overlap of expressed lincRNAs between NB cell lines and SEQC. g) Fischer exact-test p-values (non-adjusted) for lincRNAs that are only expressed in the SEQC dataset and are associated with stromal cell types according to FANTOM5. (Cancer type abbreviations: AML: Acute Myeloid leukemia; BALL: B-cell Acute Lymphoblastic Leukemia; BLCA: Bladder Carcinoma; BRCA: Breast Carcinoma; CML: Chronic Myeloid Leukemia; CRCA: Colon Adenocarcinoma; ESCC: Esophageal Squamous Cell Carcinoma; GBM: Glioblastoma; GCA: Gastric Carcinoma; HCC: Hepatocellular Carcinoma; HL: Hodgkin Lymphoma; HSNCC: Head and Neck Squamous Cell Carcinoma; LGG: Brain Lower Grade Glioma; MBM: Medulloblastoma; MEL: Melanoma; MESO: Mesothelioma; MM: Plasma cell myeloma; NB: Neuroblastoma; NHL: Non-Hodgkin Lymphoma; PACA: Pancreatic Adenocarcinoma; PRCA: Prostate carcinoma; RCC: Renal Cell Carcinoma; SCLC: Small Cell Lung Carcinoma; TALL: T-cell Acute Lymphoblastic Leukemia; THCA: Thyroid gland carcinoma.)

lincRNAs are differentially expressed between NB tumors and neuroblasts

NB predominantly exhibit a noradrenergic gene regulatory network including high expression levels of sympatho-adrenal lineage specific bHLH transcription factors such as PHOX2B and genes involved in dopamine synthesis such as tyrosine hydroxylase. We previously isolated and determined the transcriptome using gene expression arrays on microdissected human fetal neuroblasts and provided evidence for the presumed cell-of-origin for NB for these cells[46]. More recently, we have also generated human fetal neuroblast transcriptomes by RNA-seq allowing to also explore expression of lincRNAs in normal reference cells versus NB cells. Detailed validation of the samples will be described elsewhere but importantly, well-established neuronal and chromaffin markers including *TH*, *CHGA*, *BCL2* and *HNK1* were expressed in all samples (Figure 3A).

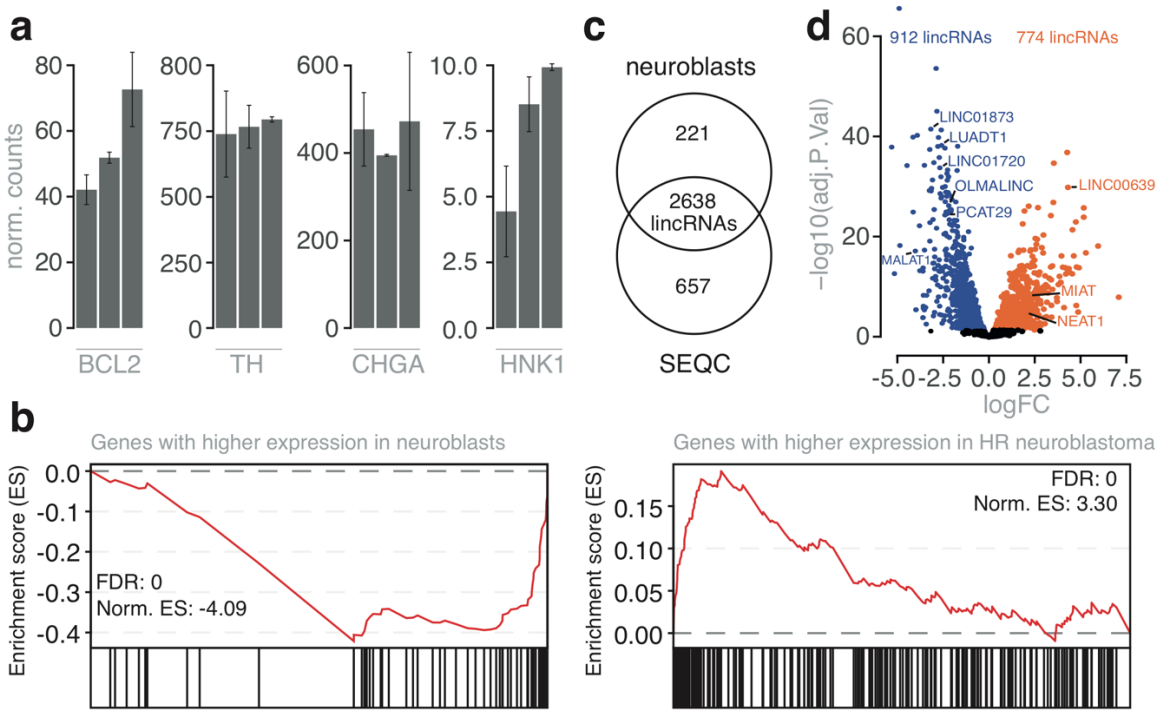


Figure 3: Establishing the neuroblast lincRNome
a) Expression profiles of neural and chromaffin markers in the neuroblast samples (mean epxr. +/- SE). b) GSEA results on a logFC ordered mRNA list, derived from differential expression analysis between neuroblasts and high-risk NB tumors, using neuroblast/HR NB specific gene sets. c) Number and overlap of expressed lincRNAs in neuroblast and NB samples. d) Volcano plot of differentially expressed lincRNAs between neuroblasts and high-risk neuroblastoma samples at q<0.05. The orange dots represent upregulated genes (774 lincRNAs) in HR NB samples, whereas the blue dots depict genes with a lower abundance (912 lincRNAs).

To further validate the RNA seq data, we analyzed mRNA gene sets that were previously reported to be differentially expressed between neuroblasts and high-risk NB tumors[46]. Gene set enrichment analysis demonstrated a strong and significant enrichment of these signatures among up- and downregulated mRNAs between neuroblasts and high-risk NB tumors, supporting the validity of the expression dataset (Figure 3B). We identified 2859 lincRNAs expressed in the neuroblast samples (Figure 3C). The neuroblast lincRNome largely overlapped the NB lincRNome, with 2638 lincRNAs in common. Differential expression analysis revealed 774 and 912 lincRNAs that were significantly up- and downregulated in high-risk NB tumors compared to neuroblasts (Figure 3D and Supplemental table 1). Of interest, the highly abundant lincRNA *MIAT* showed a 4-fold upregulation in NB tumors compared to neuroblasts. Of the 774 lincRNAs upregulated in NB tumors, all 774 were also expressed in NB cell lines and thus likely tumor derived.

DNA copy number alterations drive lincRNA expression

Several studies have shown that DNA copy number alterations can drive lincRNA expression in cancer cells[47–49]. As high-risk NB tumors are characterized by recurrent segmental copy number alterations, we first evaluated lincRNA expression in regions with recurrent copy number gain (1q, 2p, 17q) and loss (1p, 3p, 11q). We found 23.7% of the neuroblastoma lincRNAome with significant positive correlation to their DNA copy number amplitudes (Supplemental Figure 2). These results are in line with similar analysis in other cancer types and further support lincRNA dosage sensitivity.

To evaluate whether segmental copy number alterations can indirectly impact lincRNA expression, we grouped tumors based on copy number status, followed by differential expression analysis (Figure 4). In order not to confound the results with lincRNA expression differences driven by *MYCN* amplification, we excluded all *MYCN* amplified samples from the analysis. Differential lincRNAs were identified for each copy number alteration except for 1p deletions. However, when applying a more robust differential expression analysis, based on iterative subsampling, 17q gain was the only copy number alteration for which differential

lincRNAs were identified. From the 5 lincRNAs differentially expressed between tumors with and without 17q gain, 3 were higher (*lnc-BRC1-2*, *LINC02432* and *lnc-RPS6KA4-3*) and 2 were lower (*LINC02211* and *LINC01930*) in tumors with a 17q gain (Supplemental table 2). These lincRNAs were also expressed in NB cell lines, confirming they are tumor derived. None of the upregulated lincRNAs were located on 17q, suggesting that 17q gain indirectly deregulates lincRNA expression in NB tumor cells.

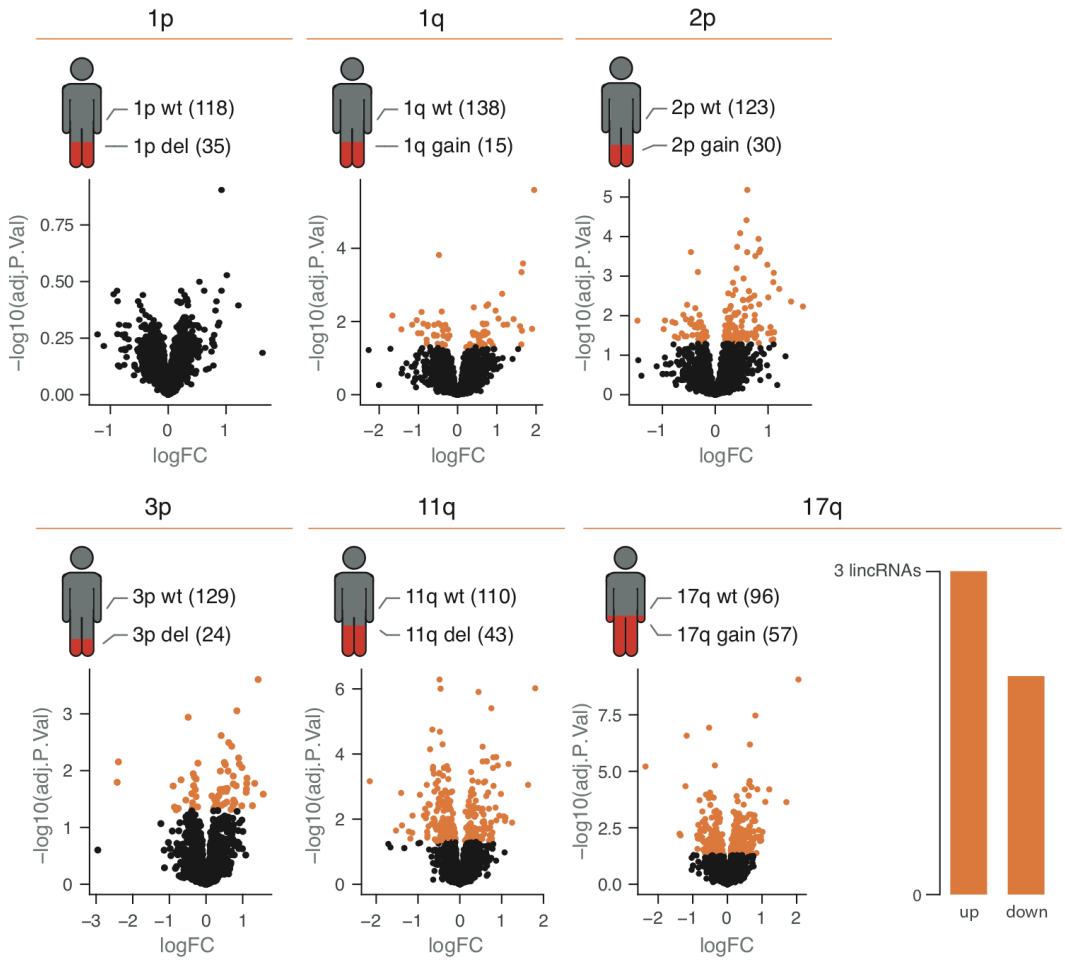


Figure 4: Altered expression levels of lincRNAs by copy number variations
Schematic representation of the number of cases with a copy number variation (1p deletion, 1q gain, 2p gain, 3p deletion, 11q deletion and 17q gain) present in our data set. The volcano plots show differentially expressed lincRNAs per copy number variation. Only 17q gain resulted in 5 robust significantly differentially expressed lincRNAs after iterative differential expression analysis (upregulated lincRNAs: *lnc-BRC1-2*, *LINC02432*, *lnc-RPS6KA4-3*; downregulated: *LINC02211*, *LINC01930*).

NB driver genes regulate lincRNA expression

NB tumors are characterized by low mutation rates[27,34,50]. As a consequence, the identification of oncogenic drivers has been challenging. *MYCN* amplification, activating mutations in the *ALK* receptor tyrosine kinase, *TERT* rearrangements, inactivating *ATRX* mutations and dominant negative mutations in *PHOX2B* are among the most recurrent genetic events that drive oncogenic signaling and tumor formation[23,25,27,51–54]. Our aim was to evaluate to what extent several of these well-established driver genes (*MYCN*, *ALK*) and neuroblastoma identity genes (*PHOX2B*) impact lincRNA expression.

Identification of lincRNAs regulated by NB driver genes in primary tumor samples

For *MYCN* and *ALK*, this was evaluated by grouping tumor samples based on driver gene status (i.e. amplified and mutated respectively) followed by differential lincRNA expression analysis (Figure 5A-B). While *MYCN* amplification status was available for each tumor sample, *ALK* mutation status was not. *ALK* mutation status was therefore determined based on RNA seq data. We identified 54 tumors with missense mutations in the *ALK* gene, with the R1275Q and F1174L mutations as the most frequent ones (Supplemental Figure 3). Differential mRNAs in tumors with varying *MYCN* or *ALK* status were strongly enriched for established mRNA gene sets previously shown to be regulated by *MYCN* or *ALK* (Supplemental Figure 3 and 4), thus validating our approach. Differential lincRNA expression analysis resulted in 1511 (773 upregulated and 738 downregulated) and 80 (55 upregulated and 25 downregulated) lincRNAs for *MYCN* and *ALK*, respectively (Figure 5B and Supplemental table 3 and 4). When applying a more robust differential lincRNA expression analysis, 536 and 1 differential lincRNA(s) for *MYCN* and *ALK* respectively were retained for further analysis (Figure 5C and Supplemental table 3 and 4). While the *ALK* associated lincRNA was also expressed in NB cell lines, expression of up to 11% of *MYCN* associated lincRNAs was restricted to NB tumor samples (Figure 5D). The latter suggests these lincRNAs may be derived from stromal cells whose abundance or composition differs between tumors with and without *MYCN* amplification. To evaluate this hypothesis, we determined immune cell type fractions in each tumor through deconvolution of the mRNA expression data. We found a clear and significant difference in immune cell infiltration between *MYCN* amplified and *MYCN* single copy tumors

for naïve B-cells, CD8+ T-cells and resting NK cells (Supplemental Figure 5). These results confirm previous observations[55] and support our hypothesis that a fraction of lincRNAs associated with *MYCN* amplification status are derived from infiltrating immune cells.

To identify *PHOX2B* associated lincRNAs, an alternative strategy based on two gene expression scores was applied. The first score reflects the activity of the *PHOX2B* core regulatory circuit (CRC) defining the noradrenergic cell state (NOR score). The second score reflects the activity of the *JUN-FOS* CRC, defining the neural crest cell state (NCC score) (Supplemental Figure 6)[56]. As the activity of both CRC inversely correlates, *PHOX2B* associated lincRNAs were defined as lincRNAs that positively and negatively correlate to the NCC and NOR score, respectively, or vice versa. A total of 603 lincRNAs for which such relationship was identified were prioritized for further analysis (Figure 5B-C and Supplemental table 5). Similar to *MYCN*, 9% of these lincRNAs are not expressed in NB cell lines (Figure 5D). We observed significant correlations between tumor NCC/NOR scores and the percentage of immune cells in the tumor biopsies, again suggesting that a fraction of these lincRNAs are stromal derived (Supplemental Figure 7). Interestingly, 120 of the NCC/NOR associated lincRNAs were also differentially expressed in *MYCN* amplified tumors (Figure 6A). Further, 2 and 1 lincRNA(s) that were differentially expressed in *MYCN* amplified tumors were also differentially expressed in tumors with a 17q gain or an *ALK* mutation, respectively.

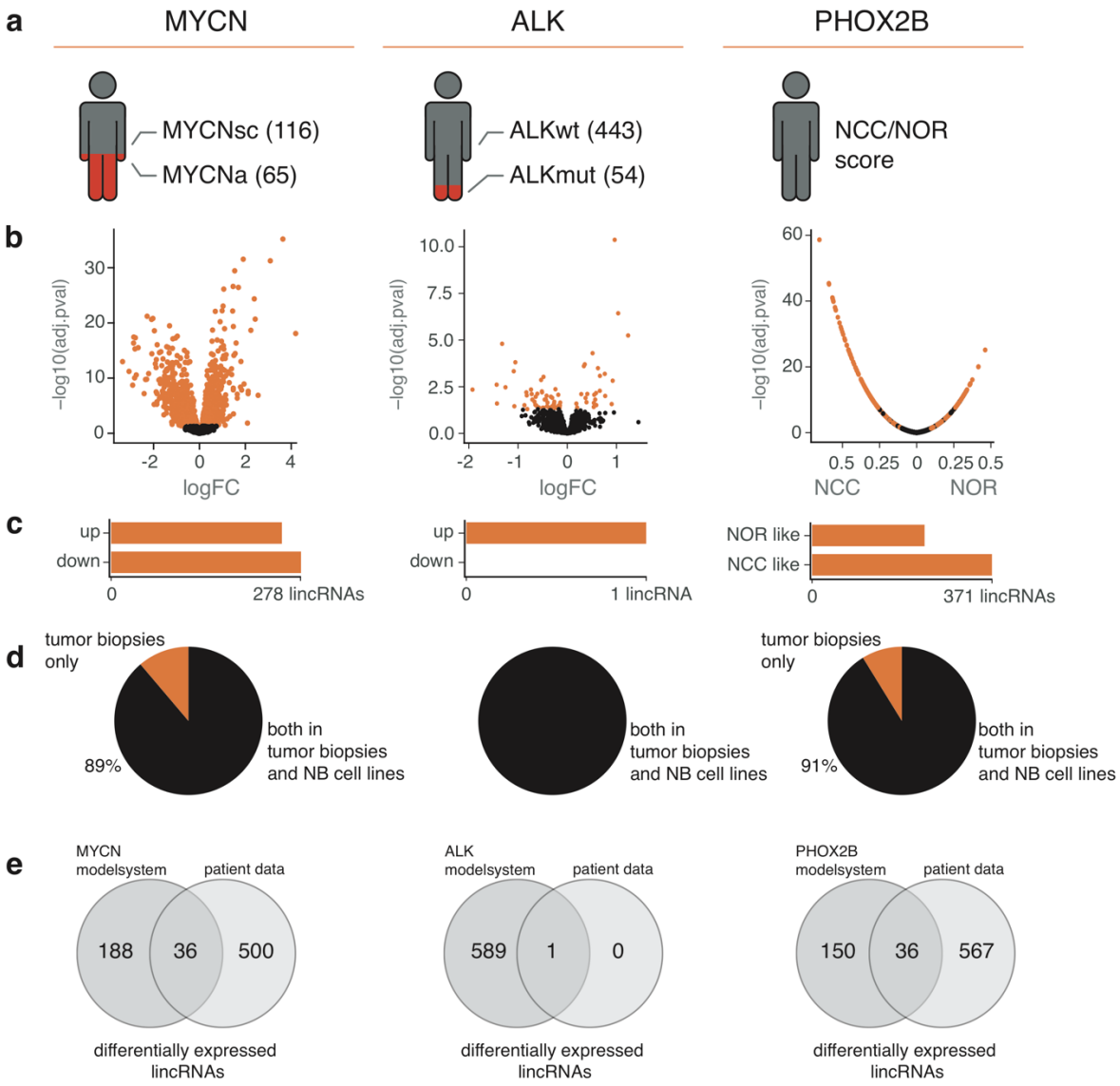


Figure 5: Regulation of lincRNAs by key driver genes

a) Depiction of number of samples in our dataset for the three subtypes of genomic aberrations. b) Volcano plot showing differentially expressed lincRNAs using all samples mentioned in (a) at q<0.05 (orange dots). In the case of *PHOX2B*, correlation coefficient and adjusted p-value are represented. Here, the orange dots represent genes with opposing signs and significance for their correlation with the CRC scores. c) Number of differentially expressed lincRNAs for *MYCN* and *ALK*. The far-right bar plot represents the number of lincRNAs found to be significantly correlated with both the *PHOX2B* CRC and *JUN*/*FOS* CRC in opposing directions. d) Percentage of differentially expressed lincRNAs that are expressed in both the CCLE NB cell lines and the SEQC data set, or solely in the tumor biopsies. e) Overlap of differentially expressed lincRNAs found in the SEQC analysis and after perturbation of the driver genes in the corresponding model systems (p<0.05).

Integration of driver gene model systems to validate lincRNA regulation

To evaluate which of the selected lincRNAs are regulated (directly or indirectly) by these driver genes, we used both in house generated data as well as publicly available RNA seq data for *MYCN*, *ALK* and *PHOX2B* perturbation model systems. For *MYCN* and *PHOX2B*, we applied inducible model systems containing a shRNA construct against *MYCN* (IMR5-75-shMYCN-TR)[57] or *PHOX2B* (CLB-GA-shPHOX2B)[56] or a *MYCN* overexpression construct[58]. For *ALK*, the *ALK* mutant NB cell line CLBGA was treated with the ALK inhibitor crizotinib. Out of 536 lincRNAs differentially expressed in *MYCN* amplified vs *MYCN* single copy tumor samples, 36 were also perturbed in at least one of two *MYCN* model systems (Supplemental table 6). For *ALK* and *PHOX2B*, 1/1 and 36/603 differential lincRNAs were also regulated in the respective model systems (Figure 5E and Supplemental table 7 and 8). These results demonstrate that NB driver genes can (directly or indirectly) regulate lincRNA expression. The majority of these lincRNAs appear to be regulated by a single driver gene (Figure 6B), with only a small subset affected by multiple drivers (i.e. *PHOX2B* and *MYCN*). We found driver gene regulated lincRNAs to be strongly associated with NB patient survival and often differentially expressed between NB tumors and neuroblasts (Figure 6C and Supplemental Figure 8). For instance, 27/36 *MYCN* regulated lincRNAs are significantly associated with patient survival and 16 are differentially expressed between neuroblasts and high-risk NB tumors (Supplemental table 9). Occasionally, these lincRNAs are located within a super-enhancer or display a NB-specific expression profile. In total, 14 lincRNAs are associated with multiple parameters, increasing their potential importance in NB biology.

One example is *lnc-GOLGA61-1*, a divergent lincRNA upstream of the *ISLR2* gene. *lnc-GOLGA61-1* is upregulated in tumor samples harboring an *ALK* mutation or a *MYCN* amplification. Treatment with crizotinib strongly represses *lnc-GOLGA61-1* expression, suggesting *ALK* is involved in *lnc-GOLGA61-1* regulation. Of interest, *lnc-GOLGA61-1* and *ISLR2* expression are strongly correlated in NB tumors (Supplemental Figure 9). *ISLR2* is an interaction partner of *NTRK1* and *RET*, both involved in regulating NB differentiation, and *RET* has been shown to be activated by mutant *ALK*[59,60]. Expression of *lnc-GOLGA61-1* and *ISLR2* is restricted to NB cells as evidenced by RNA seq data of the CCLE cohort[61] (Figure 6D and Supplemental Figure 10).

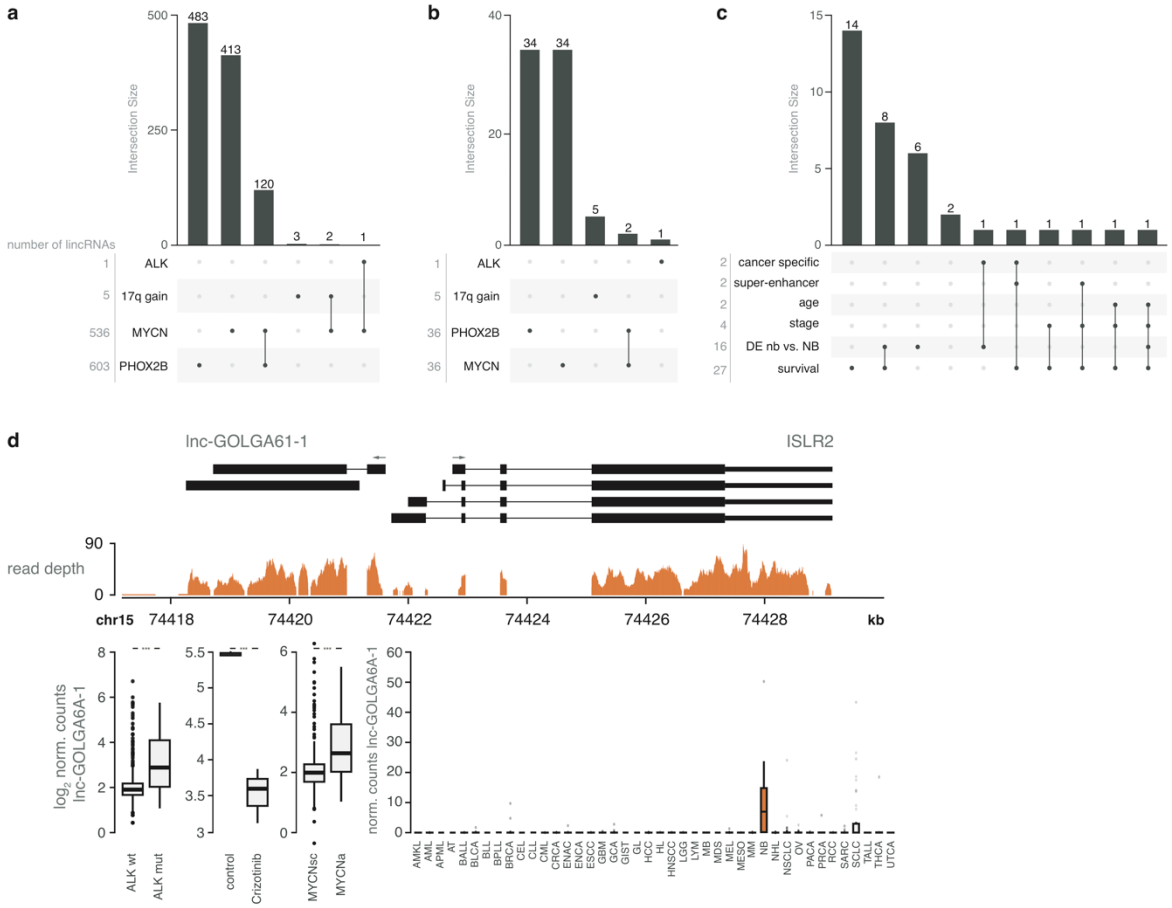


Figure 6: Association of lincRNAs with genetic and clinical parameters
a) Number of lincRNAs regulated by one or more driver gene(s) or differentially expressed upon copy number variation in the patient samples. b) Number of lincRNAs regulated by one or more driver genes or differentially expressed upon copy number variation, for lincRNAs differentially expressed in both the patient samples and model systems. c) Number of lincRNAs regulated by MYCN, associated with clinical and genetic features. d) Representation of the genomic locus and RNA-seq data of *lnc-GOLGA6A-1*. *lnc-GOLGA6A-1* expression levels in *ALK* wild type and mutated samples, and *MYCN* amplified and single copy samples are shown, together with expression of the lincRNA upon crizotinib treatment, an ALK inhibitor. The boxplot shows the expression pattern of *lnc-GOLGA6A-1* across the different cancer types.

LincRNAs as upstream regulators of neuroblastoma driver genes

LincRNAs as modulators of driver gene activity

The above described analyses demonstrate the impact of various NB driver genes on lincRNA expression. However, lincRNAs can potentially also function upstream of, or in concert with these driver genes. To uncover such lincRNAs, we applied LongHorn, a computational pipeline aimed at uncovering effector genes and target genes of individual lincRNAs[62]. The pipeline

essentially considers lincRNAs as modulators of effector proteins (i.e. transcription factors (TF) or RNA-binding proteins (RBP)) or as indirect regulators of target gene expression (Figure 7). To uncover these relationships, LongHorn integrates mRNA and lincRNA expression data with experimental data on RNA-RBP interactions (eCLIP), TF regulation (ChIP-seq and PWM) and lincRNA-DNA binding site prediction (triplex). As both *MYCN* and *PHOX2B* are transcription factors, we first evaluated if lincRNAs could modulate *MYCN* or *PHOX2B* activity. Only tumor derived (i.e. expressed in primary tumors and cell lines) lincRNAs with a median absolute deviation > 0.1 in the SEQC dataset were considered. LongHorn uncovered 25 and 36 lincRNAs that were predicted to modulate *PHOX2B* and *MYCN* activity, respectively (Figure 7A, Supplemental table 9). Importantly, none of these lincRNAs correlate with *MYCN* or *PHOX2B* expression levels directly, excluding the possibility that such correlations may confound the results. LincRNAs predicted to modulate *MYCN* activity included - amongst others - *MIAT*, *TSIX* and *MEG3*. Of note, 7 lincRNAs were found to modulate both *MYCN* and *PHOX2B* activity. Targets that were predicted to be affected by lincRNA modulation with *MYCN* or *PHOX2B* as effectors were subsequently evaluated for enrichment of hallmark gene sets. We observed enrichments for, amongst others, the TP53 pathway (8 lincRNAs), STAT signaling (24 lincRNAs), KRAS signaling (14 lincRNAs) and the apoptosis pathway (9 lincRNAs) (Figure 7B). We identified *TSIX* as a modulator of *MYCN* activity, driving expression of KRAS signaling genes including *BIRC3* and *CCND2* (Figure 7C). Alternatively, *TTTY15* was found to modulate *MYCN* activity towards repression of the apoptosis pathway genes *GADD45B* and *BTG2* (Figure 7C).

LincRNAs as regulators of driver gene activity

Next to direct modulation of *MYCN* or *PHOX2B* activity, lincRNAs may also modulate effectors of *MYCN*, *PHOX2B* or *ALK* expression (Figure 7D). We identified 65 lincRNAs that were predicted to regulate *MYCN* expression through modulation of one or more effectors. Several effectors of *MYCN* expression, including *TCF3*, *NAP1L1*, *HMGA1* and *CREB3L4* were predicted to be modulated by multiple lincRNAs (Figure 7D-E). No lincRNAs were predicted to regulate *PHOX2B* or *ALK* expression. Taken together, these analyses demonstrate that lincRNAs may indeed regulate or modulate the expression or activity of one or multiple NB driver genes.

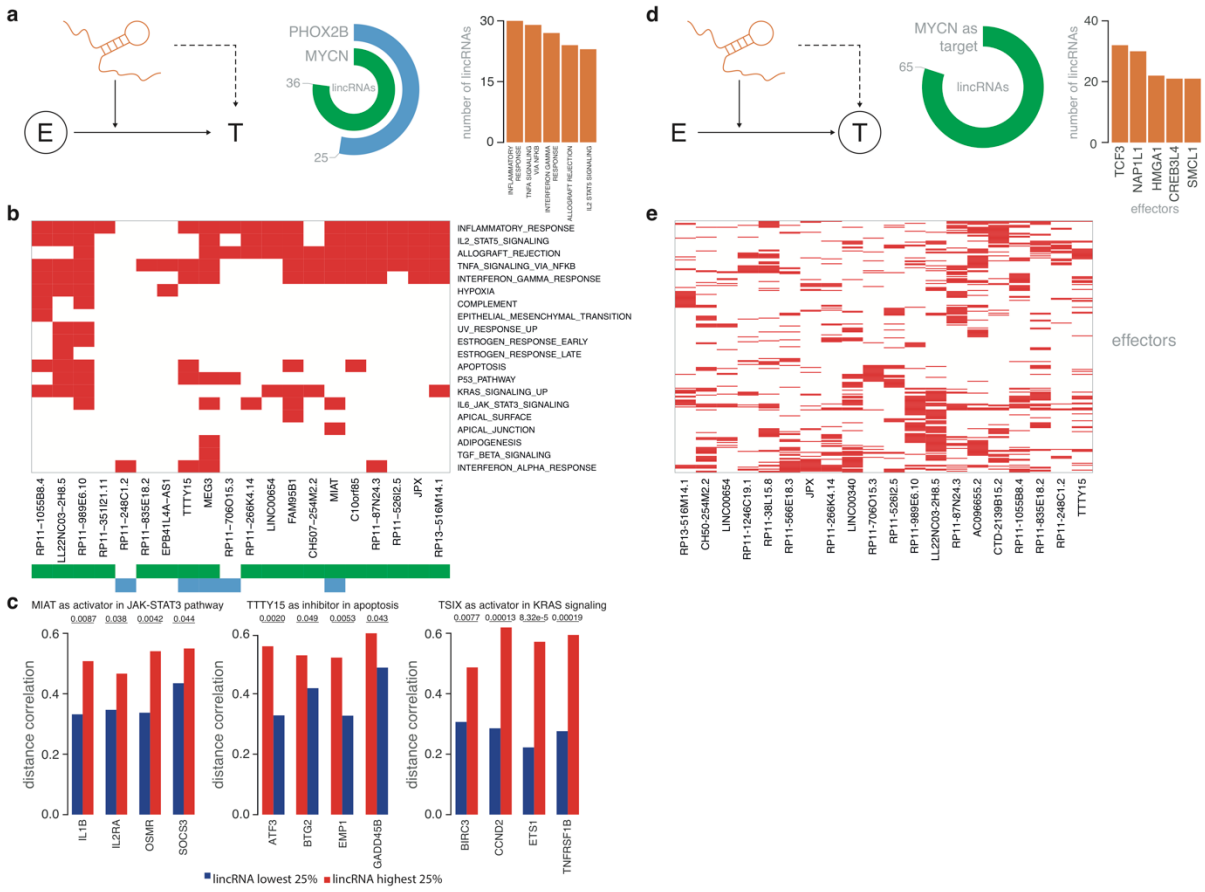


Figure 7: Identification of lincRNAs as modulators of activity or expression of driver genes

a) Schematic representation of the investigated lincRNAs, modulating the activity of the effector proteins, MYCN and PHOX2B. The circular plots display the number of lincRNAs, expressed in NB cell lines, found to be modulators of MYCN and PHOX2B targets. Bar plots show the top 5 hallmarks that were significantly enriched (Fisher exact test, $p<0.001$). b) The heatmap visualizes the clustering of the significantly enriched hallmarks for the top 20 modulating lincRNAs. c) Differences in distance correlation between the samples of low and high abundance of the lincRNAs. The presented targets are genes enriched in the hallmark gene sets. d) Schematic representation of the investigated lincRNAs, regulating the activity of the target protein MYCN. The circular plots display the number of activating and inhibiting lincRNAs modulating *MYCN* expression. The top 5 effectors targeting *MYCN* are shown in the bar plot. e) Clustering of the top 20 regulating lincRNAs with MYCN as target is visualized in the heatmap.

Discussion

We have evaluated RNA seq data of primary NB tumors, human fetal neuroblasts and various cellular perturbation model systems to reveal alterations in lincRNA expression patterns invoked by driver mutations, amplifications or DNA copy number variations. Through various prioritization strategies, we provide a core set of lincRNAs with a potential role in NB tumor biology, up- or downstream of the key NB driver genes *MYCN*, *ALK* and *PHOX2B*.

Integration of RNA seq data from primary tumors and cell lines revealed that a fraction of the lincRNAs expressed in tumor biopsies were not detected in cell lines. Cell lines are known to have higher levels of CpG hypermethylation than primary tumors[63,64], potentially explaining why expression of certain lincRNAs was restricted to tumor samples. In addition, the tumor sample cohort is more heterogeneous in nature compared to the NB cell lines which are typically derived from high-risk tumors only. However, tumor biopsies also have a stromal component that is absent in cell lines, leading us to hypothesize that some of these lincRNAs are stromal in origin. Tissue ontology enrichment analysis provided support for this hypothesis. Moreover, we could demonstrate that stromal composition, and more specifically immune cell infiltration, correlated with *MYCN* amplification status and *PHOX2B* core regulatory circuit activity. This confounded our differential lincRNA expression analysis between *MYCN* amplified and *MYCN* single copy tumors and lincRNA *PHOX2B* CRC correlation analysis. As a result, many lincRNAs that were prioritized as differentially expressed or correlated, were undetected in NB cell lines and thus likely stromal cell derived. Cell type composition of tumor biopsies can be elucidated using computational deconvolution methods[65–67]. In addition, RNA seq data from various human cell types is becoming increasingly available[3]. Integrating this type of information when performing (differential) lincRNA (or mRNA) expression analysis on tumor biopsies could help elucidate the cell of origin of RNA molecules and assist in prioritizing key players in tumor biology.

By combining RNA seq data from primary tumors with model systems for *MYCN*, *ALK* and *PHOX2B*, we could demonstrate that each of these driver genes regulate a core set of lincRNAs. Whether these lincRNAs are regulated directly (e.g. through binding of *MYCN* or

PHOX2B transcription factors in the lincRNA promoter) or indirectly remains to be determined. Several of these lincRNAs were found to be differentially expressed between NB tumors and precursor neuroblasts and/or associated with patient survival and disease stage. To further prioritize driver gene regulated lincRNAs, we evaluated their link with super-enhancers and NB expression specificity. Several lincRNAs that play a role in tumor biology, including *CCAT1-L* and *SAMMSON*[42,68] are associated with these features. This core set of NB driver gene regulated lincRNAs should be further explored by genetic perturbation experiments to investigate their impact on the cellular and molecular phenotype.

As driver genes themselves could be under the control of one or more lincRNAs, we applied a state-of-the-art computational workflow aimed at prioritizing lincRNAs that modulate driver gene activity or expression. This resulted in 36 and 25 lincRNAs that modulate *MYCN* and *PHOX2B* activity, respectively. Functional associations between *MYC* and several lincRNAs predicted to modulate *MYCN* activity, have been demonstrated previously for *MEG3*[69] and *TSIX*[70] amongst others. *MIAT*, one of the most abundant lincRNAs in NB, was identified as a modulator of both *MYCN* and *PHOX2B*. Target genes that were affected as a result of this modulation were significantly enriched in the IL6-JAK-STAT3 pathway. Interestingly, *MIAT* has been described to enhance *STAT3* expression by acting as a molecular sponge for *miR-181b*[71], a miRNA upregulated in *MYCN* amplified NB tumors[72]. *MIAT* was also identified, together with 48 additional lincRNAs, as a candidate to modulate the activity of effectors of *MYCN* expression. *TCF3*, *HMGA1* and *CREB3L4* were among the most recurrent effectors of *MYCN*. *HMGA1* is able to regulate *MYCN* expression in NB cells[73], while *TCF3* has been shown to regulate *MYC* expression in colorectal cancer cells[74]. Moreover, *TCF3* has previously been identified as a master regulator in *MYCN* amplified NB tumors[75].

In summary, we identified a comprehensive catalogue of lincRNAs up- and downstream of key NB driver genes. These lincRNAs could play an important role in tumor initiation and progression and may serve as a solid starting point for further experimental validation.

Material and Methods

Annotation and quantification

The TARGET fastq files were downloaded from the Genomics Data Commons Data portal. Kallisto (v0.42.4) was used to quantify gene expression in the samples, using the hg38 human assembly (GRCh38.p10), encompassing 37,297 lincRNA and 180,869 protein coding transcripts. Only long intergenic non-coding RNAs were considered, as RNA seq data for both data sets was unstranded. The transcripts were classified according to the Ensembl biotype annotation (GRCh38.p10).

Cancer cell line encyclopedia

RNA sequencing data of the Cancer Cell Line Encyclopedia was reprocessed using Kallisto with the hg38 human genome assembly (GRCh38.p10). We used this RNA seq data set to filter out lincRNAs not expressed in NB cell lines (i.e. tumor specific). Uberon and cell types assigned to lincRNAs were downloaded from FANTOM5[3]. Only lincRNAs with Ensembl gene IDs were selected. Enrichment of stromal cell types was determined based on tumor specific lincRNAs (Fisher Exact test, $p<0.05$).

Specificity of genes was calculated based on a minimum fold change >3 between median expression of each cancer type per gene.

Neuroblasts

Ethical approval was obtained for the collection of fetal adrenal glands from fetuses aborted for clinical reasons and informed consent was obtained for the use of all samples (Ethics committee Erasme Hospital, Brussels, Belgium; approval no.: OM021). All methods were carried out in accordance with relevant guidelines and regulations. Neuroblasts were isolated from 3 fetal adrenal glands from 13-16 week gestation embryos using laser capture microdissection. We extracted RNA from 6 neuroblast clusters and 3 areas of adjacent normal adrenal cortex as controls using the PicoPure kit (ARCTURUS). Samples were PCR amplified (SMART-Seq v4 ultra low input RNA kit, Takara Bio) and sequenced on the Illumina Hiseq 4000 platform to create a unique resource of neuroblast mRNA and lincRNA expression data.

Copy number analysis

Copy numbers status was determined using array CGH. Copy number amplitudes (CNA) higher than 2.5 and lower than 1.5 were annotated as aberrant. Each segment was annotated by its corresponding chromosome arm, allowing classification according to known chromosome arm gains and deletions in NB tumors. To assess of dosage sensitivity of lincRNAs, a CNA was assigned to each gene per sample, based on its chromosomal location. Correlation with expression was calculated using Pearson's method, p values were adjusted using the Benjamini-Hochberg method[76] ($q<0.05$).

RNA sequencing data based mutation analysis

The SEQC and TARGET data set were aligned to the human hg19 assembly, using TopHat (v2.10). Mutations were identified by means of the Genome Analysis ToolKit (v3.2-2) using the RNA-seq best practices workflow. Only mutations in protein coding genes deemed damaging or possibly damaging by means of Polyphen and Sift, were retained. Variants with a prevalence of more than 0.1% according to gnomAD, having a total read-depth below 5 or a read-depth for the mutant allele below 3, were filtered out.

Differential expression analysis

Limma voom (v3.36.5) was used to assess differential expression between neuroblast and high-risk neuroblastoma samples, mutated and non-mutated samples for *ALK,* and amplified and non-amplified samples with INSS stage 4 for *MYCN* and gain/deletions and wild type samples for the CNVs, for genes expressed in at least half of the SEQC samples. Genes were classified as differentially expressed based on their adjusted p-value ($q<0.05$). For a more robust differential expression analysis for *ALK* and *MYCN* differentially expressed genes, the SEQC data set was divided into two subgroups, having an equal number of mutated or amplified samples. Differential expression analysis was performed for both subgroups and genes differentially expressed in both groups were identified ($q<0.05$). This workflow was repeated 100 times, and only genes differentially expressed in more than 80 of the repeats were classified as being truly differentially expressed genes. Differential expression in the model systems was calculated using limma voom, with a threshold of $p<0.05$ (non-adjusted p-value).

Gene set enrichment analysis

Gene set enrichment analysis[77] was performed using the Java GSEA application of the Molecular Signatures Database (*1000 permutations, classic analysis*). To validate differentially expressed genes between high-risk NB samples and neuroblasts, a list of mRNAs was analyzed through pre-ranked GSEA, using gene sets built out of known differentially expressed genes between these two groups of samples[46]. GSEA was also performed on mRNA lists resulting from differential expression analysis between *MYCN* amplified and *MYCN* single copy tumors (INSS stage 4) and mRNA list from treated and untreated *MYCN* model systems, using all curated gene sets in the Molecular Signatures Database (Supplemental Figure 11). For the GSEA on mRNA list from differential expression analysis for the CNVs, only the positional genesets were used (Supplemental Figure 12). In the case of *ALK*, no gene sets are available. As such, we created our own gene sets based on a validated *ALK* mRNA signature[59]. Significant enrichment was defined at FDR<0.05. All mRNA lists were ordered based on the log-transformed fold change.

Defining stromal cell composition

CIBERSORT, a computational method to estimate cell type fractions from bulk RNA-seq data, was used to define the cell type composition of the primary tumors. Gene expression data sets with raw counts were used as input. The algorithm (v1.04) was run in R (v3.5.0) with the default signature matrix at 100 permutations. Statistical significance between the *MYCN* amplified and *MYCN* single copy subset was calculated with a Mann-Whitney test. For samples associated with the *PHOX2B* or the *JUN*/*FOS* core regulatory circuit, correlations of the percentages of immune cell types and the ranksums were calculated with Spearman's rank correlation method. P-values were adjusted using the Benjamini-Hochberg method[76]. Significance cut-off was set at q<0.05.

Cell line perturbation models

Four neuroblastoma model systems were used in this study. For *MYCN*, publicly available RNA seq data sets for IMR5-75-shMYCN-TR (ArrayExpress E-MTAB-6568) and SHEP-MYCN-TET (Gene Expression Omnibus GSE83327) cells were used. CLBGA-shPHOX2B cells[56] were treated

with doxycycline for 5 days (n=2), together with a shControl cell line (n=2) (Supplemental Figure 13). *ALK* mutant CLBGA cells were treated with *ALK* inhibitor crizotinib at a concentration of 500 nM for 24 hours. Over the course of 3 weeks, matched treatment of 1 sample and 1 control sample was performed per week.

cDNA library prep and sequencing

Total RNA was extracted from the CLB-GA cell line for the ALK model system using TRIzol Reagent (Invitrogen) and the miRNeasy Mini Kit (Qiagen). For the CLB-GA-shPHOX2B system, total RNA was extracted from fresh cells using TRIzol® Reagent (Invitrogen) and the AllPrep DNA/RNA Mini Kit (Qiagen). All samples were subjected to quality control on a Bioanalyzer instrument and all RNA exhibited a RIN (RNA Integrity Number) > 8. All RNA sequencing libraries were prepared from 200 ng of total RNA using the Illumina TruSeq Stranded mRNA Library preparation kit. Kappa qPCR quantification was used to perform equimolar pooling. The concentration of the pooled library was measured with Qubit. Sequencing of 1.2 pM of pooled library was performed with the Illumina NextSeq 500 instrument using 2 x 75 cycles (paired-end) for all samples (high output sequencing kit). Transcripts were quantified by means of Kallisto using the human GRCh38 transcriptome as a reference.

Survival analysis

Overall survival analysis was performed on the SEQC data set using a Kaplan-Meier analysis. The curves were created by dichotomizing the RNA-expression data, using the median expression value as a cutoff. The log-rank test was used to compare the two curves and generate a significance level of the impact of expression on overall survival for each gene. Multiple testing correction was performed using the Benjamini-Hochberg method. Genes with $q<0.05$ were considered to be associated with survival.

Super-enhancers

H3K27ac ChIP-sequencing and super-enhancer analysis using LILLY was performed as described[78] (Supplemental Figure 14). A gene was classified to be located in a super-enhancer if that region was called in a minimum of 14 different NB cell lines. The lincRNAs were ordered based on their mean rank over all NB cell lines.

LONGHORN

LongHorn searches for sequence patterns in proximal promoters that are predictive of RNA-DNA triplex structures identified by Triplexator[79], and expression-based evidence for modulation of transcription factor (TF) activity. To predict modulation, we first collected candidate TF-target pair interactions and then tested for evidence of their modulation. For all candidate TF-target pairs, we required that each TF-target candidate has a significant nonlinear correlation ($p<1E-11$) as estimated by distance correlation (dCor), and either TF binding evidence from ENCODE ChIP-Seq assays or predicted interactions based on published TF binding-site motifs. To collect evidence for the modulation of these TF-target interaction candidates, we used delta dCor within a triplet composed of a lncRNA, a TF, and a protein-coding target. Specifically, for each lncRNA, we partitioned all tumor samples into four groups based on the expression profile of this lncRNA, from lowest to highest. To avoid circularity, for each triplet, we added an independence constraint by requiring that the lncRNA was not correlated with the TF ($p>0.1$) and a range constraint by requiring a minimum of 2x fold-change between the lncRNA's average expression in the two sets (low vs high). Then, comparing the sample groups with highest and lowest lncRNA expression, we required a nonparametric $p<0.05$ for the delta dCor between the TF and the target against a bootstrapping-based null hypothesis. These p-values were integrated across all significant triplets using Fisher's method to identify significant lncRNA-target pairs at an adjusted $p<0.01$. Enrichment of the cancer hallmarks gene sets of the Molecular Signatures Database was calculated using Fisher's exact test ($q<0.05$)[80]. The top 20 lincRNAs modulating *MYCN* or *PHOX2B* effect were selected based on the minimum calculated adjusted p-value, regardless of the specified effector. LincRNAs with *MYCN* as target in the heatmap were selected based on the minimum p-value. All selected lincRNAs were expressed in NB cell lines.

Data availability

RNA sequencing data of the ALK and PHOX2B model systems have been deposited in GEO with the accession codes GSE124450 and GSE124451.

Acknowledgements

Supported by grants from Fonds Wetenschappelijk Onderzoek (FWO – Grant Number 131579), Kom Op Tegen Kanker (Grant Number – 00000000116000000099) and Stichting Tegen Kanker.

Author contributions

J.V., S.L., P.M. and D.R. designed the project. F.S., K.D.P., G.D., N.D. and S.R. collected and provided the neuroblast samples. M.F. provided the SEQC RNA-seq data. M.F. and C.B. provided aCGH data. C.E. reanalyzed the CCLE data and the model system RNA-seq data. B.D. and D.R. performed the superenhancer analysis. A.P. and I.J. performed the treatment and provided RNA of the PHOX2B model system. H.S.C. and P.S. performed the LongHorn analysis. N.Y. performed the library preps of the model systems. D.R. analyzed the data, with the help of S.L. and P.M. P.M. and S.L. supervised the project. D.R. made all the figures. D.R., S.L. and P.M. wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

References

1. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101 (2012).
2. The ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science (80-.).* **306**, 636 LP-640 (2004).
3. Hon, C.-C. *et al.* An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* **543**, 199 (2017).
4. Volders, P.-J. *et al.* LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.* **41**, D246–D251 (2013).
5. Volders, P.-J. *et al.* LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.* gky1031-gky1031 (2018).
6. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
7. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
8. Fang, Y. & Fullwood, M. J. Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics. Proteomics Bioinformatics* **14**, 42–54 (2016).
9. Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43**, 904–914 (2011).
10. Marchese, F. P., Raimondi, I. & Huarte, M. The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* **18**, 206 (2017).
11. Schmitt, A. M. & Chang, H. Y. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* **29**, 452–463 (2016).
12. Chen, X., Yan, C. C., Zhang, X. & You, Z.-H. Long non-coding RNAs and complex diseases: from experimental results to computational models. *Brief. Bioinform.* **18**, 558–576 (2017).
13. Sanchez Calle, A., Kawamura, Y., Yamamoto, Y., Takeshita, F. & Ochiya, T. Emerging roles of long non-coding RNA in cancer. *Cancer Sci.* **109**, 2093–2100 (2018).
14. Prensner, J. R. & Chinnaiyan, A. M. The emergence of lncRNAs in cancer biology. *Cancer Discov.* **1**, 391–407 (2011).
15. Chen, S. *et al.* LincRNa-p21: function and mechanism in cancer. *Med. Oncol.* **34**, 98 (2017).
16. KOTAKE, Y. *et al.* Long Non-coding RNA, PANDA, Contributes to the Stabilization of p53 Tumor Suppressor Protein. *Anticancer Res.* **36**, 1605–1611 (2016).
17. Iaccarino, I. lncRNAs and MYC: An Intricate Relationship. *Int. J. Mol. Sci.* **18**, 1497 (2017).
18. Ashouri, A. *et al.* Pan-cancer transcriptomic analysis associates long non-coding RNAs with key mutational driver events. *Nat. Commun.* **7**, 13197 (2016).
19. Kawasaki, Y. *et al.* MYU, a Target lncRNA for Wnt/c-Myc Signaling, Mediates Induction of CDK6 to Promote Cell Cycle Progression. *Cell Rep.* **16**, 2554–2564 (2016).
20. Doose, G. *et al.* MINCR is a MYC-induced lncRNA able to modulate MYC's transcriptional network in Burkitt lymphoma cells. *Proc. Natl. Acad. Sci.* **112**, E5261 LP-E5270 (2015).
21. Maris, J. M. Recent Advances in Neuroblastoma. *N. Engl. J. Med.* **362**, 2202–2211 (2010).
22. Raabe, E. H. *et al.* Prevalence and functional consequence of PHOX2B mutations in neuroblastoma. *Oncogene* **27**, 469 (2007).

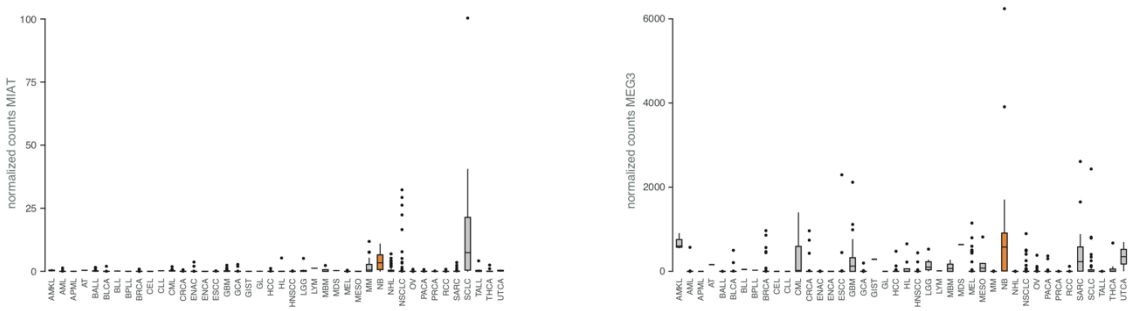
23. Trigg, R. M. & Turner, S. D. ALK in Neuroblastoma: Biological and Therapeutic Implications. *Cancers (Basel).* **10**, 113 (2018).
24. Chen, Y. *et al.* Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* **455**, 971 (2008).
25. Cheung, N.-K. V *et al.* Association of age at diagnosis and genetic mutations in patients with neuroblastoma. *JAMA* **307**, 1062–1071 (2012).
26. Olsen, R. R. *et al.* MYCN induces neuroblastoma in primary neural crest cells. *Oncogene* **36**, 5075 (2017).
27. Peifer, M. *et al.* Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **526**, 700 (2015).
28. Pandey, G. K. *et al.* The Risk-Associated Long Noncoding RNA NBAT-1 Controls Neuroblastoma Progression by Regulating Cell Proliferation and Neuronal Differentiation. *Cancer Cell* **26**, 722–737 (2014).
29. Yu, M.; Ohira, M.; Li, Y.; Niizuma, H.; Oo, M. L.; Zhu, Y.; Ozaki, T.; Isogai, E.; Nakamura, Y.; Koda, T.; Oba, S.; Yu, B.; Nakagawara, A. High expression of ncRAN, a novel non-coding RNA mapped to chromosome 17q25.1, is associated with poor prognosis in neuroblastoma. *Int. J. Oncol.* **34**, 931–938 (2009).
30. Atmadibrata, B. *et al.* The Novel Long Noncoding RNA linc00467 Promotes Cell Survival but Is Down-Regulated by N-Myc. *PLoS One* **9**, e88112 (2014).
31. Liu, P. Y. *et al.* Effects of a Novel Long Noncoding RNA, lncUSMycN, on N-Myc Expression and Neuroblastoma Progression. *JNCI J. Natl. Cancer Inst.* **106**, dju113-dju113 (2014).
32. Consortium, S.-I. *et al.* A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* **32**, 903 (2014).
33. Zhang, W. *et al.* Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol.* **16**, 133 (2015).
34. Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nat. Genet.* **45**, 279 (2013).
35. Avila Cobos, F. *et al.* Zipper plot: visualizing transcriptional activity of genomic regions. *BMC Bioinformatics* **18**, 231 (2017).
36. (DGT), T. F. C. and the R. P. M. I. and C. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462 (2014).
37. Briggs, J. A., Wolvetang, E. J., Mattick, J. S., Rinn, J. L. & Barry, G. Mechanisms of Long Non-coding RNAs in Mammalian Nervous System Development, Plasticity, Disease, and Evolution. *Neuron* **88**, 861–877 (2015).
38. Romero-Barrios, N., Legascue, M. F., Benhamed, M., Ariel, F. & Crespi, M. Splicing regulation by long noncoding RNAs. *Nucleic Acids Res.* **46**, 2169–2184 (2018).
39. Liu, Z. *et al.* Long non-coding RNA MIAT promotes growth and metastasis of colorectal cancer cells through regulation of miR-132/Derlin-1 pathway. *Cancer Cell Int.* **18**, 59 (2018).
40. Lai, I.-L. *et al.* Long noncoding RNA MIAT promotes non-small cell lung cancer proliferation and metastasis through MMP9 activation. *Oncotarget* **8**, 98148–98162 (2017).
41. Li, Y.; Wang, K.; Wei, Y.; Yao, Q.; Zhang, Q.; Qu, H.; Zhu, G. lncRNA-MIAT regulates cell biological behaviors in gastric cancer through a mechanism involving the miR-29a-3p/HDAC4 axis. *Oncol. Rep.* 3465–3472 (2017).

doi:https://doi.org/10.3892/or.2017.6020
42. Leucci, E. *et al.* Melanoma addiction to the long non-coding RNA SAMMSON. *Nature* **531**, 518 (2016).
43. Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**, 199–208 (2015).
44. Böttcher, R. *et al.* Novel long non-coding RNAs are specific diagnostic and prognostic markers for prostate cancer. *Oncotarget* **6**, 4036–4050 (2015).
45. Yang, L. *et al.* lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* **500**, 598–602 (2013).
46. De Preter, K. *et al.* Human fetal neuroblast and neuroblastoma transcriptome analysis confirms neuroblast origin and highlights neuroblastoma candidate genes. *Genome Biol.* **7**, R84–R84 (2006).
47. Guan, Y. *et al.* Amplification of PVT1 Contributes to the Pathophysiology of Ovarian and Breast Cancer. *Clin. Cancer Res.* **13**, 5745 LP-5755 (2007).
48. Volders, P.-J. *et al.* Targeted Genomic Screen Reveals Focal Long Non-Coding RNA Copy Number Alterations in Cancer Cell Lines. *Non-coding RNA* **4**, 21 (2018).
49. Heselmeyer-Haddad, K. *et al.* Genomic amplification of the human telomerase gene (TERC) in pap smears predicts the development of cervical cancer. *Am. J. Pathol.* **166**, 1229–1238 (2005).
50. Molenaar, J. J. *et al.* Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature* **483**, 589 (2012).
51. Huang, M. & Weiss, W. A. Neuroblastoma and MYCN. *Cold Spring Harb. Perspect. Med.* **3**, a014415–a014415 (2013).
52. Mosse, Y. P. *et al.* Germline PHOX2B Mutation in Hereditary Neuroblastoma. *Am. J. Hum. Genet.* **75**, 727–730 (2004).
53. Reiff, T. *et al.* Neuroblastoma Phox2b Variants Stimulate Proliferation and Dedifferentiation of Immature Sympathetic Neurons. *J. Neurosci.* **30**, 905 LP-915 (2010).
54. Trochet, D. *et al.* Germline Mutations of the Paired–Like Homeobox 2B (PHOX2B) Gene in Neuroblastoma. *Am. J. Hum. Genet.* **74**, 761–764 (2004).
55. Zhang, P. *et al.* MYCN Amplification Is Associated with Repressed Cellular Immunity in Neuroblastoma: An In Silico Immunological Analysis of TARGET Database. *Front. Immunol.* **8**, 1473 (2017).
56. Boeva, V. *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* **49**, 1408 (2017).
57. Muth, D. *et al.* Transcriptional Repression of SKP2 Is Impaired in MYCN-Amplified Neuroblastoma. *Cancer Res.* **70**, 3791 LP-3802 (2010).
58. Schulte, J. H. *et al.* MYCN regulates oncogenic MicroRNAs in neuroblastoma. *Int. J. Cancer* **122**, 699–704 (2007).
59. Lambertz, I. *et al.* Upregulation of MAPK Negative Feedback Regulators and RET in Mutant ALK Neuroblastoma: Implications for Targeted Treatment. *Clin. Cancer Res.* **21**, 3327 LP-3339 (2015).
60. Lopez-Delisle, L. *et al.* Activated ALK signals through the ERK–ETV5–RET pathway to drive neuroblastoma oncogenesis. *Oncogene* **37**, 1417–1429 (2018).
61. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of

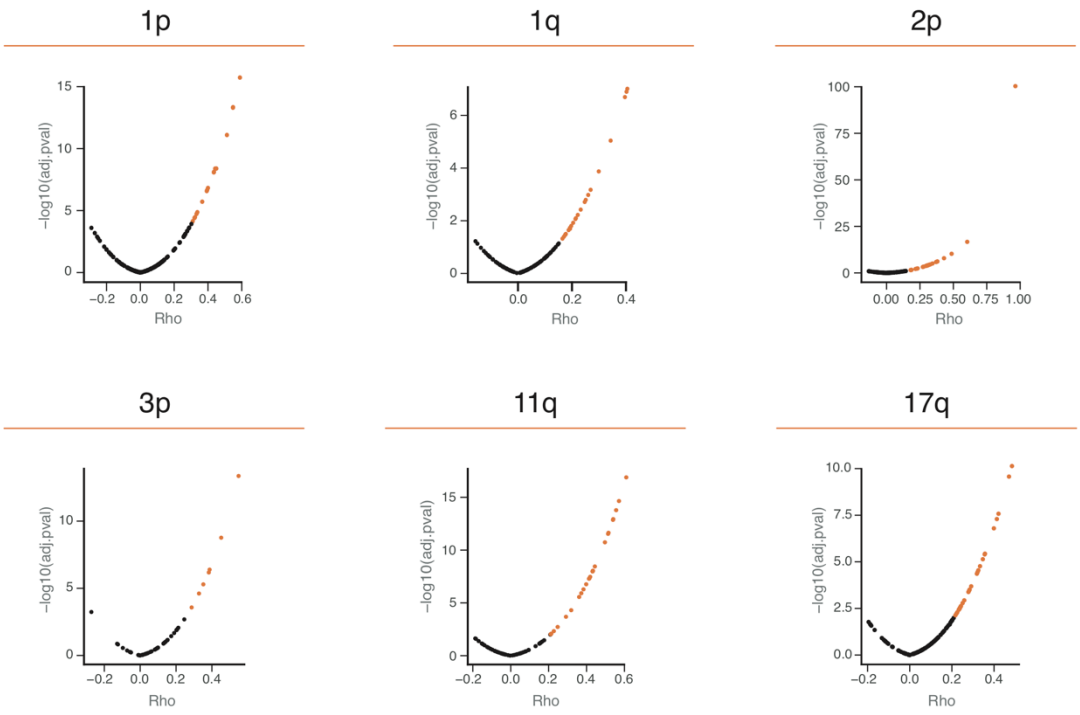
anticancer drug sensitivity. *Nature* **483**, 603 (2012).
62. Chiu, H.-S. *et al.* Pan-Cancer Analysis of lncRNA Regulation Supports Their Targeting of Cancer Genes in Each Tumor Context. *Cell Rep.* **23**, 297–312.e12 (2018).
63. Smiraglia, D. J. *et al.* Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. *Hum. Mol. Genet.* **10**, 1413–1419 (2001).
64. Paz, M. F. *et al.* A Systematic Profile of DNA Methylation in Human Cancer Cell Lines. *Cancer Res.* **63**, 1114 LP-1121 (2003).
65. Erkkilä, T. *et al.* Probabilistic analysis of gene expression measurements from heterogeneous tissues. *Bioinformatics* **26**, 2571–2577 (2010).
66. Li, B. *et al.* Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* **17**, 174 (2016).
67. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453 (2015).
68. Xiang, J.-F. *et al.* Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.* **24**, 513 (2014).
69. Yan-hua, L., Xiang-lei, L., Hong, L. & Jian-jun, W. Long noncoding ribonucleic acids maternally expressed gene 3 inhibits lung cancer tumor progression through downregulation of MYC. *Indian J. Cancer* **52**, 190–193 (2015).
70. Navarro, P. *et al.* Molecular coupling of Tsix regulation and pluripotency. *Nature* **468**, 457 (2010).
71. Zhong, X. *et al.* MIAT promotes proliferation and hinders apoptosis by modulating miR-181b/STAT3 axis in ox-LDL-induced atherosclerosis cell models. *Biomed. Pharmacother.* **97**, 1078–1085 (2018).
72. Chen, Y. & Stallings, R. L. Differential Patterns of MicroRNA Expression in Neuroblastoma Are Correlated with Prognosis, Differentiation, and Apoptosis. *Cancer Res.* **67**, 976 LP-983 (2007).
73. Zaatiti, H.; Abdallah, J.; Nasr, Z.; Khazen, G.; Sandler, A.; Abou-Antoun, T. J. Tumorigenic proteins upregulated in the MYCN-amplified IMR-32 human neuroblastoma cells promote proliferation and migration. *Int. J. Oncol.* 787–803. (2018). doi:https://doi.org/10.3892/ijo.2018.4236
74. Shah, M., Rennoll, S. A., Raup-Konsavage, W. M. & Yochum, G. S. A dynamic exchange of TCF3 and TCF4 transcription factors controls MYC expression in colorectal cancer cells. *Cell Cycle* **14**, 323–332 (2015).
75. Nozato, M.; Kaneko, S.; Nakagawara, A.; Komuro, H. Epithelial-mesenchymal transition-related gene expression as a new prognostic marker for neuroblastoma. *Int. J. Oncol.* 134–140 (2013). doi:https://doi.org/10.3892/ijo.2012.1684
76. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
77. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545 LP-15550 (2005).
78. Decaesteker, B. *et al.* TBX2 is a neuroblastoma core regulatory circuitry component enhancing MYCN/FOXM1 reactivation of DREAM targets. *Nat. Commun.* **9**, 4866 (2018).
79. Buske, F. A., Bauer, D. C., Mattick, J. S. & Bailey, T. L. Triplexator: Detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res.* **22**, 1372–1381 (2012).

80. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).

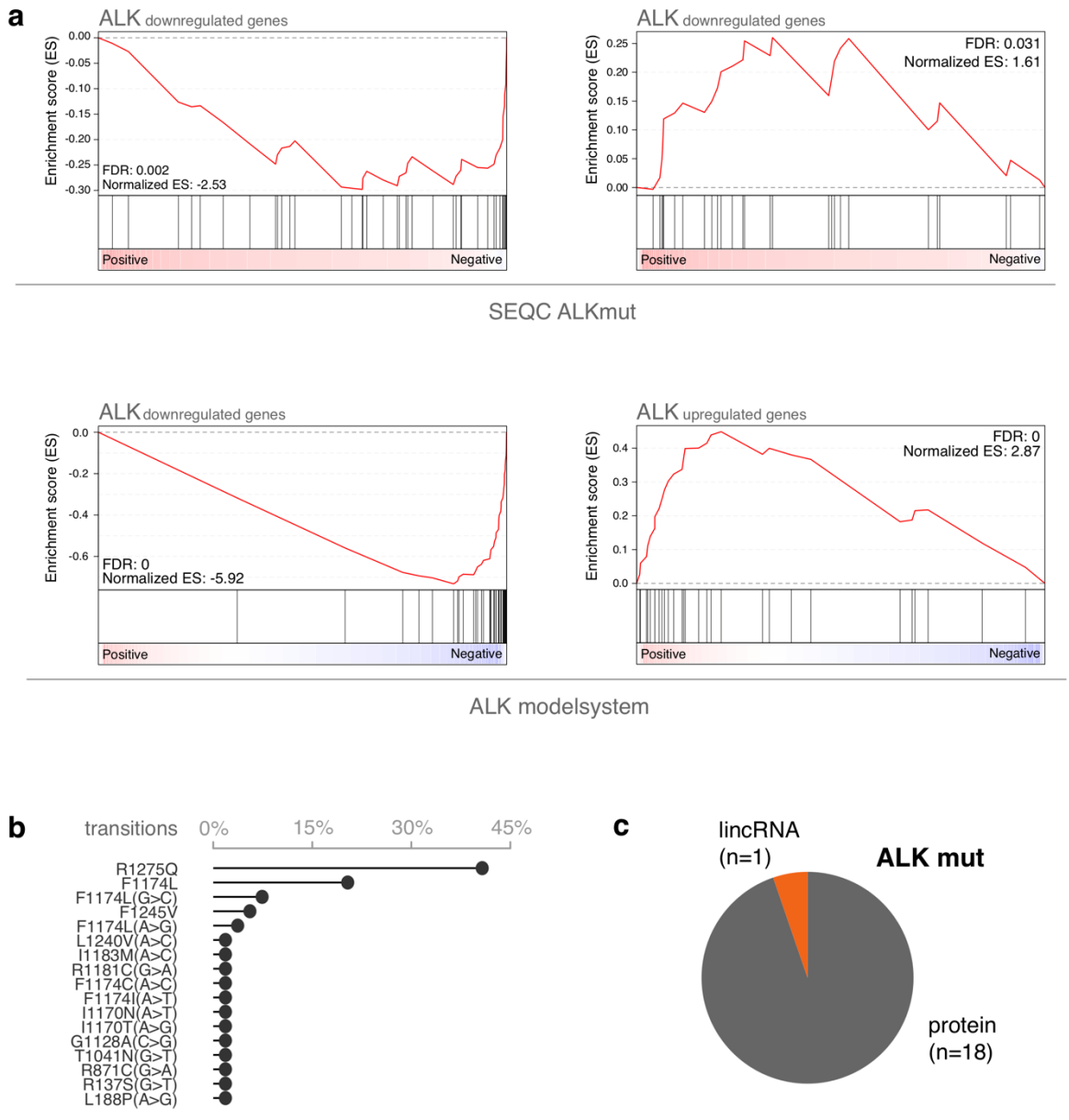
Supplementary information



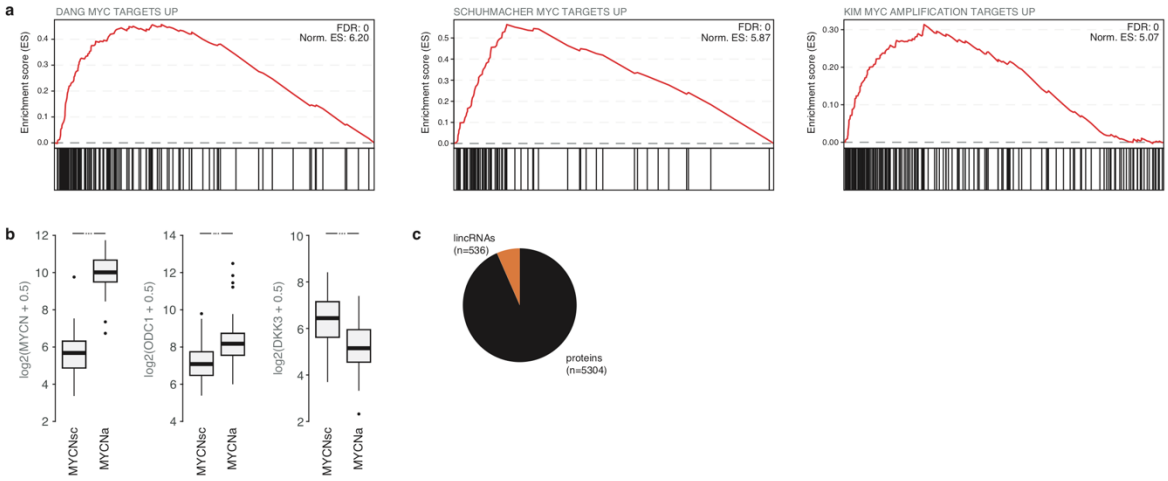
Supplemental Figure 1: Boxplots showing expression levels of *MIAT* and *MEG3* across different cancer cell lines



Supplemental Figure 2: Volcano plots with correlation coefficients and p-values for lincRNA expression and copy number variations, for lincRNAs situated on the respective chromosomal arm.

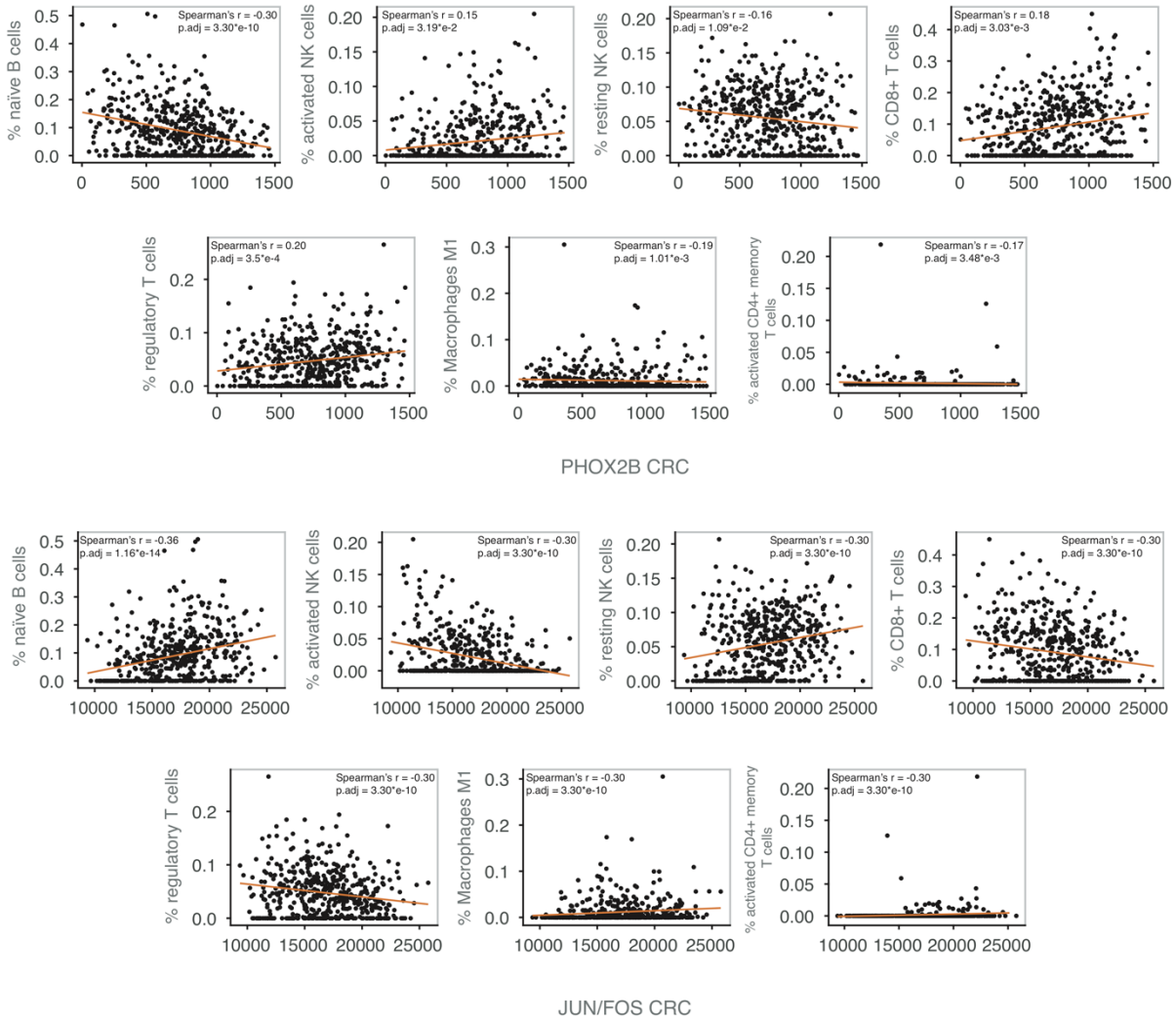


Supplemental Figure 3: a) GSEA plots for the differentially expressed genes between *ALK* mutated and wild samples and *ALK* treated and control samples, respectively. b) Percentage of samples harboring selected mutations found in the SEQC data. c) Number of differentially expressed lincRNAs and mRNAs after robust differential expression analysis.

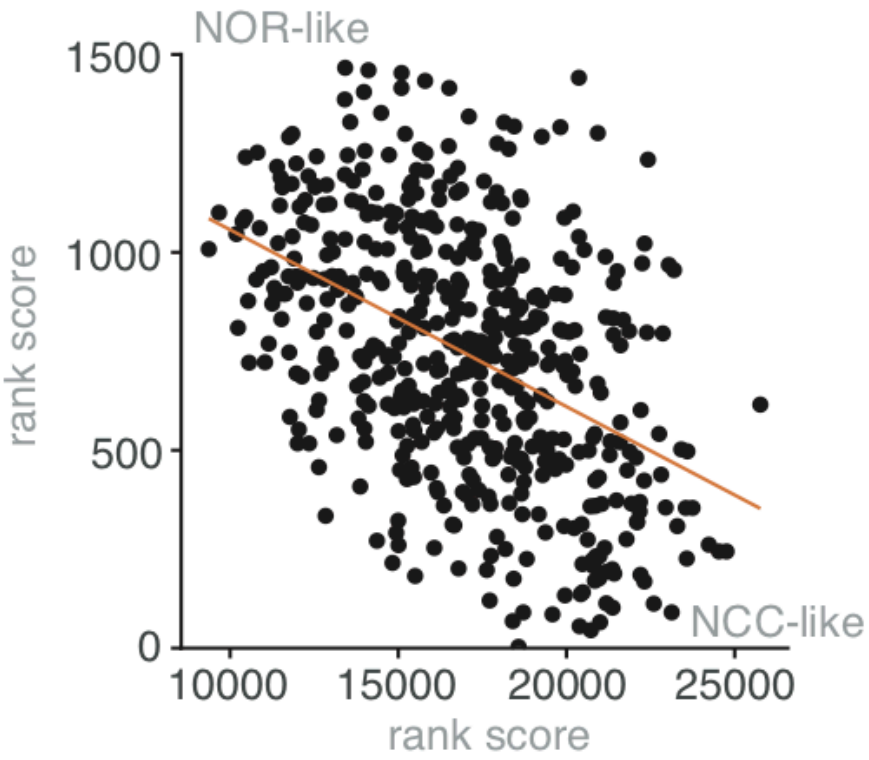


Supplemental Figure 4: a) GSEA showing enrichment for MYC(N) targets, after differential expression analysis between *MYCN* amplified and *MYCN* single copy samples. b) Boxplot displaying the expression levels of *MYCN* and two bona fide target genes, *ODC1* and *DKK3.* c) Number of lincRNAs and mRNAs robustly differentially expressed.

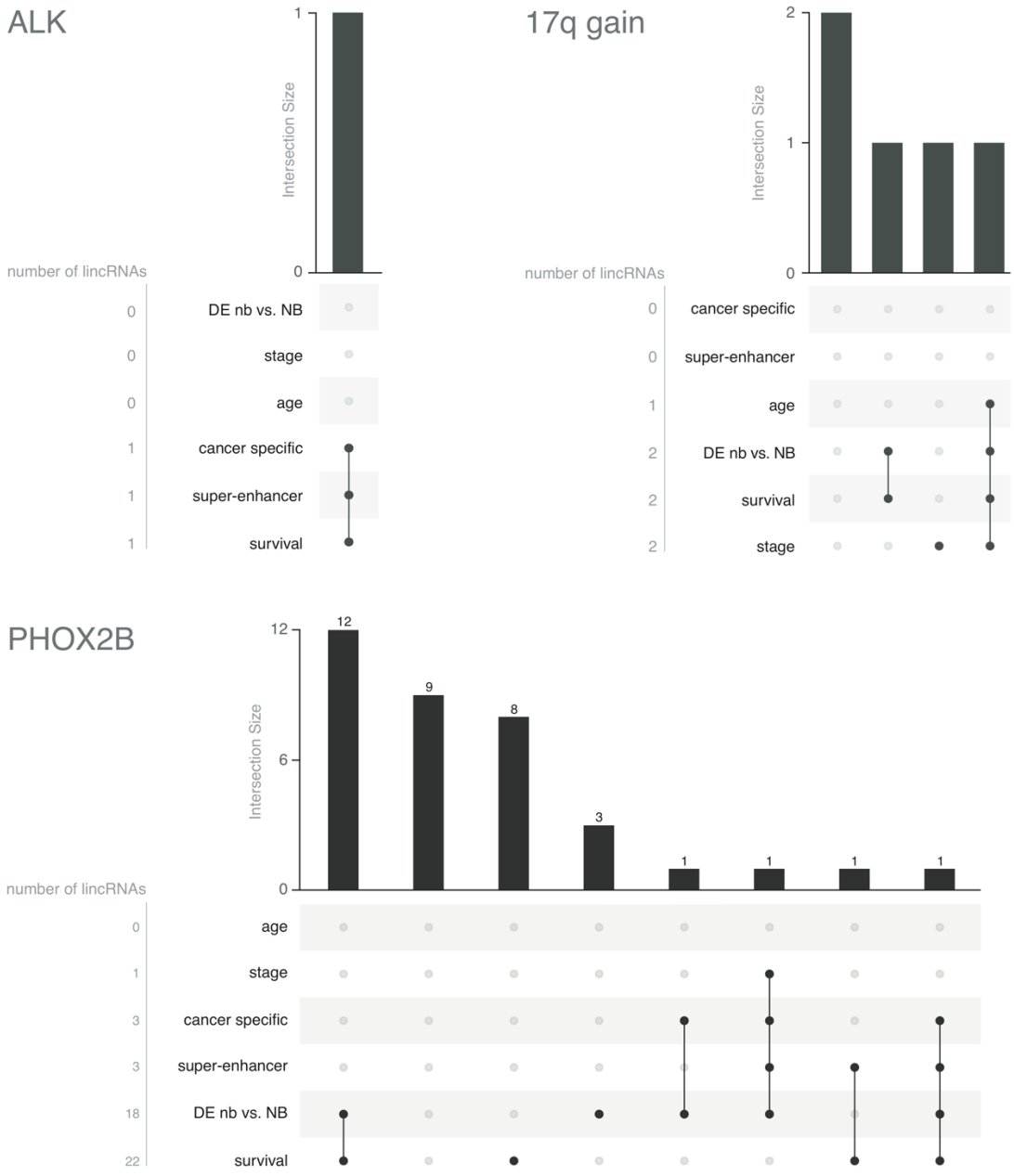
Supplemental Figure 5: The percentage of immune infiltration of 8 immune cell types between *MYCN* amplified and *MYCN* single copy tumors. Statistical significance was calculated using a Mann-Whitney test ($q<0.05$).



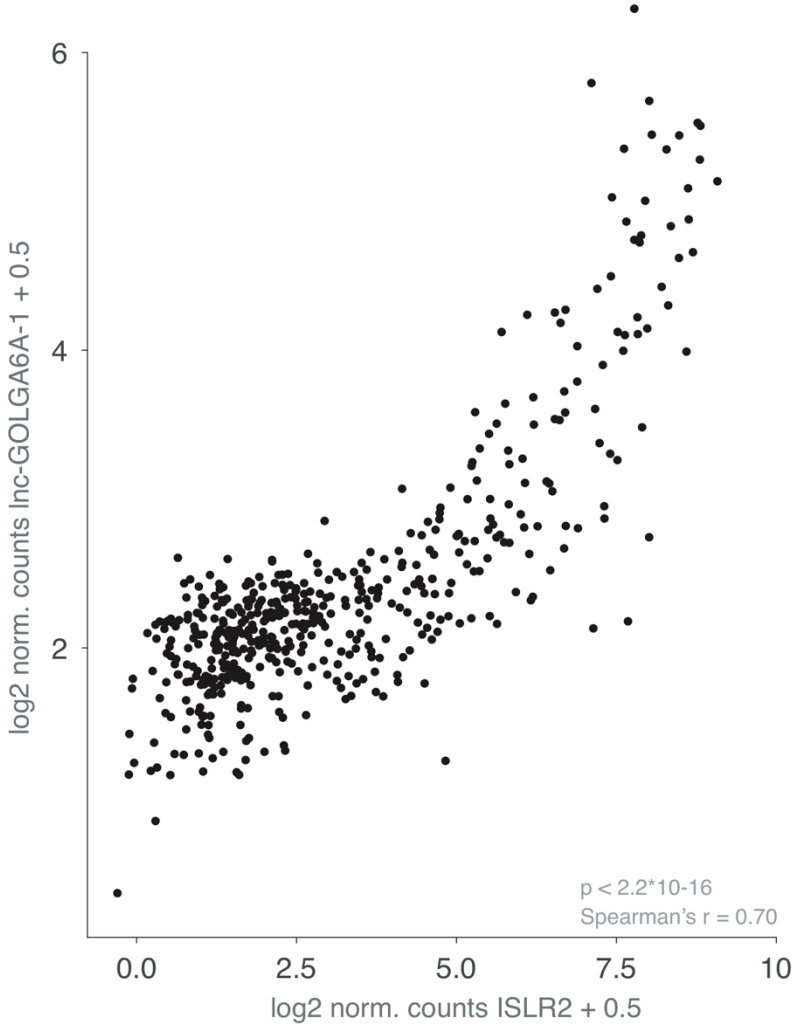
Supplemental Figure 6: Correlation plots of 7 immune cell types with significant correlation with the *PHOX2B* CRC or *JUN/FOS* CRC score. Correlation coefficient and p-value were determined using Spearman's method.



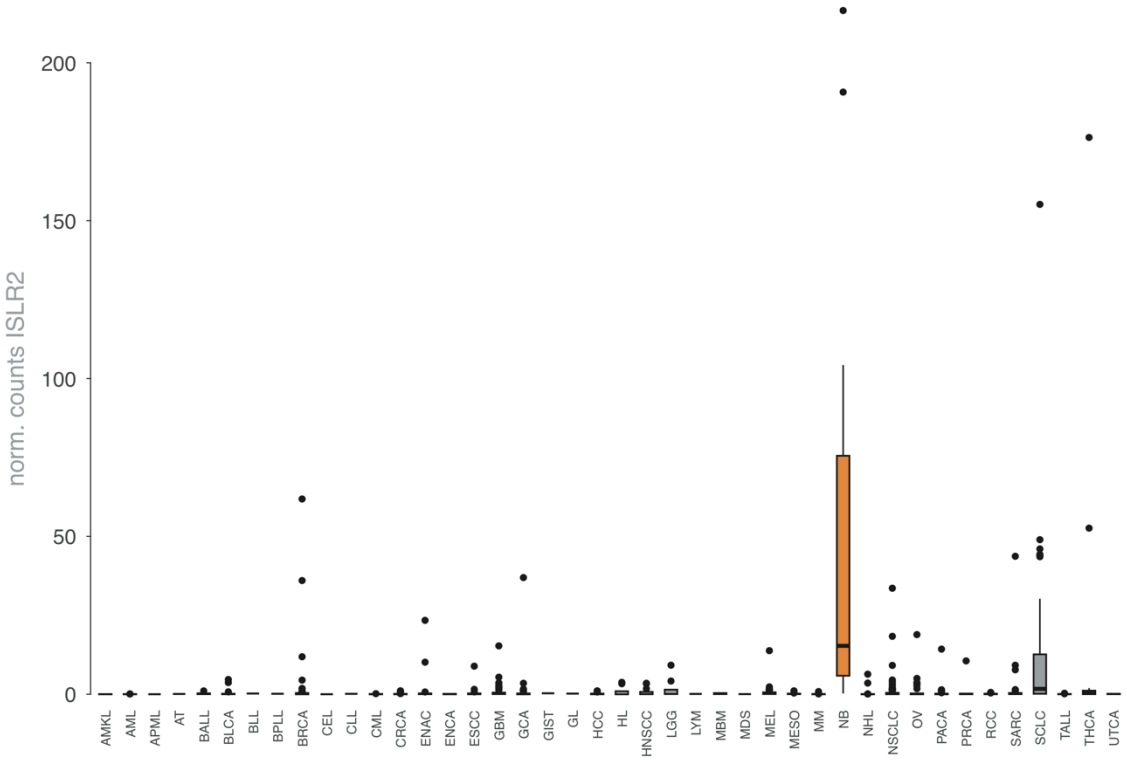
Supplemental Figure 7: Scatterplot visualizing the inverse correlation between the rank score calculated for the *PHOX2B* (NOR-like) and *JUN/FOS* CRC (NCC-like).



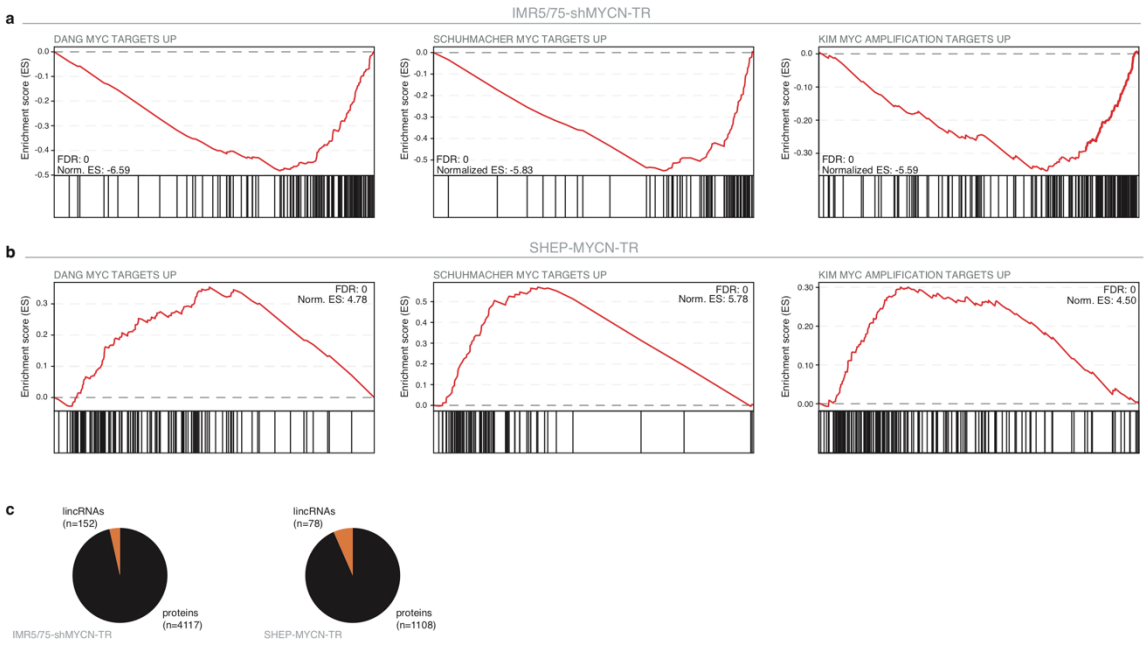
Supplemental Figure 8: Upset plots showing associations with clinical and genetic parameters for lincRNAs found to be regulated by *PHOX2B* or *ALK*, or influenced by a copy number gain of 17q.



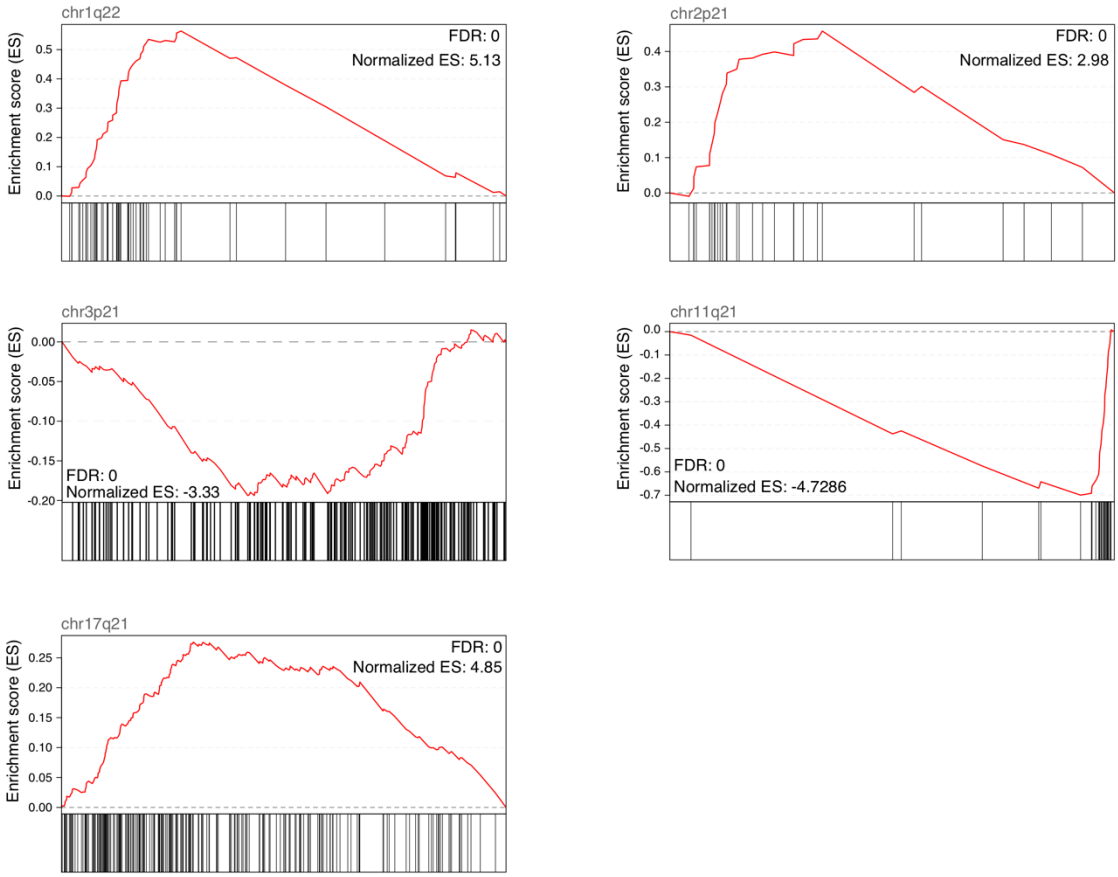
Supplemental Figure 9: Correlation plot of *lnc-GOLGA6A-1* and *ISLR2* expression. Calculation of the correlation coefficient and statistical significance was done using Spearman's method.



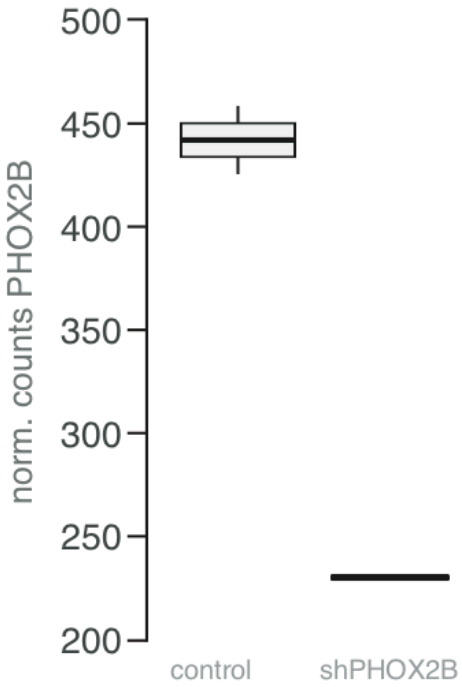
Supplemental Figure 10: Boxplot showing expression pattern of *ISLR2* across different cancer cell lines.



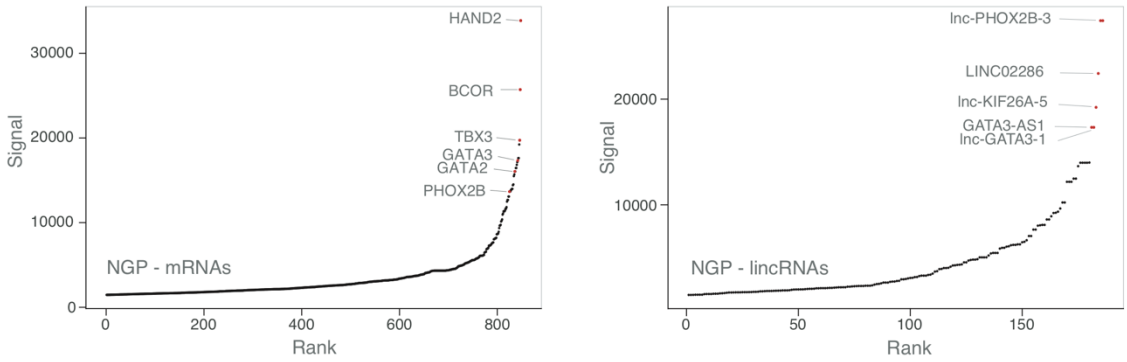
Supplemental Figure 11: a+b) GSEA plots of the *MYCN* model systems, showing for depletion and enrichment for three MYC gene sets upon treatment of IMR5/75-shMYCN-TR and SHEP-MYCN-TR, respectively. c) Number of lincRNAs and mRNAs differentially expressed in the model systems.



Supplemental Figure 12: GSEA plots showing enrichment/depletion for chromosomal arm gene sets.

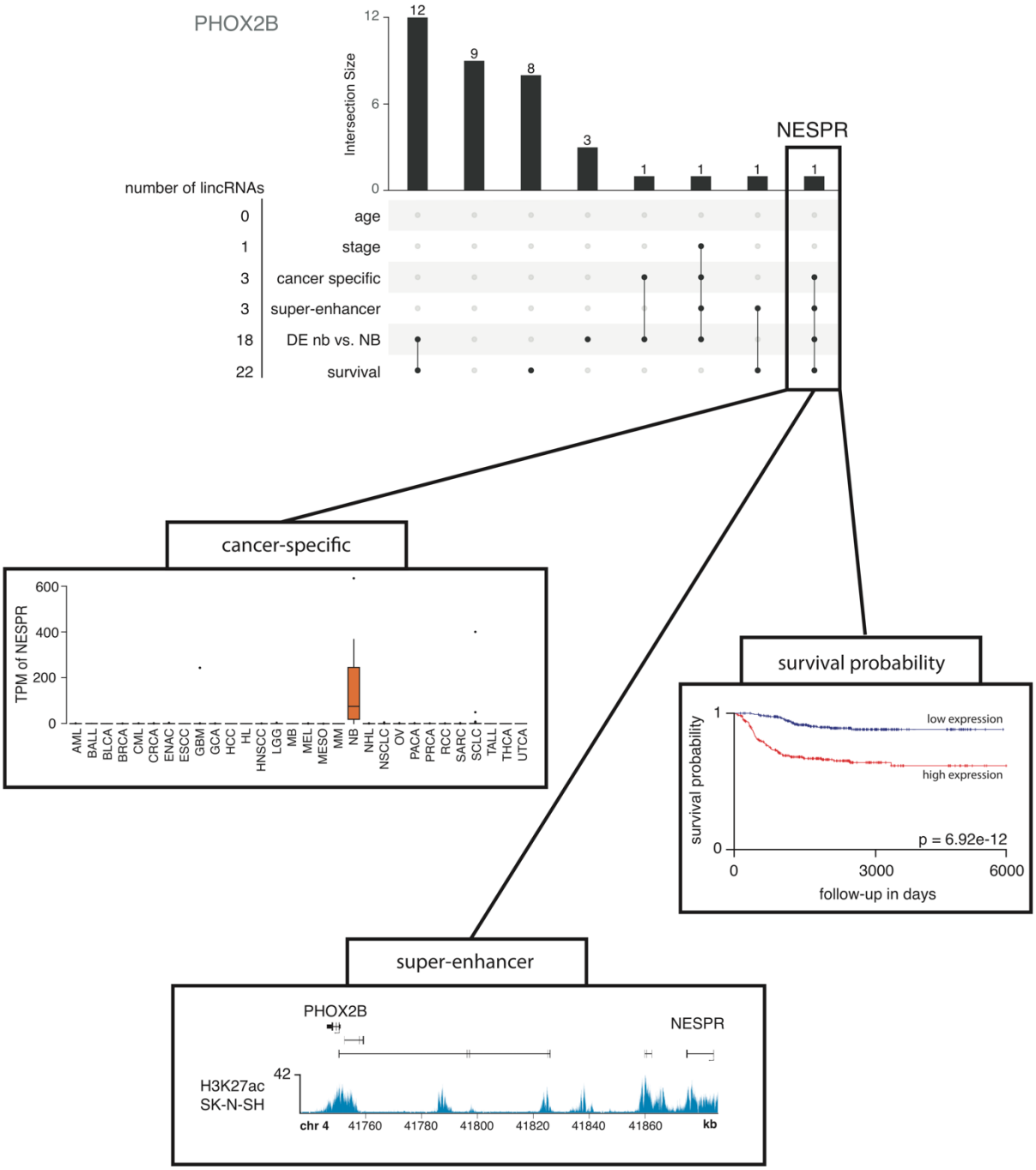


Supplemental Figure 13: Boxplot showing expression levels of *PHOX2B* in samples treated with doxycycline (shPHOX2B) or without doxycycline (control) of the CLB-GA cell line.



Supplemental Figure 14: Hockey stick plot of genes associated with super-enhancers in NGP for mRNAs and lincRNAs, respectively.

3.2 NESPR is a neuroblastoma-specific lincRNA downstream of PHOX2B



Our analyses result in a core set of lincRNAs implicated in neuroblastoma (paper 1). The top candidate, coined *NESPR*, is a lincRNA regulated by PHOX2B. *NESPR* shows cancer-specific expression, together with an association to survival probability, where a higher expression is associated with a worse outcome. Furthermore, the lincRNA is located in a super-enhancer, which are known to drive expression of cell identity genes. These criteria sparked our interest to functionally investigate the lincRNA in neuroblastoma biology.

3.3 **Research paper 2:** The neuroblastoma-specific lncRNA NESPR is a component of the noradrenergic core regulatory circuit and is essential for neuroblastoma cell survival

Dries Rombaut, Louis Delhaye, Eric J De Bony De Lavergne, Tiago França Brazao, Eva D'haene, Giorgio Milazzo, Roberto Ciaccio, Nurten Yigit, Agathe Peltier, Isabelle Janoueix-Lerosey, Celine Everaert, Matthias Fischer, Thorsten Simon, Johan Van Nes, Rogier Versteeg, Björn Menten, Giovanni Perrini, Jo Vandesompele, Sven Eyckerman, Steve Lefever, Pieter Mestdagh

Manuscript based on submitted patent (submission number: EP18154813.2; PCT: PCT/EP2019/051757), 2019

Contributions: The candidate contributed in part or in whole to design of the work, data acquisition and interpretation and the data analysis. Aside from the ChIRP-MS and ChIRP-sequencing and RNA fractionation experiments, the candidate performed most of the wet lab experiments. For the matching paragraphs in the manuscript, the candidate contributed to the interpretation of the data and the revision of the text. The remainder of the manuscript was drafted and revised by the candidate.

The neuroblastoma-specific lncRNA NESPR is a component of the noradrenergic core regulatory circuit and is essential for neuroblastoma cell survival

Dries Rombaut[1,2], Louis Delhaye[2,3], Eric J De Bony De Lavergne[1,2], Tiago França Brazao[1,2], Eva D'haene[1,2], Giorgio Milazzo[4], Roberto Ciaccio[4], Nurten Yigit[1,2], Agathe Peltier[4,5], Isabelle Janoueix-Lerosey[4,5], Celine Everaert[1,2], Matthias Fischer[8], Thorsten Simon[8], Johan Van Nes[7], Rogier Versteeg[7], Björn Menten[1,2], Giovanni Perrini[4], Jo Vandesompele[1,2], Sven Eyckerman[2,3], Steve Lefever[1,2], Pieter Mestdagh[1,2]

1. Cancer Research Institute Ghent (CRIG), Ghent, Belgium
2. Center for Medical Genetics, Ghent University, Ghent, Belgium
3. Medical Biotechnology Center, VIB, Albert Baertsoenkaai 3, Ghent, Belgium
4. Department of Pharmacy and Biotechnology (FaBiT), University of Bologna, Bologna, Italy
5. Institut Curie, PSL Research University, Inserm U830, Equipe Labellisée contre le Cancer, F-75005, Paris, France.
6. SIREDO: Care, Innovation and Research for Children, Adolescents and Young Adults with Cancer, Institut Curie, F-75005, Paris, France.
7. Department of Oncogenomics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands
8. Department of Pediatric Oncology and Hematology, University of Cologne, Köln, Germany

Abstract

Long non-coding RNAs (lncRNAs) form a novel class of RNA molecules that are often characterized by an exquisite tissue-specificity, making them extremely attractive as targets for therapeutic intervention. By reprocessing RNA-sequencing data from over 900 cancer cell lines, we identified several neuroblastoma-specific lncRNAs including *NESPR* (NEuroblastoma Specific Phox2B Regulatory rna). *NESPR* is located in the super-enhancer region upstream of the familial neuroblastoma gene *PHOX2B*. Unlike many lncRNAs, *NESPR* is abundantly expressed, efficiently spliced and highly conserved in mammals. *NESPR* expression specificity was confirmed in a cohort of over 10,000 tumor samples representing 34 cancer types. In neuroblastoma, high *NESPR* expression is associated with high stage disease, MYCN amplification and poor patient survival, suggesting *NESPR* may function as a lineage survival oncogene. To assess the function of *NESPR* in neuroblastoma, antisense oligonucleotides (ASOs) and siRNA pools (siPOOLs) were used to knock down its expression. While ASOs were capable of reducing both the nuclear and cytoplasmic fraction of *NESPR*, siPOOLs only reduced the cytoplasmic fraction. Notably, knockdown of the nuclear fraction, but not the cytoplasmic

fraction, resulted in a significant decrease in colony formation and cell growth, as evidenced by cell viability assays and real-time cell monitoring. These effects were accompanied by an increase in apoptosis and were validated using independent ASO sequences. RNA-sequencing of ASO-treated neuroblastoma cell lines revealed a significant reduction of several neuroblastoma master regulators including *PHOX2B*, *PHOX2A*, *DACH1* and *ZNF536* while expression of *CHD5* was significantly induced. Using 4C-sequencing we could demonstrate a long-range interaction between the *NESPR* locus and the *PHOX2B* promoter, suggesting that the nuclear fraction of *NESPR* acts as a cis-regulator of *PHOX2B* expression. Silencing of *NESPR* transcription using CRISPR-interference verified this interaction. *NESPR* pulldown followed by DNA sequencing (ChIRP-seq) revealed binding of NESPR to genomic regions also bound by *PHOX2B, GATA3 and ISL1*, two transcription factors of the noradrenergic core regulatory circuit. Our results uncover *NESPR* as a key component in the transcriptional circuit defining neuroblastoma cell identity and suggest that ASO-mediated targeting of NESPR may present a novel neuroblastoma-specific treatment option. *In vivo* validation of the observed phenotype is currently ongoing.

Introduction

Neuroblastoma (NB) is a rare, aggressive childhood cancer, affecting around 1 per 100,000 children under 15 years[1]. Arising from an aberrant differentiation of precursor cells from the sympathetic nervous system, the tumor can present itself all along the peripheral nervous system, with the highest incidence in the abdomen. Although NB is rare, it accounts for 15% of deaths caused by cancer in children, with overall survival rates less than 40% for high risk NB patients[1–3]. Most efforts to elucidate the genetic heterogeneity of NB focused on protein coding genes, resulting in the identification of key genetic events such as amplifications of *MYCN*[4], rearrangements of TERT[5], mutations of *ALK*[6] and in a small percentage of cases, mutations of *PHOX2B*[7–9], a developmental regulator of neural crest cells.

Recently, two independent studies revealed the presence of two cellular identities with differing gene expression profiles in primary neuroblastoma tumors[10,11]. These cellular identities were identified based on divergent epigenetic profiles, in particular super-enhancers, and are defined by core regulatory circuitries (CRC) containing multiple super-

enhancer associated transcription factors. The sympathetic noradrenergic cell type is defined by a CRC containing *PHOX2B*, *HAND2* and *GATA3* whereas the undifferentiated neural crest cell-like cell type is driven by a CRC composed of AP-1 transcription factors, among which proteins from the JUN or FOS families. The control of the epigenetic landscape of neuroblastoma cells and their intratumoral heterogeneity is mediated by these two networks, which determine cell lineage in normal development[10,11]. Most tumors belong to the noradrenergic identity and the transcription factors driving this identity typically show a neuroblastoma-specific expression pattern.

Large-scale sequencing efforts have demonstrated that our genome is pervasively transcribed, with only ~1.2% giving rise to protein coding genes[12]. The majority of the transcriptome consists of non-coding RNA transcripts, with long non-coding RNAs (lncRNAs) as the largest group[13]. LncRNAs show high tissue-specificity[14] and are implicated in all cancer hallmarks[15,16]. They play a role in various biological processes and function as regulators of chromatin architecture, gene expression, splicing and translation[17]. Deregulated expression of long non-coding RNAs has been implicated as a cause of tumor initiation and development[15,18]. Insights in the role of lncRNAs in neuroblastoma are limited, yet for some, oncogenic or tumor suppressor functions have been demonstrated[19]. Notable examples are *LINC00467*[20] and *lncUSMycN*[21] that modulate or regulate MYCN target genes or MYCN expression, and *NBAT-1*[22], which functions as tumor suppressor by inhibiting proliferation and invasion. A role for lncRNAs in the neuroblastoma CRCs has not yet been described.

Here, we present *NESPR*, a highly conserved neuroblastoma-specific lncRNA regulating *PHOX2B*. We provide evidence that *NESPR* has a role in the noradrenergic CRC and demonstrate that ASO-mediated downregulation of *NESPR* impairs cell growth and induces apoptosis.

NESPR is a conserved neuroblastoma-specific lincRNA

To prioritize lncRNAs with a potential role in neuroblastoma identity, we first identified lncRNAs with a neuroblastoma-restricted expression pattern. We reanalyzed RNA sequencing data from 934 cancer cell lines representing 42 tumor types, including 15 NB cell lines[23]. In this RNA-seq data repository, 9862 lncRNAs (5381 lincRNAs and 4481 antisense RNAs) are

robustly expressed, with more than five counts in half of all the cell lines per cancer type. Expression specificity of mRNAs and lncRNAs was calculated by means of a specificity score based on the Jensen-Shannon (JS) divergence metric[14]. Several mRNA transcripts demonstrated high specificity for neuroblastoma cell lines, including the known noradrenergic neuroblastoma specific genes *PHOX2B*[24] and *PHOX2A*[25]. *PHOX2A* and *PHOX2B* showed the highest specificity in neuroblastoma cells, followed by *LINC00682*, a long intergenic non-coding RNA we will further refer to as *NESPR* (NEuroblastoma Specific PHOX2B Regulatory RNA) (Fig. 1a). To confirm the neuroblastoma-specific expression pattern, we combined RNA-sequencing data of 497 primary neuroblastoma tumor samples (SEQC)[26,27] with RNA-sequencing data from The Cancer Genome Atlas (TCGA), representing 11,377 samples across 34 tumor types. *NESPR* expression was restricted to neuroblastoma, pheochromocytoma and paraganglioma, all originating from the same lineage[28] (Fig. 1b).

We next evaluated *NESPR* expression in relation to clinical and genetic parameters of neuroblastoma tumors. *NESPR* expression levels were significantly higher in MYCN amplified tumor samples (MYCNa) compared to samples with single haploid *MYCN* copy numbers (MYCNsc) (Fig 1c. adj. p-val<0.05). In addition, we observed a significant association between *NESPR* expression and overall survival of patients, with a higher expression profile implying a significant lower chance of survival (p-value = 6.92e-12, Fig. 1c).

NESPR consists of multiple transcripts, with the longest and most abundant transcript being 555 nucleotides long, located 122 kb downstream of *PHOX2B* on chromosome 4. The *PHOX2B-NESPR* locus coincides with a super-enhancer, illustrated by multiple H3K27ac peaks (Fig. 1d). The annotated *NESPR* transcription start site coincides with a H3K4me3 chromatin mark, indicative for active transcription. In addition, CAGE-sequencing data from NB cell lines[29] (CHP-134, NB-1, NH-12 and NBsusSR) demonstrates the presence of a peak at the *NESPR* TSS, confirming *NESPR* is independently transcribed (Fig. 1e). In contrast to many human lncRNAs, the *NESPR* transcript is highly conserved, with a sequence conservation of 84% over 46 vertebrate species (Fig. 1e). RNA-seq data from murine TH-MYCN neuroblastoma tumors validated the expression of an orthologous two-exon *NESPR* transcript (Supplementary Fig. 1).

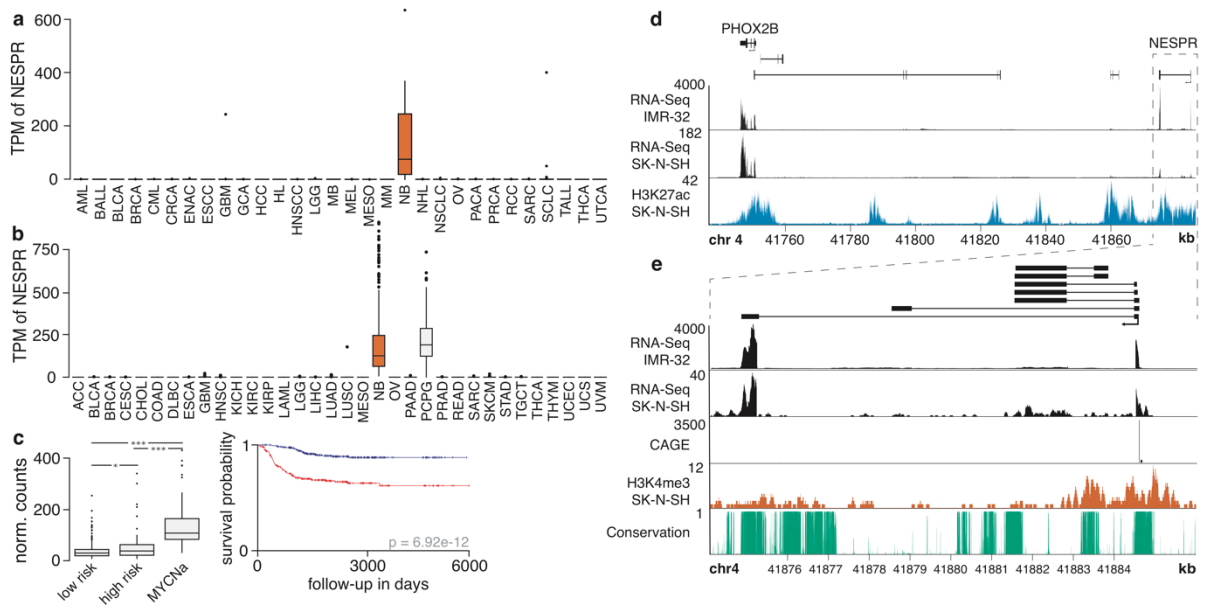


Figure 1: *NESPR* is a neuroblastoma specific lncRNA located in the *PHOX2B* locus
a) Expression (TPM) of *NESPR* across different cancer types in the CCLE RNA-seq data repository. b) Expression of *NESPR* (TPM) across different cancer types in the TCGA RNA-seq data repository complemented with RNA-seq data from neuroblastoma tumors. c) MYCNa tumors show significant higher expression of *NESPR* compared to low- and high-risk tumors. Higher levels of *NESPR* are indicative of lower chances of overall survival (red: high *NESPR* expression, blue: low *NESPR* expression). d) *NESPR* is located in the *PHOX2B* locus and resides in a super-enhancer as marked by H3K27ac. e) *NESPR* is two-exon transcript, independently transcribed and efficiently spliced, with high sequence conservation across mammals.

The mouse transcript covers 85% of its human counterpart with 96% sequence conservation. Notably, zebrafish neuroblastoma tumors also express a positionally conserved two-exon *NESPR* orthologue which did not show any sequence conservation to the human transcript (Supplementary Fig. 1). To establish the subcellular localization of *NESPR*, we applied RNA *in situ hybridization*. Visualization of the lncRNA in SHEP and SKNBE(2c) cell lines using RNAscope shows the lack of *NESPR* expression in the neural crest cell like cell line SHEP, whereas a clear presence of *NESPR* in SKNBE(2c) is demonstrated (Fig. 2a-b). Furthermore, RT-qPCR on RNA isolated from murine neural crest cells (NCC) and tumors confirms the absence of *NESPR* and *PHOX2B* expression in these NCCs, whereas the tumors exhibit expression of both genes (Fig. 2c). *NESPR* is expressed in both the nucleus and cytoplasm, which we confirmed through public cellular fraction RNA-seq data (data not shown) in SK-N-SH[30] and in-house qPCR-data of RNA fractionation assays in SKNBE(2c) (Supplemental Fig. 2).

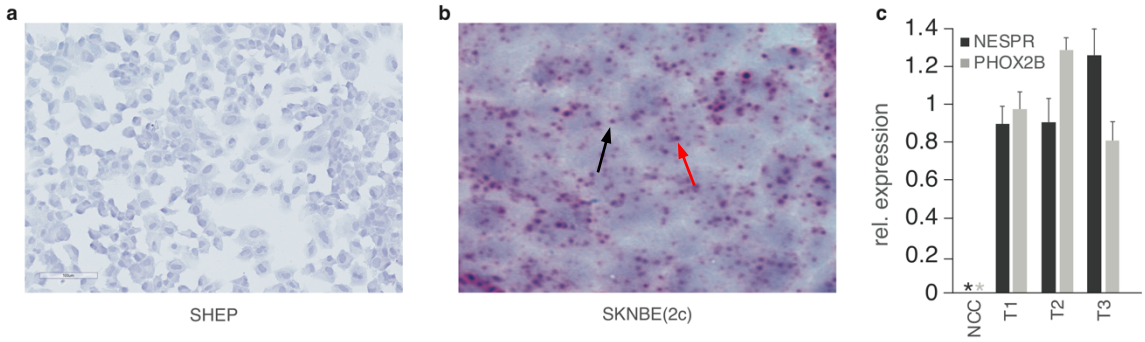


Figure 2: *NESPR is located in both the cellular and nuclear fraction*
a) *NESPR* localization in SHEP, a *NESPR*-negative cell line, using RNAscope confirms the lack of *NESPR* expression. b) *NESPR* is located in both cellular fractions of SKNBE(2c), visualized by RNAscope. The nuclei are stained with DAPI (blue), the cytoplasm is lighter. Purple dots show the location of *NESPR* in both the DAPI stained nuclei (red arrow) and the lighter space between them, i.e. the cytoplasm (black arrow). c) Relative expression values of *NESPR* and *PHOX2B* quantified by RT-qPCR. RNA was extracted from three murine neuroblastoma tumors, with neural crest cells (NCC) as a negative control.

NESPR regulates *PHOX2B* expression in cis

LincRNAs are known to be involved in cis-regulation of neighboring genes. As *NESPR* is located just downstream of *PHOX2B*, we hypothesized *NESPR* may regulate *PHOX2B* expression in cis. This hypothesis was supported by a strong and significant correlation between *NESPR* and *PHOX2B* expression in neuroblastoma cell lines (Spearman's r = 0.96, p<2.2e-16) and primary tumor samples (Spearman's r = 0.44, p<2.2e-16) (Fig. 3a). Furthermore, a recent study from our lab showed that *NESPR* expression is correlated with the *PHOX2B* CRC[31]. To further validate the potential cis-regulatory action of *NESPR*, we treated 2 neuroblastoma cell lines with *NESPR*-specific antisense oligonucleotides, significantly reducing *NESPR* expression by at least 55**%** (Fig. 3b). Knockdown of *NESPR* lead to a decrease in *PHOX2B* levels, as quantified by RNA-sequencing. Moreover, we identified another 780 genes that were significantly differentially expressed in both cell lines and with both *NESPR* ASOs (|logFC|>0.5, q-value<0.05) (Fig. 2c). Together with *PHOX2B*, several noradrenergic cell identity genes including *PHOX2A*, *DACH1* and *ZNF536*[11] were also downregulated. In contrast, the neural crest cell identity genes[10,11] *FOS*, *APOE* and *ERG1* were significantly upregulated.

Next, we performed gene set enrichment analysis (GSEA) to determine the pathways in which *NESPR* is involved. Several gene sets involved in neuronal development, embryogenesis and epithelial-to-mesenchymal transition (EMT) showed positive enrichment with upregulated genes after *NESPR* knockdown, including the EGF, HOXA5 and NRG1 gene sets, the latter containing upregulated targets after NRG1 overexpression, a protein implicated in neuronal differentiation, survival and migration (Supplementary Fig. 3). Also *CHD5,* a neuroblastoma tumor suppressor gene implicated in normal neuronal development, cell proliferation and differentiation[32], was found to be upregulated after *NESPR* KD. When performing GSEA using mesenchymal/neural crest cell-like and noradrenergic gene sets[11], a clear depletion of the noradrenergic gene signature can be observed (Supplementary Fig. 4). However, we were unable to discern a consistent, positive enrichment of the mesenchymal gene sets.

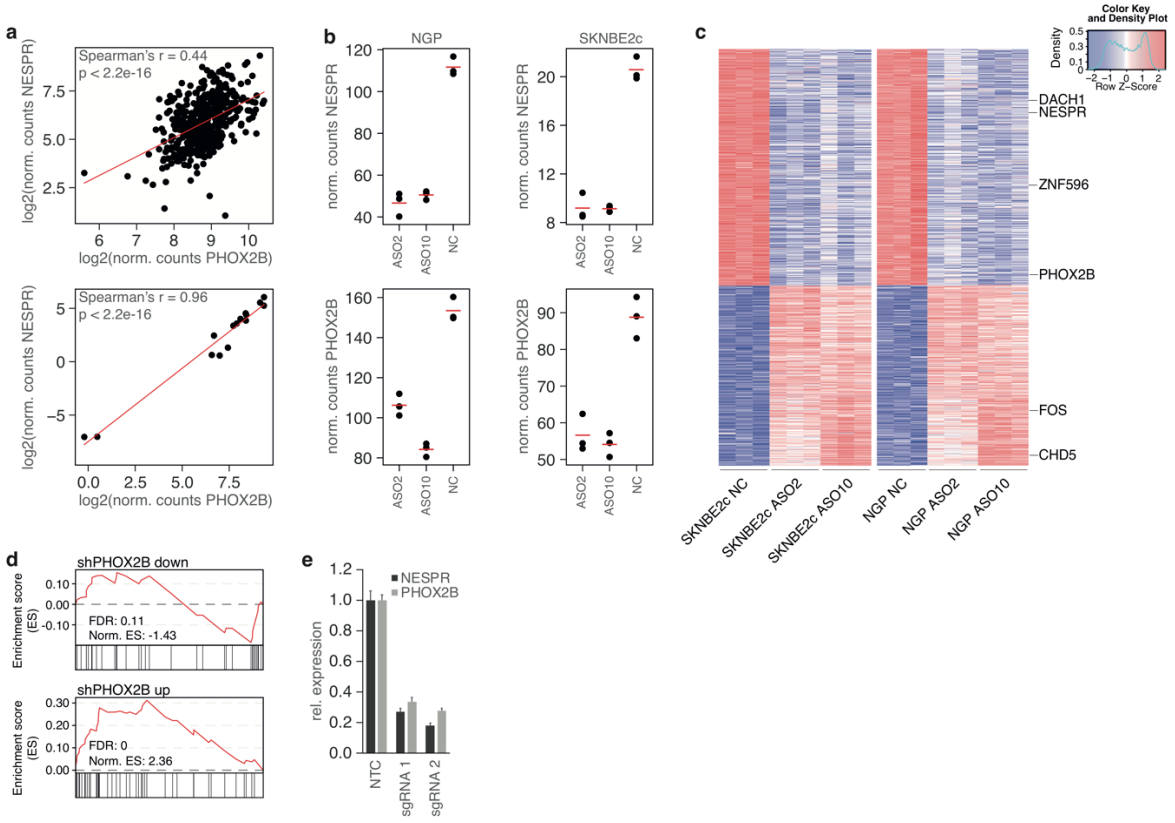


Figure 3: *NESPR* regulates *PHOX2B* transcription
a) Correlation between *PHOX2B* levels and *NESPR* abundance can be observed in primary tumor samples (spearman's r = 0.44, p<2.2e-16), as well as in neuroblastoma cell lines (spearman's r = 0.96, p<2.2e-16). b) Loss of *NESPR* results in downregulation of *PHOX2B* in NGP and SKNBE(2c) using two ASOs. c) ASO-mediated downregulation of *NESPR* results in differential expression of 782 genes (|logFC|>0.5, q-value < 0.05). d) GSEA, using two in-house generated gene sets based on knockdown of *PHOX2B*, shows significant positive enrichment with the upregulated genes after *PHOX2B* downregulation and borderline significant negative enrichment with downregulated *PHOX2B* targets. e) CRISPRi, using two sgRNAs, results in similar expression patterns as ASO-mediated knockdown.

To validate the impact of *NESPR* KD on *PHOX2B* expression, we used a *NESPR* siPOOL as an alternative method to silence *NESPR* expression. While the siPOOL resulted in knockdown of *NESPR*, there was no apparent effect on *PHOX2B* expression nor any other gene (Supplemental Figure 2 and data not shown). We therefore speculated that the cis-regulatory action of *NESPR* was mediated by nuclear *NESPR* transcripts and that the siRNAs in the siPOOL were only affecting cytoplasmic *NESPR* expression. RNA fractionation of ASO and siPOOL transfected cells followed by quantitative polymerase chain reaction with reverse transcription (RT-qPCR) demonstrated that *NESPR* ASOs reduced both cytoplasmic and nuclear *NESPR* transcripts, whereas the *NESPR* siPOOL only reduced cytoplasmic *NESPR* (Supplementary Figure 2). In line with previous findings, *PHOX2B* expression was only altered after ASO transfection, whereas the siPOOL did not affect *PHOX2B* levels. We therefore used CRISPR inactivation (CRISPRi) to silence *NESPR* transcription, with two sgRNA designs targeting the *NESPR* TSS. Transfection of the sgRNAs in SKNBE(2c) cells with inducible dCAS9-KRAB expression resulted in a knockdown of both *NESPR* and *PHOX2B*, further validating the cis-regulatory function of *NESPR* (Fig. 3e).

NESPR and PHOX2B are part of a noradrenergic-specific insulated neighborhood

To further evaluate the mechanisms underlying the cis-regulatory function of *NESPR*, we evaluated the chromatin conformation of the *NESPR-PHOX2B* locus. Using publicly available HiC-data in the neuroblastoma cell line SK-N-DZ, we identified interactions between the *PHOX2B* and *NESPR* loci, suggesting that both genes are embedded in an insulated neighborhood (Fig. 4). To further validate the chromatin loop structure of this insulated neighborhood, we performed Chromosome Conformation Capture on Chip sequencing (4C-seq)[33] with viewpoints in the *NESPR* or *PHOX2B* promoter. Both viewpoints revealed an identical interaction pattern that was validated in 2 *NESPR* positive neuroblastoma cell lines. Chromatin contacts were overlapping H3K27ac sites and binding sites for CTCF and the cohesion subunits RAD21 and SMC3, as evidenced by ChIP-sequencing data.

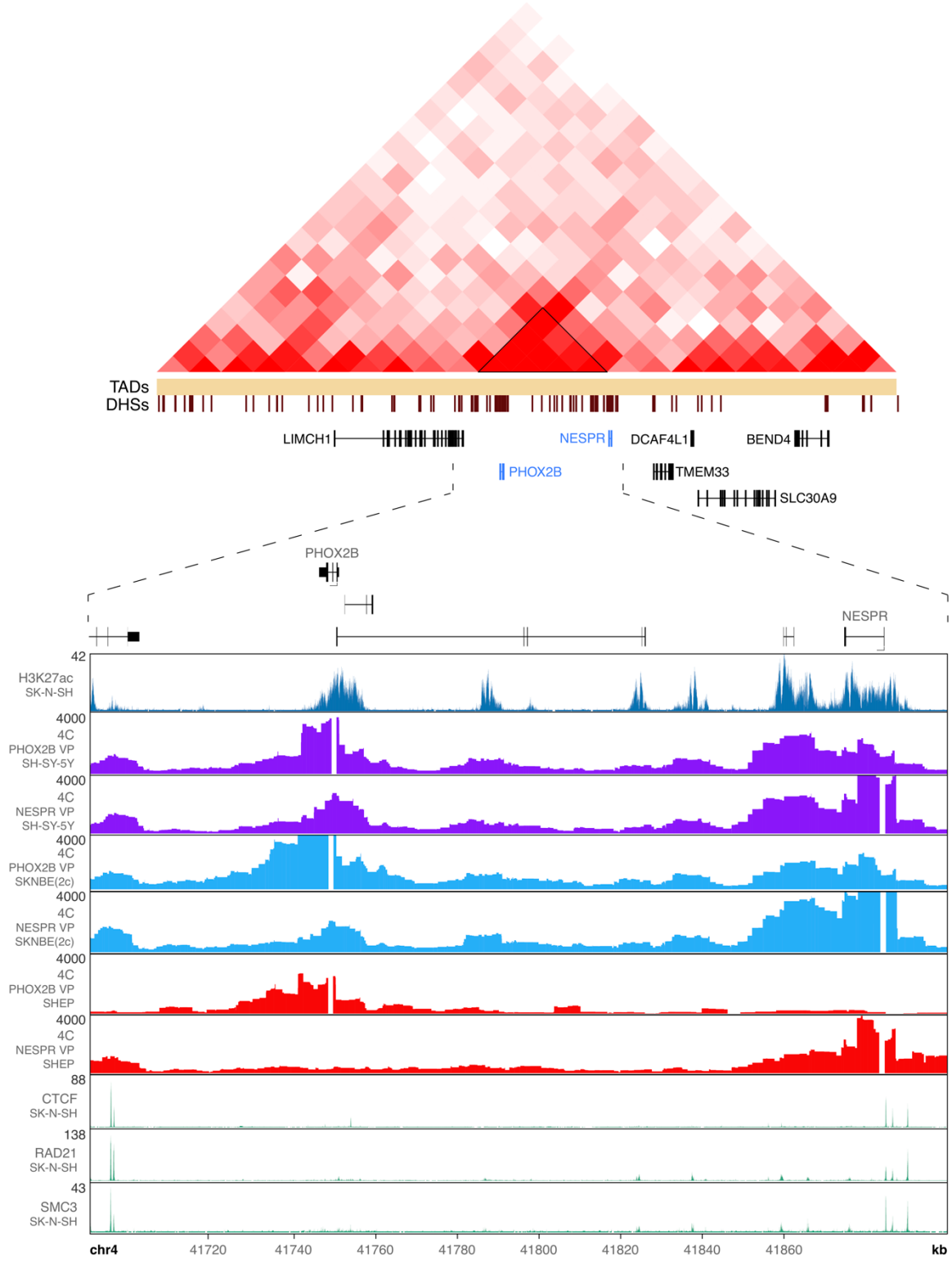


Figure 4: *NESPR* is involved in formation of an insulated neighborhood
Hi-C data of an extended genomic region shows an insulated neighborhood encompassing *NESPR* and *PHOX2B*. Through 4C-sequencing, a chromatin loop could be distinguished in SKNBE(2c) and NGP, whereas in a *NESPR* negative cell line SHEP, the chromatin structure was not present. Regional binding of CTCF and SMC3 and RAD21, two cohesin subunits, further corroborates the presence of an insulated neighborhood.

Notably, the neural crest like cell line SHEP, which is negative for both *PHOX2B* and *NESPR*, did not show evidence for chromatin looping, suggesting that the *NESPR-PHOX2B* insulated neighborhood is confined to neuroblastoma cells with a noradrenergic identity (Fig. 4).

NESPR and PHOX2B are involved in a feed-forward loop

The effect of *NESPR* KD on the expression levels of *PHOX2B* and the involvement of *PHOX2B* in the CRC of the noradrenergic lineage suggests a potential role for *NESPR* in this CRC. As CRC components typically regulate each other's activity to create a feed-forward loop, we evaluated if PHOX2B could also regulate *NESPR* expression. We therefore generated RNA-seq data of a cellular *PHOX2B* perturbation model system, with an inducible shRNA construct against *PHOX2B* (CLB-GA-shPHOX2B). Next to a significant downregulation of *PHOX2B* (logFC = -0.96, p<7.2*10^-9), we also observed a significant reduction in *NESPR* expression (logFC = -0.5, p<0.003) (Fig. 5a). In addition, PHOX2B ChIP-seq data demonstrates PHOX2B binding to the *NESPR* promoter, further substantiating the existence of a feed-forward loop between *NESPR* and *PHOX2B* (Fig. 5b).

NESPR can function in trans to regulate gene expression independent of *PHOX2B*

To evaluate if *NESPR* can regulate gene expression independent of *PHOX2B*, we integrated RNA-seq data from the *NESPR* knockdown and *PHOX2B* knockdown experiments. Two PHOX2B gene sets were created representing upregulated genes and downregulated genes after *PHOX2B* knockdown, respectively. The gene expression profiles of the *NESPR* knockdown experiments showed significant enrichment for the upregulated genes after *PHOX2B* knockdown. Interestingly, when calculating enrichment with the genes downregulated upon *PHOX2B* knockdown, we only reached borderline significance, suggesting that *NESPR* may also function independent of *PHOX2B* by regulating gene expression in trans (Fig. 3).

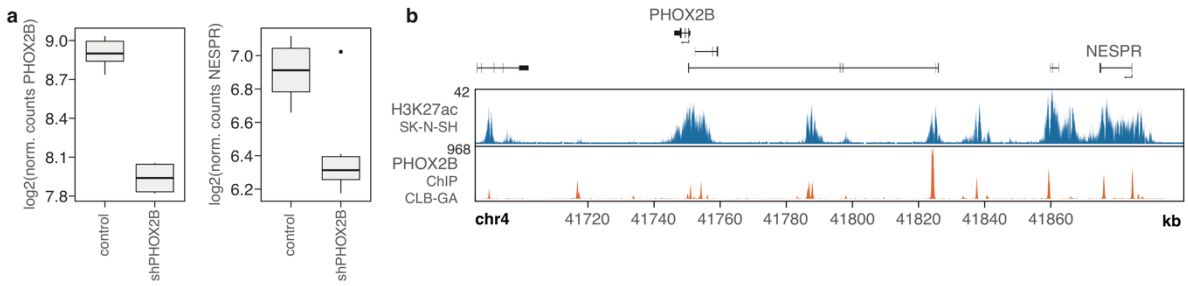


Figure 5: *PHOX2B* regulates *NESPR* transcription
a) Knockdown of *PHOX2B* in an inducible cellular perturbation model system leads to *NESPR* downregulation. b) Transcriptional regulation of *NESPR* through *PHOX2B* seems to be direct, as several binding sites of PHOX2B were detected in the genomic region, overlapping with H3K27ac site.

To further evaluate this hypothesis, we applied chromatin isolation by RNA purification[34] followed by DNA-sequencing (ChIRP-seq) to identify potential *NESPR* binding sites. ChIRP-seq was performed on the noradrenergic neuroblastoma cell line IMR-32 with a *NESPR*-specific probe set and a probe set designed against LacZ as a negative control. Enrichment of *NESPR* transcripts in the *NESPR* pull down compared to LacZ pull down was validated using RT-qPCR (Supplemental Fig. 5). DNA-sequencing of the isolated chromatin fragments revealed 3220 DNA binding sites that were significantly enriched in the *NESPR* pull down compared to the LacZ pull down. HOMER motif analysis[35] of these candidate *NESPR* binding sites uncovered enrichment of DNA motifs for several transcription factors, including GATA3 and ISL1[10], two members of the noradrenergic CRC. ChIP-seq data for GATA3 and ISL1 revealed that a subset of *NESPR* binding sites were indeed bound by GATA3 and ISL1. Moreover, several of these binding sites were also bound by PHOX2B itself.

In total, 84 loci showed binding of both PHOX2B, GATA3 and ISL1, whereas in 232, 169 and 298 cases *NESPR* binding overlapped with PHOX2B, GATA3 and ISL1 binding, respectively (Fig. 6b-c). Interestingly, we also discovered binding of *NESPR, PHOX2B, GATA3* and *ISL1* in the *NESPR* locus itself, with those binding sites overlapping enhancer elements (Fig. 6a). In 85 instances, *NESPR* binding sites were located in or in the vicinity of a differentially expressed gene after *NESPR* downregulation. 79 of those 85 genes were downregulated after loss of *NESPR* expression, implying a direct role in their transcriptional regulation. Taken together, our results suggest that *NESPR* is not only controlling expression of *PHOX2B* in cis, but could

also function as a trans-acting regulatory RNA, in close association with the core noradrenergic transcription factors GATA3, ILS1 and PHOX2B itself.

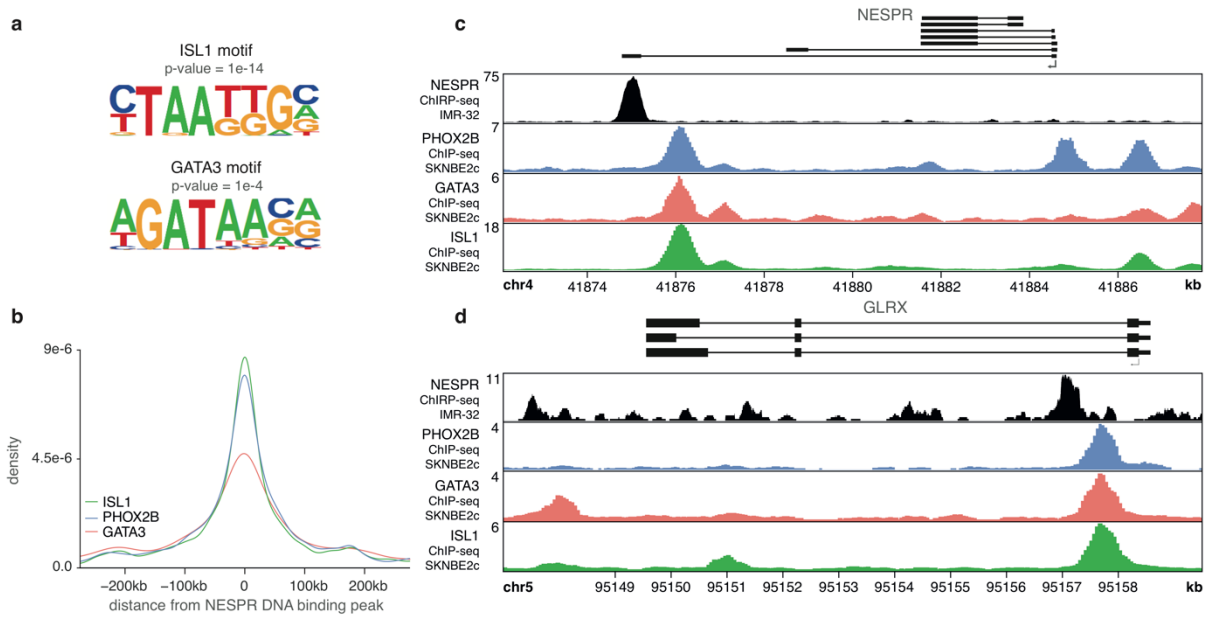


Figure 6: *NESPR* binding sites coincide with binding regions of members of the noradrenergic CRC
a) HOMER motif analysis reveals significant enrichment of ISL1 and GATA3 binding motifs (p = 1e-14 and p = 1e-4, respectively) b) The density plot shows the enrichment of binding sites of the three transcription factors at the *NESPR* binding regions. c) *NESPR* binds to its endogenous locus, together with PHOX2B, GATA3 and ISL1, three transcription factors of the noradrenergic CRC. d) In 84 loci PHOX2B, GATA3 and ISL1 bind the same region as *NESPR*, exemplified for GLRX.

NESPR downregulation results in growth arrest and apoptosis

Components of the noradrenergic CRC define cell identity and have previously been shown to be essential for neuroblastoma cell survival[10]. To evaluate the role of *NESPR* in neuroblastoma cell survival, we treated two neuroblastoma cell lines with 2 independent *NESPR* ASOs and monitored cell proliferation in real time. *NESPR* knockdown significantly reduced proliferation in both cell lines (Fig. 7a). Furthermore, we observed a significant reduction in colony formation capacity (Fig. 7c). These effects were accompanied by an induction of apoptosis, as measured by caspase 3/7 activity (Fig. 7b). In contrast, cells transfected with the *NESPR* siPOOL did not show any effects on apoptosis. Together, these results demonstrate that nuclear, but not cytoplasmic, *NESPR* transcripts are essential for neuroblastoma cell survival.

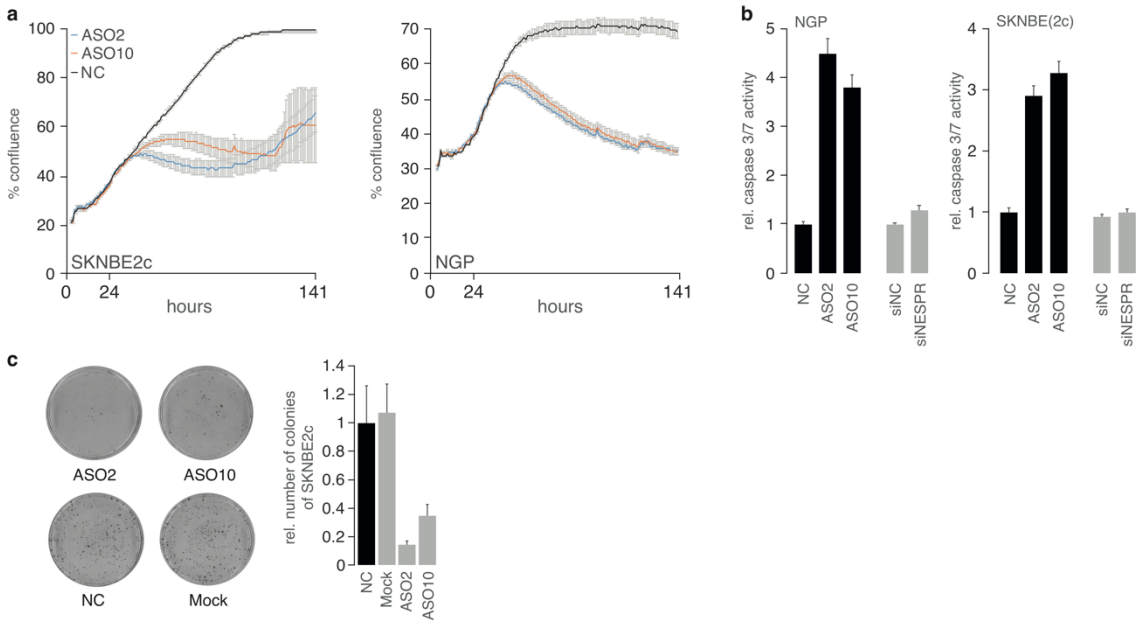


Figure 7: *NESPR* plays a role in neuroblastoma cell proliferation and survival
a) ASO-mediated knockdown of *NESPR* leads to a reduction of cell proliferation as evidenced by the confluency plots produced through live-cell imaging using the IncuCyte. b) An induction of caspase 3/7 activity is observed after downregulation of *NESPR* using two LNA-modified ASOs, whereas no difference in caspase 3/7 activity was detected after siPOOL transfection. c) A drop in colony formation capacity was noticed after ASO-mediated knockdown in SKNBE(2c).

Discussion

Large scale RNA-sequencing efforts have identified thousands of lncRNA transcripts whose functions are largely unknow. Nevertheless, evidence is accumulating that lncRNAs can have regulatory potential and play important roles during development and disease, including cancer. Here, we present a novel neuroblastoma-specific lncRNA, called *NESPR*, and report on its role as a co-regulator of neuroblastoma cell identity. Neuroblastoma cell identity is heterogenous and defined by two distinct core regulatory circuits, each consisting of a select set of transcription factors[10,11]. Thus far, lncRNAs have not been reported as components of these CRCs. Our results demonstrate that *NESPR* functions as a cis-regulator of *PHOX2B* expression, one of the transcription factors driving noradrenergic cell identity. We found *NESPR* and *PHOX2B* to be part of the same insulated neighborhood and demonstrated binding of *NESPR* to enhancer elements within that insulated neighborhood. Insulated neighborhoods are functional components in gene regulation and contain genes involved in determination of

cell identity during development. As we could only identify this insulated neighborhood in *NESPR* positive noradrenergic neuroblastoma cell lines, *NESPR* could be involved in the establishment or maintenance of the chromatin loop. Positionally conserved lncRNAs (pcRNAs) are often associated with transcription factors implicated in development. More than half of pcRNAs are involved in chromatin formation and are called topological anchor point RNAs (tapRNAs), which often show higher sequence conservation than other lncRNAs[36]. We found *NESPR* to be highly conserved across mammals. Although an absence of conservation does not imply a lack of function[37], stable and crucial functions across species can be ascribed to evolutionary conserved lncRNAs[38]. Several lncRNAs are already implicated in interacting with chromatin modifying complexes such as PRC2, including *HOTAIR*[39] and *XIST*[40] hereby directly contributing to chromatin modifications. *NESPR* could act in a similar manner, assisting in local genome structure modulation through direct tethering of chromatin remodelers to their specific chromatin site. This is further substantiated by binding of *NESPR* in the vicinity of the anchor points of the insulated neighborhood, as evidenced by the ChIRP-seq data. Another lncRNA implicated in insulated neighborhood formation is *Wrap53*, a lncRNA in control of *TP53* expression upon DNA damage[41]. *Wrap53* interacts with the RNA binding region of CTCF, bridging two monomers into multimers, influencing CTCF driven loop formation. CTCF binding in the promoter region of *NESPR*, similar to *Wrap53*, could indicate the functional importance of the lncRNA transcript in the formation of the insulated neighborhood, through interaction with and generation of CTCF multimers. Remodeling of the *PHOX2B-NESPR* locus might be initiated by *NESPR* itself, bringing enhancers located in that region in proximity to the promoter of *PHOX2B*, resulting in transcriptional activity. To validate these premises, 4C experiments after ASO-mediated KD of *NESPR* are planned to evaluate the effect of *NESPR* on the insulated neighborhood formation. Furthermore, RIP/CLIP-seq of CTCF could provide evidence of binding between CTCF and *NESPR*, whereas ChIRP-MS might provide a broader view of potential interaction partners active in chromatin remodeling.

In case of *NESPR* not being in control of the insulated neighborhood formation, the lncRNA could still play a role in the transcriptional regulation[42] of *PHOX2B* through recruitment of co-factors or transcription factors. Since the regional chromatin remodeling ensures the two genes to be in the vicinity of one another, as evidenced by the 4C seq data and CTCF, SMC3 and RAD21 ChIP seq data, *NESPR* could act as a guide for transcriptional regulators or

cofactors. Recent studies[10,11] unraveled two distinct cell identity types, governed by CRCs containing super-enhancer associated TFs. At least three TFs of the noradrenergic CRC including GATA3, ISL1 and PHOX2B, bind in the second exon of the longest and most abundant transcript of *NESPR*, alongside its own binding site, implying a functional role of the lncRNA in their recruitment. Establishing *NESPR* as a guide in transcriptional regulation warrants further research. Through ChIRP-MS, we will be able to discern all interacting proteins, revealing its mechanism of action.

Comparative analysis of *NESPR* and *PHOX2B* RNA-seq data brought about the notion that *NESPR* has other functionalities, independent of *PHOX2B*, as downregulation of *NESPR* resulted in differentially expressed genes which could not be confirmed upon *PHOX2B* knockdown. Moreover, ChIRP-seq data revealed multiple *NESPR* binding sites across the genome, often coinciding with GATA3 or ISL1 binding sites. Similar to its own locus, *NESPR* can be a guide for these noradrenergic transcription factors, aiding in consolidating the noradrenergic cell identity. Motif analysis also revealed enrichment of the PRDM1 binding sequence, a protein involved in fate specification of neural crest cells[43]. Interestingly, PRDM1 is strongly associated with the neural crest cell-like (or mesenchymal) lineage in neuroblastoma.

The impact of *PHOX2B* on neuroblastoma cell growth is well described[44]. Continuing on the relation we established between *NESPR* and *PHOX2B* and the association of *NESPR* levels with overall survival in patients, *in vitro NESPR* knockdown resulted in a decrease in cell growth and confluency. The increase in Caspase 3/7 led us to conclude that this decline was caused by apoptosis. Furthermore, a considerable reduction in colony formation capacity was observed after knockdown of *NESPR.*

In conclusion, we identified *NESPR* as a novel component in the noradrenergic CRC and a contributing factor to maintenance and survival of neuroblastoma cells. While further validation in the mechanism of action is required, we also demonstrate that a lincRNA can have both cis and trans regulatory activity. Finally, this study also presents *NESPR* as a potential therapeutic target with high specificity for neuroblastoma cells.

Material and Methods

Specificity score

The specificity scores, based on the Jensen-Shannon divergence, were calculated using TPM values of the Cancer Cell Line Encyclopedia, as described in the manuscript by Cabili et al. Counts were generated with Kallisto, with Ensembl 91 as a reference transcriptome. Genes with counts higher than 5 in more than half of the samples per cancer type were retained for further analysis.

Conservation

The conservational status of *NESPR* was assessed using the vertebrate 46-way PhyloP scores[45], which measure evolutionary conservation in 46 different species of vertebrates at the individual nucleotide level. Percentage of conservation of 46 different vertebrate species was calculated based on the mean PhyloP score over the entirety of the most abundant *NESPR* transcript. Conservation between the mouse and human *NESPR* sequence was determined using BLASTN.

Cell culture

The NB cell lines, SKNBE(2c) and NGP, used in this study were grown in RPMI1640 medium supplemented with 10% fetal bovine serum (FBS), 2 mM L-Glutamine and 100 IU/ml penicillin/streptavidin (complete medium) at 37 °C in a 5% CO_2 incubator. Cell culture reagents were obtained from Life Technologies. SKNBE(2c) and NGP cells (3 replicates) were transfected with Lipofectamine 2000 according to manufacturer instructions with 100 nM of LNA-modified ASOs (Exiqon/Qiagen) or siPOOLs (siTOOLs Biotech). As a control, cells were transfected with a negative control ASO (Negative control A, Exiqon/Qiagen) or negative control siPOOL, or treated with transfection reagent without addition of ASOs or siPOOLs.

Three samples of CLB-GA-shPHOX2B and CLB-GA-shControl cells were treated with doxycycline for 5 days. The experiment was repeated at a different time point to generate a total of 6 *PHOX2B* knockdown samples and 6 control samples.

Antisense oligonucleotides and siPOOLs

The ASOs were designed using the design pipeline of Exiqon. All ASOs are produced in a gapmer configuration with a phosphorothioate backbone with flanking LNA molecules. Standard desalted, lyophilized ASOs (5 nmol) were resuspended in 250 µL of RNAse-free water. siPOOLs are designed by siTOOLs Biotech. One batch of siPOOLs contains 30 siRNAs targeting the same RNA molecule. Lyophilized siPOOLs (5 nmol) were resuspended in 250 µL of RNAse-free water.

ASO2: GCTCGAATGATCTTTA; ASO10: AAAACTGATCGACGCA

Murine neural crest cells (NCC) and MYCN-driven tumors

Murine NCCs and 3 murine MYCN-driven tumors were isolated, RNA was extracted and was reverse transcribed into cDNA as described by Olsen et al.[46]. Concentration of the cDNA samples was approximately 10 ng/µL. RT-qPCR was used to measure *PHOX2B* and *NESPR* expression levels.

Real-time quantitative PCR

Total RNA was extracted post-transfection after 24 or 48 hours from fresh NGP and SKNBE(2c) cells using Qiazol Reagent and the miRNeasy kit (Qiagen) according to the manufacturer's instructions, including on-column DNase treatment. For the CLB-GA-shPHOX2B cells, total RNA was extracted from fresh cells using TRIzol® Reagent (Invitrogen) and the AllPrep DNA/RNA Mini Kit (Qiagen). All samples were subjected to quality control on a Bioanalyzer instrument and all RNA exhibited a RIN (RNA Integrity Number) > 8. Concentration was determined with the Nanodrop (Thermo Scientific). Reverse transcription was performed using the iScript Advanced cDNA synthesis kit (BioRad). Quantitative PCR was carried out using 5 ng of cDNA, 2.5 ul SsoAdvanced SYBR qPCR supermix (Bio-Rad). A volume of 0.25 µl forward and reverse primer (5 µM, IDT) was added. Expression levels were measured on a LightCycler 480 (Roche). Normalization was based on relative expression levels of 3 reference genes (SDHA, YWHAZ, TBP) and analyzed using qBase+ software (Biogazelle).

Primer sequences used are: *NESPR* (primer assay 1) forward, GTTGGAGTCTGCACAGTTGG; reverse, CTTTGCCCACTTTCTGACCC; *NESPR* (primer assay 2) forward, AATGCGCGCACCTTCAAC; reverse, CTTTGCCCACTTTCTGACCC; murine *NESPR* (primer assay 1)

forward, GTTGGAGTCTGCACAGTTGG; reverse, CTTTGCCCACTTTCTGACCC; murine *NESPR* (primer assay 2) forward, CCTTTCAGGCCAACAATGCT; reverse, ACTGCATTCACAACCCATGG; murine *NESPR* (primer assay 3) forward, GGGTCAGAAAGTGGGCAAAG; reverse, CTCACTGCATTCACAACCCA; *PHOX2B* forward, TAACTGGCTGCAGAGAAATC; reverse, GTTCGGATCATTCCAACAGA; murine *PHOX2B* forward, GCACTACCCTGACATCTACACC; reverse, CTGCTTGCGAAACTTAGC; *SDHA* forward, TGGGAACAAGAGGGCATCTG; reverse, CCACCACTGCATCAAATTCATG; *YWHAZ* forward, ACTTTTGGTACATTGTGGCTTCAA; reverse, CCGCCAGGACAAACCAGTAT; *TBP* forward, CACGAACCACGGCACTGATT; reverse, TTTTCTTGCTGCCAGTCTGGAC; murine TBP forward: CCCCACAACTCTTCCATTCT; reverse: GCAGGAGTGATAGGGGTCAT; murine PPIA forward: CAGACGCCACTGTCGCTTT; reverse: TGTCTTTGGAACTTTGTCTGCAA.

Analysis of RNA-sequencing data

All RNA sequencing libraries were prepared from 200 ng of total RNA using the Illumina TruSeq Stranded mRNA Library preparation kit. Kappa qPCR quantification was used to perform equimolar pooling. The concentration of the pooled library was measured with Qubit. Sequencing of 1.2 pM of pooled library was performed with the Illumina NextSeq 500 instrument using 2 x 75 cycles (paired-end) for all samples (high output sequencing kit). Transcripts were quantified by means of Kallisto using the human Ensembl 91 transcriptome as a reference.

Gene set enrichment analysis

Gene set enrichment analysis was performed using the Java GSEA application of the Molecular Signatures Database (*1000 permutations, classic analysis*). To identify enriched pathways based on differentially expressed genes between samples with or without *NESPR* knockdown, lists of mRNAs, pre-ranked on log-transformed fold change, were analyzed through GSEA, using all curated gene sets in the Molecular Signatures Database (C2 collection). Significant enrichment was defined at FDR<0.05. To validate overlap of differentially expressed genes found in the *PHOX2B* model system and the *NESPR* knockdown samples, we produced our own gene sets, based on the genes differentially expressed upon *PHOX2B* knockdown in the shPHOX2B samples. Genes with a logFC higher or lower than 1.5 and a significance cut-off

(q<0.05), were divided into an upregulated or downregulated shPHOX2B gene set, respectively.

Cytoplasmic/nuclear RNA fractionation assays

$1x10^6$ SKNBE(2c) cells were harvested and collected with 5 ml of DMEM in a tube, centrifuged at 230g for 5 minutes and subsequently washed twice with 1 ml of ice-cold phosphate-buffered saline (PBS 1X). Cells were later resuspended again with 1 ml of PBS 1X and dived in two aliquots of 500 µL each to perform Total RNA purification (using Tri-reagent protocol by Sigma Aldrich company) and Cytoplasmic/nuclear RNA fractionation. To perform the subcellular fractionation, cells were centrifuged and later gently resuspended with 300 µL of Lysis Buffer[47] for 20 times and sedimented for 4 minutes at 800g (4°C). The supernatant (Cytoplasmic fraction) was collected and next resuspended with 1 ml of Tri-reagent to perform Cytoplasmic RNA purification, while pellet (nuclei) were resuspended in 500 µL of Lysis buffer and left on ice for 30 minutes. Nuclei were then washed by 2 consecutive centrifugations through 350 µl sucrose cushions in lysis buffer 20% and 30% (w/w) and span at 900g (4°C) for 10 minutes.

The purified nuclei were resuspended with 500 µl of Tri-reagent solution to perform nuclear RNA purification. Total-cytoplasmic-nuclear RNA integrity was then verified by Agarose gel electrophoresis (gel concentration: 1,5%) using Ethidium bromide as intercalating agent.

Generate stable CRISPRi cell line

Cloning

To create pLV-TRE-dCas9-KRAB>PGK-Neo (MP-I-1081), we performed a restriction digest with XhoI and XbaI on pLV-TRE-dCas9-BirA*>PGK-Neo (MP-I-1014, VectorBuilder) to remove the BirA* module. KRAB was PCR amplified with overlapping primer extensions using Pfu polymerase (Thermo) according to manufacturer's instructions. KRAB was inserted in the digested backbone by InFusion (Takara Bio) cloning. Chemical transformation was performed in in-house made Stbl3 chemically competent cells to avoid recombination events frequently seen with lentiviral LTRs. Plasmid sequence was verified by control restriction digest and sequencing.

Lentivirus production

6.5 x 10^6 HEK293T cells were seeded in a T75 per lentiviral construct, and transfected with 24 µg lentiviral vector, 18 µg pCMVR8.74 (Addgene #22036), and 7.2 µg pMD2.G (Addgene #12259) one day after seeding with calcium phosphate. Medium was refreshed one day after transfection, and virus was harvested on two and three days after transfection. Virus was concentrated by ultracentrifugation at 22.000 rpm for 2 h at 4°C. A colony formation assay was performed to determine viral titer, in brief: HEK293T cells were infected with a serial dilution of 2 µL concentrated virus. Selection was performed until the uninfected control was completely dead. Surviving colonies were stained using crystal violet, and counted to calculate viral titer.

Generate stable cell line

400.000 SK-N-BE(2c) or SH-SY5Y cells were seeded in a T25 in complete RPMI1640, and infected in two consecutive rounds with concentrated lentivirus at a MOI of 0.3. First, the Tet-On regulator pLV-EF1a-rtTA-T2A-tTS>PGK-Hygro was infected and selected for two weeks using 200 µg/mL hygromycin. In a second round, cells were infected with pLV-TRE-dCas9-KRAB>PGK-Neo and selected on 1 mg/mL G418 for two weeks. 8 µg/mL polybrene was added to the infection medium to enhance lentiviral transduction. Uninfected wild-type control cells were selected in parallel to ensure selection was sufficient. Cells were cultured to not exceed a confluency of 60-65%.

Transient CRISPRi

sgRNAs were designed with an online tool (CRISPR.MIT.EDU) and synthesized with the Guide-it sgRNA In Vitro Transcription kit (Takara) according to the manufacturer's protocol and purified with the Guide-it IVT RNA Clean-Up kit (Takara) according to the manufacturer's protocol. sgRNA yield was quantified with the Nanodrop 2000 (ThermoFischer). SK-N-BE(2c)-dCas9-KRAB and SH-SY5Y-dCas9-KRAB cells were cultured in RPMI + 10% FBS + antibiotics. For transfections, cells were seeded in 12-well plates at 2x10^5 cells/well the previous day. SK-N-BE(2c)-dCas9-KRAB cells were transfected with FuGene HD Transfection Reagent (Promega) according to the manufacturer's protocol. Briefly, 1µg sgRNA was diluted in 80µl OPTI-MEM I; 5µl FuGene HD Transfection Reagent was added to the mix; the mix was vortexed for 10 seconds and was left at room temperature for 10 minutes before adding to the cells. SH-SY5Y-

dCas9-KRAB cells were transfected with the NEON nucleofector (ThermoFischer) 10µl kit according to the manufacturer's protocol. Briefly, per well, cells were trypsinized and washed with PBS; $2x10^5$ cells were resuspended in 10µl buffer R and 1µg sgRNA was added to the mix; the mix was electroporated at 1100v, 40ms, 1 pulse. After electroporation cells were immediately added to 12-well plates containing 500µl pre-warmed medium per well.

4C-sequencing

4C templates were prepared according to the protocol by Van de Werken et al.[48]. In brief, for each template 1x107 cells were detached, counted, resuspended and crosslinked by incubating them with 2% formaldehyde for 10 min at room temperature. Following cell lysis, crosslinked DNA was digested with 400U of DpnII restriction enzyme (NEB # R0543L) and nearby DNA fragments were ligated using 50U of T4 DNA ligase (Roche # 10799009001). Ligated DNA circles were de-crosslinked overnight using proteinase K and purified with NucleoMag P-Beads (Macherey-Nagel) to obtain an intermediate 3C template. A second round of digestion and ligation, using 50U of Csp6I restriction enzyme (Thermo Scientific # ER0211), resulted in 4C templates.

Adaptor-containing reading and non-reading primers, specific to the viewpoints of interest, were designed to amplify all captured, interacting DNA fragments[48] (Table S1 & S2). For each viewpoint, 16 PCR reactions, each using 200 ng of input 4C template, were pooled. Resulting 4C sequencing libraries were purified using High Pure PCR Product Purification kit (Roche # 11732676001) and QIAquick PCR Purification kit (Qiagen # 28106). Approximately 15-20 different 4C sequencing libraries were pooled and sequenced simultaneously on an Illumina NextSeq 500 (single-end, 75 nt, loading concentration 1.6 pM).

Chromatine immunoprecipitation by RNA Purification (ChIRP) assay followed by mass spectrometry with label-free quantification

75-100 X 10^7 IMR32 cells were cultured in 145 cm^2 dishes, washed once with PBS, and UV cross-linked in PBS at 254 nm with an increasing intensity up to 400 mJ/cm^2. Cells were scraped in PBS, and split equally among eight microcentrifuge tubes. ChIRP lysis buffer[49] (20 mM Tris-HCl pH7.5, 200 mM NaCl, 2.5 mM $MgCl_2$, 0.05% NP-40, 0.1% SDS) was supplemented with fresh 0.1% sodiumdeoxycholate, 60 U/mL Superase-In RNase inhibitor (Invitrogen), 1 mM DTT, 0.5 mM PMSF, and protease inhibitor cocktail (Roche). Each cell pellet was suspended in supplemented ChIRP lysis buffer, and sonicated with a Bioruptor (Diagenode) until lysates appeared clear. 5 µL lysate was treated with proteinase K (Sigma), and gDNA was purified using a PCR clean-up kit (Machery-Nagel). Purified gDNA was ran on a 1% agarose gel to ensure proper fragment size (<500 bp). 10% of the ChIRP sample was used for RNA extraction of input material. Thereafter, 6.23 µL of 50 µM *NESPR* and LacZ raPOOLs (siTOOL Biotechnologies) were bound to 100 µL of equilibrated RNase-free Dyna-One C1 magnetic beads (Thermo) per sample and were incubated overnight at 4°C. Next day, ChIRP lysates were pre-cleared with 30 µL RNase-free by end-to-end rotation for 30 min at 4°C. After pre-clearing, raPOOL-bound probes were added and lysates were rotated for 3 h at 4°C. Bead-bound fractions were washed three times with supplemented ChIRP lysis buffer. 10% of the sample was used for RNA extraction to validate RNA pulldown on RT-qPCR. Next, beads were washed three times with RNase-free trypsin digestion buffer (20 mM Tris-HCl pH7.5, 2 mM $CaCl_2$), and were ultimately resuspended in 20 µL 20 mM Tris-HCl pH7.5. 750 ng trypsin was added directly on the beads, and digestion was left overnight at 37°C. Next day, beads were magnetized and the supernatant was transferred to a new microcentrifuge tube, an additional 250 ng trypsin was added and incubated for 3 h at 37°C. Peptides were acidified with 20% formic acid and transferred to a MS vial. The LC-MS/MS was operated by the VIB Proteomics Core facility. Peptides were separated over a 1.5 h gradient and ran on a Q-Executive HF (Thermo) Orbitrap MS system. Proteins were searched by MaxQuant using the UniProt Human Proteome as a reference and differential proteomic analysis was performed using Perseus software.

Chromatine immunoprecipitation by RNA Purification (ChIRP) assay followed by high-throughput sequencing and motif analysis

20-30 X 10^6 IMR32 cells per sample were cultured in T175 flasks and harvested by scraping in PBS. PBS was aspirated and *in vivo* cross-linking was done by resuspending the cell pellets in 1% glutaraldehyde (Sigma) in PBS and letting it rotate for 10 min at room temperature. The reaction was quenched by adding 1.25 M glycine. Pellets were lysed in SDS lysis buffer[50] (50 mM Tris-HCl pH7.5, 10 mM EDTA, and 1% SDS) supplemented fresh with 60 U/mL Superase-In RNase inhibitor, 1 mM DTT, 0.5 mM PMSF, and protease inhibitor cocktail. All following steps were performed as described before. DNA was eluted in two consecutive rounds by treating with a RNase cocktail (10 mg/mL RNase A (Sigma) and 10 U/µL RNase H (Roche)) in DNA elution buffer (50 mM $NaHCO_3$ and 1% SDS). Next, supernatant was treated with proteinase K at 50°C for 45 min, and DNA was purified using a 5PRIME heavy phase lock gel tubes and PhOH:Chloroform:Isoamyl. The aqueous phase was transferred, and 3 µg GlycoBlue, 30 µL 3 M NaOAc, and 900 µL 100% EtOH was added and stored overnight at -20°C. Next day samples were spinned at 16.100 x g for 30 min at 4°C. Supernatant was removed and the pellet was washed in 1 mL 70% EtOH. Subsequently, the pellet was air dryed, and resuspended in sterile in 30 µL nuclease-free water.

Raw sequencing data of ChIRP-seq were evaluated using FastQC and subsequently mapped to the human reference genome (hg19) with bowtie2. Peakcalling was performed using Macs2 only retaining peaks with $q<0.05$ and filtered using the encode blacklist code. Amplified regions on chr2p were manually removed. Homer was used for enrichment of known motifs and discovery of *de novo* motifs.

Phenotypic experiments

Confluency of SKNBE(2c) and NGP after *NESPR* knockdown was measured in real-time (2 hours interval, magnification ×4, whole well picture) using the IncuCyte® Live Cell imaging system (Essen BioScience). A density of $2.5x10^3$ of SKNBE(2c) and $5x10^3$ of NGP cells was used to seed in a 96 well plate (Corning costar 3596) containing complete medium. Plates were placed in the IncuCyte for 24 hours to ensure attachment. Transfection with 2 ASOs and a negative control ASO was performed after 24 hours, after which the plates were placed back (3

replicates per condition). Masking for cell counting was done using the IncuCyte® ZOOM Software. Caspase 3 and 7 activity was measured using Caspase-Glo 3/7 luciferase assay (Promega) and VICTOR X4 Reader (PerkinElmer) 48 and 72 h after transfection of ASOs and siPOOLs. For the colony formation assays, 2000 SKNBE(2c) cells per condition (2 *NESPR* ASOs, NTC, mock and untreated cells) were seeded in a 6-cm dish in a total volume of 5 ml complete medium. The dishes were placed in a humid incubator at 37 °C and not disturbed for 14 days. After visual inspection, the colonies were washed with PBS, fixed and stained with 0.1% crystal violet in a 10% ethanol solution and counted using ImageJ.

Statistical analysis

Normalization and differential expression analysis of the *NESPR* ASO RNA-seq data were performed using limma voom (threshold of $q<0.05$). P-values were adjusted using the Benjamini-Hochberg method.

Overall survival analysis was performed on the SEQC data set using a Kaplan-Meier analysis. Patients were grouped based on *NESPR* expression, using the median expression value as a cutoff. Significance was assessed based on the log-rank test.

Mean-centered data standardization was used for allow comparison between separate colony formation assays. Assessment of statistical significance concerning different transfection conditions for the colony formation assay was executed by a non-parametric Mann–Whitney test using R (version 3.5.1).

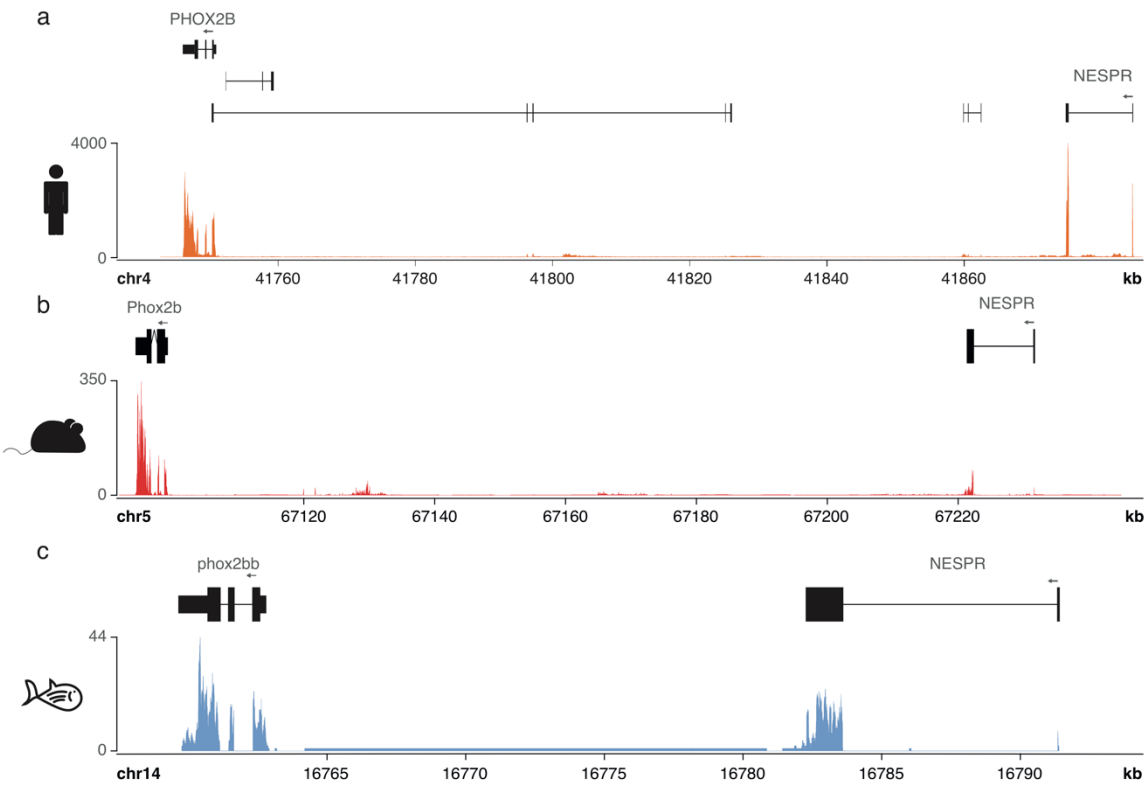
References

1. Maris, J. M. Recent Advances in Neuroblastoma. *N. Engl. J. Med.* **362**, 2202–2211 (2010).
2. Mueller, S. & Matthay, K. K. Neuroblastoma: Biology and staging. *Curr. Oncol. Rep.* **11**, 431 (2009).
3. Maris, J. M., Hogarty, M. D., Bagatell, R. & Cohn, S. L. Neuroblastoma. *Lancet* **369**, 2106–2120 (2007).
4. Schwab, M. *et al.* Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* **305**, 245–248 (1983).
5. Peifer, M. *et al.* Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **526**, 700 (2015).
6. Chen, Y. *et al.* Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* **455**, 971 (2008).
7. Trochet, D. *et al.* Germline Mutations of the Paired–Like Homeobox 2B (PHOX2B) Gene in Neuroblastoma. *Am. J. Hum. Genet.* **74**, 761–764 (2004).
8. Mosse, Y. P. *et al.* Germline PHOX2B Mutation in Hereditary Neuroblastoma. *Am. J. Hum. Genet.* **75**, 727–730 (2004).
9. Raabe, E. H. *et al.* Prevalence and functional consequence of PHOX2B mutations in neuroblastoma. *Oncogene* **27**, 469 (2007).
10. Boeva, V. *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* **49**, 1408 (2017).
11. van Groningen, T. *et al.* Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat. Genet.* **49**, 1261 (2017).
12. Consortium, I. H. G. S. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931 (2004).
13. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101 (2012).
14. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
15. Schmitt, A. M. & Chang, H. Y. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* **29**, 452–463 (2016).
16. Gutschner, T. & Diederichs, S. The hallmarks of cancer. *RNA Biol.* **9**, 703–719 (2012).
17. Fang, Y. & Fullwood, M. J. Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics. Proteomics Bioinformatics* **14**, 42–54 (2016).
18. Raveh, E., Matouk, I. J., Gilon, M. & Hochberg, A. The H19 Long non-coding RNA in cancer initiation, progression and metastasis - a proposed unifying theory. *Mol. Cancer* **14**, 184 (2015).
19. Pandey, G. K. & Kanduri, C. Long noncoding RNAs and neuroblastoma. *Oncotarget* **6**, 18265–18275 (2015).
20. Atmadibrata, B. *et al.* The Novel Long Noncoding RNA linc00467 Promotes Cell Survival but Is Down-Regulated by N-Myc. *PLoS One* **9**, e88112 (2014).
21. Liu, P. Y. *et al.* Effects of a Novel Long Noncoding RNA, lncUSMycN, on N-Myc Expression and Neuroblastoma Progression. *JNCI J. Natl. Cancer Inst.* **106**, dju113–dju113 (2014).
22. Pandey, G. K. *et al.* The Risk-Associated Long Noncoding RNA NBAT-1 Controls

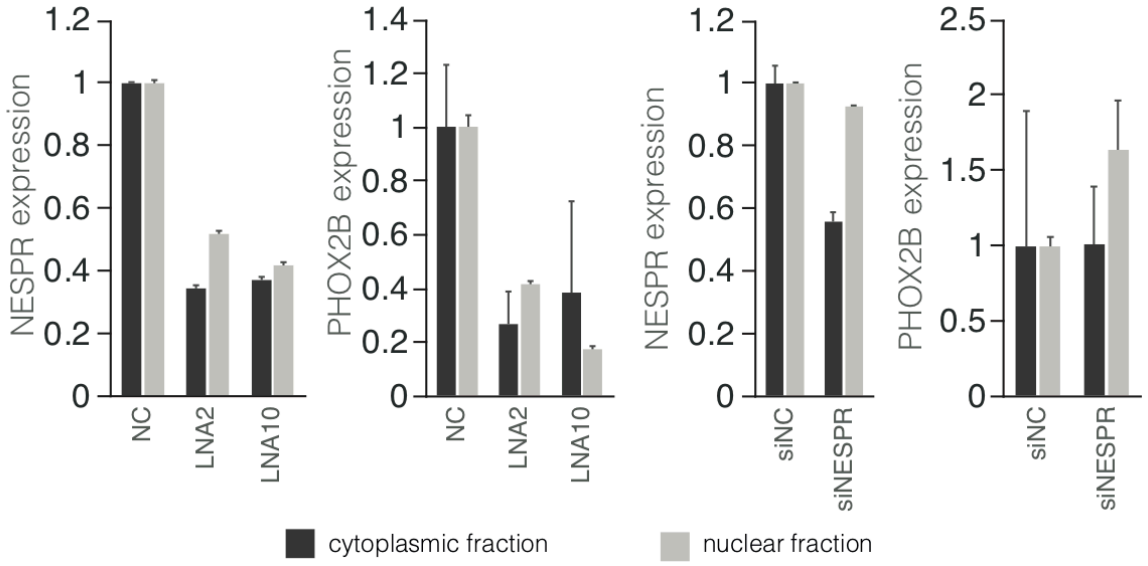
Neuroblastoma Progression by Regulating Cell Proliferation and Neuronal Differentiation. *Cancer Cell* **26**, 722–737 (2014).
23. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603 (2012).
24. Bielle, F. *et al.* PHOX2B Immunolabeling: A Novel Tool for the Diagnosis of Undifferentiated Neuroblastomas Among Childhood Small Round Blue-cell Tumors. *Am. J. Surg. Pathol.* **36**, (2012).
25. Longo, L. *et al.* PHOX2A and PHOX2B genes are highly co-expressed in human neuroblastoma. *Int. J. Oncol.* **33**, 985–991 (2008).
26. Consortium, S.-I. *et al.* A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* **32**, 903 (2014).
27. Zhang, W. *et al.* Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol.* **16**, 133 (2015).
28. Maguire, L. H., Thomas, A. R. & Goldstein, A. M. Tumors of the neural crest: Common themes in development and cancer. *Dev. Dyn.* **244**, 311–322 (2015).
29. (DGT), T. F. C. and the R. P. M. I. and C. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462 (2014).
30. Consortium, T. E. P. An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **489**, 57–74 (2012).
31. Rombaut, D. *et al.* Integrative analysis identifies lincRNAs up- and downstream of neuroblastoma driver genes. *Sci. Rep.* **9**, 5685 (2019).
32. Thompson, P. M., Gotoh, T., Kok, M., White, P. S. & Brodeur, G. M. CHD5, a new member of the chromodomain gene family, is preferentially expressed in the nervous system. *Oncogene* **22**, 1002 (2003).
33. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nat. Genet.* **38**, 1348 (2006).
34. 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. *Mol. Cell* **44**, 667–678 (2011).
35. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
36. Amaral, P. P. *et al.* Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. *Genome Biol.* **19**, 32 (2018).
37. Johnsson, P., Lipovich, L., Grandér, D. & Morris, K. V. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim. Biophys. Acta* **1840**, 1063–1071 (2014).
38. Li, D. & Yang, M. Q. Identification and characterization of conserved lncRNAs in human and rat brain. *BMC Bioinformatics* **18**, 489 (2017).
39. 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).
40. Plath, K. *et al.* Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science (80-.).* **300**, 131 LP – 135 (2003).
41. Saldaña-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* **28**, 723–734

(2014).
42. Bhat, S. A. *et al.* Long non-coding RNAs: Mechanism of action and functional utility. *Non-coding RNA Res.* **1**, 43–50 (2016).
43. Hernandez-Lagunas, L. *et al.* Zebrafish narrowminded disrupts the transcription factor prdm1 and is required for neural crest and sensory neuron specification. *Dev. Biol.* **278**, 347–357 (2005).
44. Di Zanni, E. *et al.* Targeting of PHOX2B expression allows the identification of drugs effective in counteracting neuroblastoma cell growth. *Oncotarget* **8**, 72133–72146 (2017).
45. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
46. Olsen, R. R. *et al.* MYCN induces neuroblastoma in primary neural crest cells. *Oncogene* **36**, 5075 (2017).
47. Carneiro, M. & Schibler, U. Accumulation of rare and moderately abundant mRNAs in mouse L-cells is mainly post-transcriptionally regulated. *J. Mol. Biol.* **178**, 869–880 (1984).
48. van de Werken, H. J. G. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9**, 969 (2012).
49. Leucci, E. *et al.* Melanoma addiction to the long non-coding RNA SAMMSON. *Nature* **531**, 518 (2016).
50. Chu, C., Quinn, J. & Chang, H. Y. Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.* 3912 (2012). doi:10.3791/3912

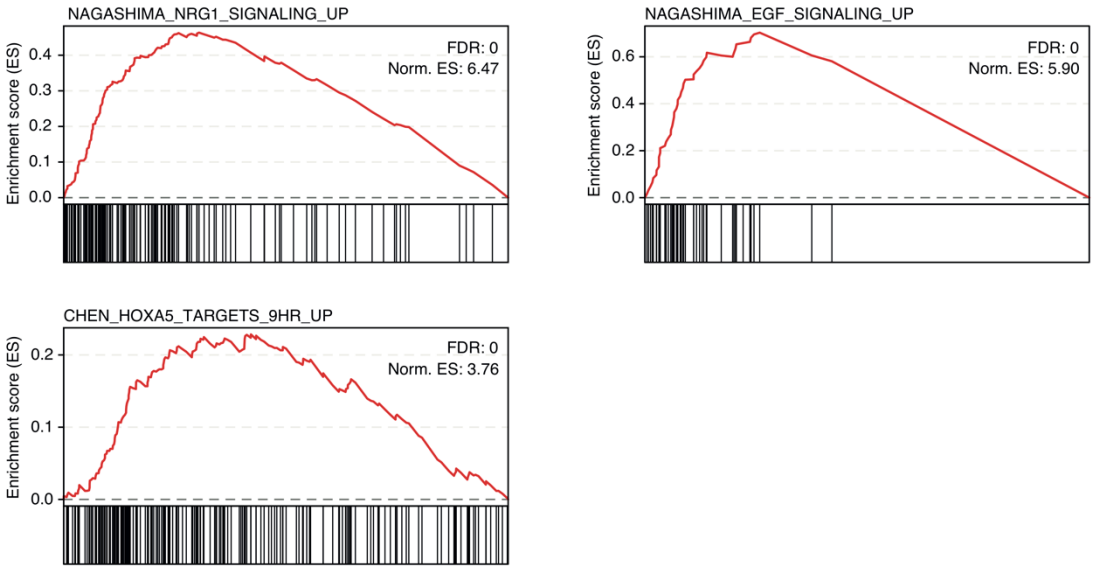
Supplemental information



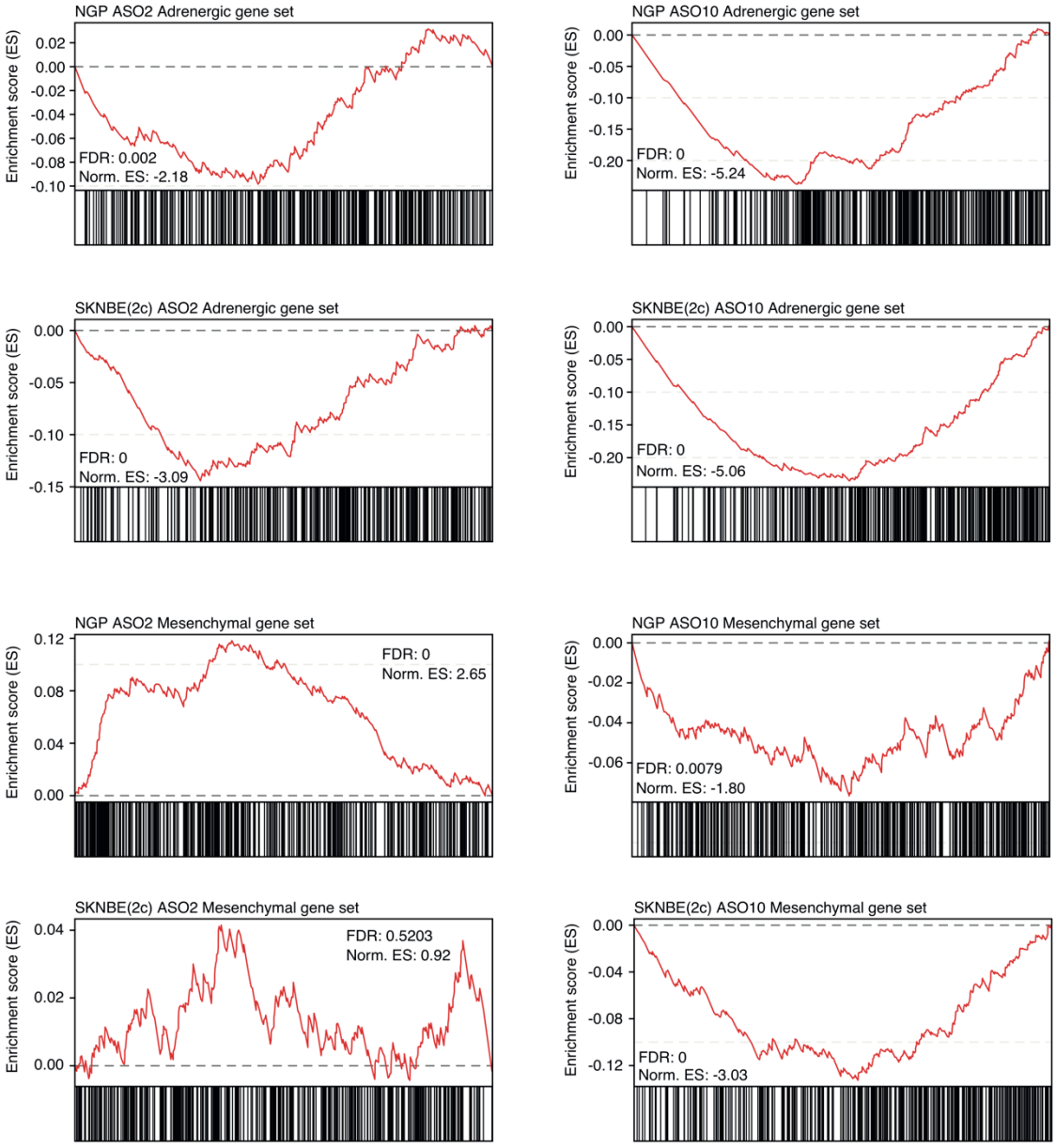
Supplemental Figure 1: *NESPR* expression is conserved in mouse and zebrafish
NESPR expression can be observed in human, mouse and zebrafish. The transcript is also positionally conserved, as it is located next to *PHOX2B* in all three species.



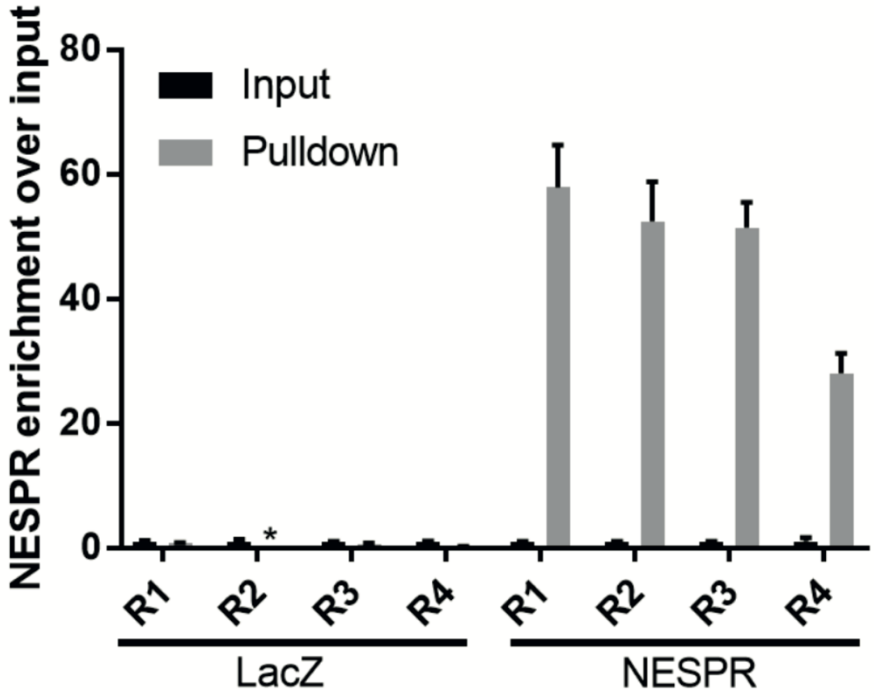
Supplemental Figure 2: The nuclear fraction of NESPR regulates PHOX2B expression.
Relative expression values of NESPR and PHOX2B after cellular RNA fractionation on ASO- and siPOOL-transfected cells. ASO-mediated KD of NESPR results in loss of PHOX2B expression. KD through siPOOLs only revealed a reduction in the cytoplasmic fraction of NESPR, without similar loss of PHOX2B.



Supplemental Figure 3: GSEA on all MSigDB gene sets reveals positive enrichment of gene sets implicated in neuronal development and EMT.



Supplemental Figure 4: Negative enrichment of adrenergic gene sets is calculated with GSEA using ordered mRNA lists after *NESPR* downregulation.
Significant negative enrichment of the adrenergic gene set is detected using GSEA on ordered mRNA lists from two cell lines transfected with two ASOs. Consistent, positive enrichment of the mesenchymal gene set is not clear, with only the ASO2 treated NGP cell line showing significant positive enrichment.



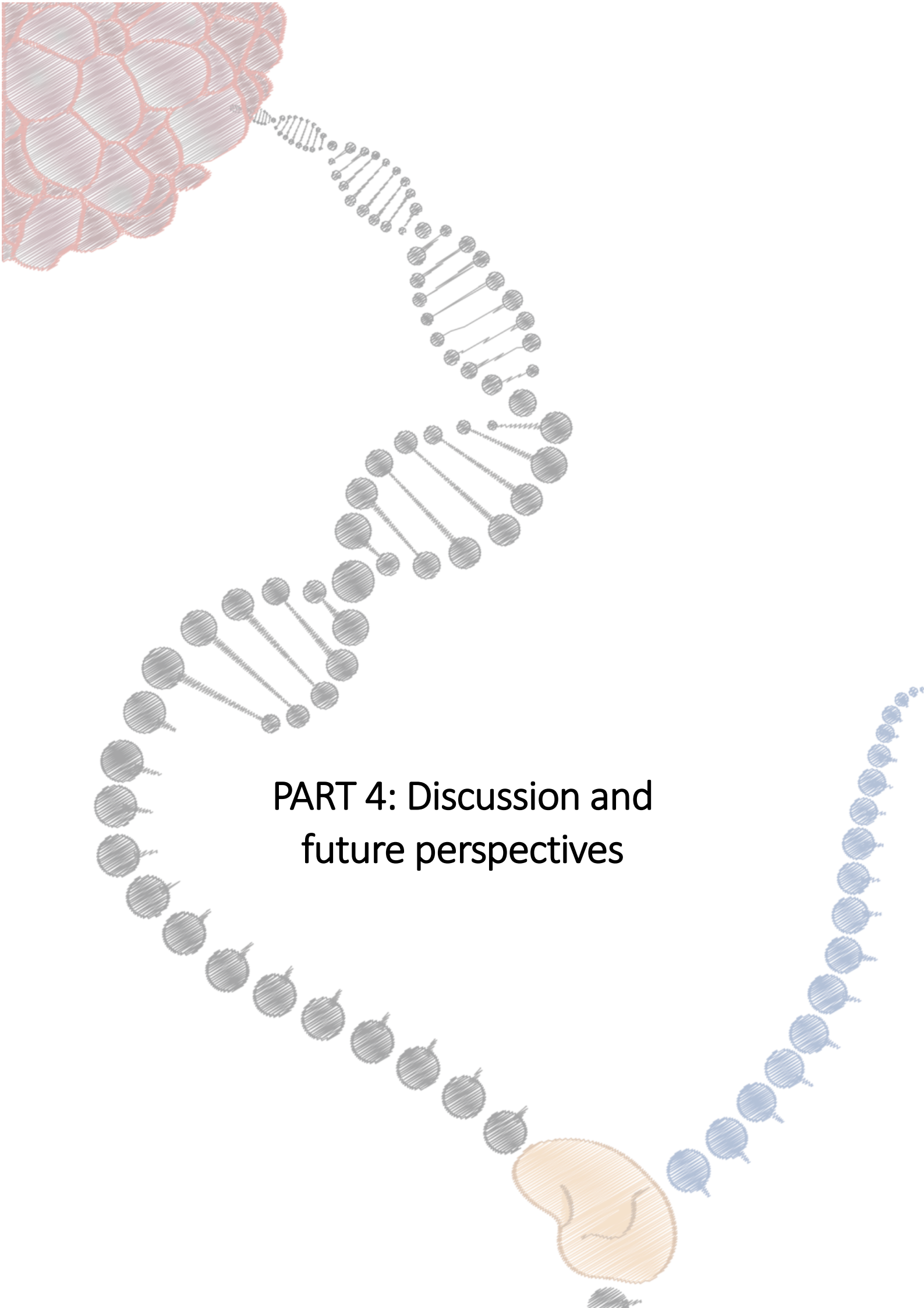
Supplemental Figure 5: Pulldown of NESPR in a ChIRP-experiment results in an enrichment of NESPR in comparison to the LacZ control.

Rank	Motif	Name	P-value	log P-pvalue	q-value (Benjamini)	# Target Sequences with Motif	% of Targets Sequences with Motif	# Background Sequences with Motif	% of Background Sequences with Motif
1		Isl1(Homeobox)/Neuron-Isl1-ChIP-Seq(GSE31456)/Homer	1e-14	-3.340e+01	0.0000	512.0	27.72%	9724.5	20.12%
2		Rfx5(HTH)/GM12878-Rfx5-ChIP-Seq(GSE31477)/Homer	1e-12	-2.866e+01	0.0000	91.0	4.93%	1017.7	2.11%
3		Bapx1(Homeobox)/VertebralCol-Bapx1-ChIP-Seq(GSE36672)/Homer	1e-11	-2.739e+01	0.0000	486.0	26.31%	9461.9	19.58%
4		PRDM1(Zf)/Hela-PRDM1-ChIP-Seq(GSE31477)/Homer	1e-8	-1.880e+01	0.0000	170.0	9.20%	2824.6	5.84%
5		Nkx3.1(Homeobox)/LNCaP-Nkx3.1-ChIP-Seq(GSE28264)/Homer	1e-7	-1.735e+01	0.0000	506.0	27.40%	10634.4	22.01%
6		Gata6(Zf)/HUG1N-GATA6-ChIP-Seq(GSE51936)/Homer	1e-5	-1.216e+01	0.0003	227.0	12.29%	4433.6	9.17%
7		GATA3(Zf)/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-4	-1.147e+01	0.0005	358.0	19.38%	7566.1	15.66%
8		Foxh1(Forkhead)/hESC-FOXH1-ChIP-Seq(GSE29422)/Homer	1e-4	-1.044e+01	0.0013	154.0	8.34%	2889.1	5.98%
9		COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer	1e-4	-1.026e+01	0.0014	296.0	16.03%	6191.4	12.81%
10		IRF8(IRF)/BMDM-IRF8-ChIP-Seq(GSE77884)/Homer	1e-4	-1.008e+01	0.0015	93.0	5.04%	1578.3	3.27%
11		Meis1(Homeobox)/MastCells-Meis1-ChIP-Seq(GSE48085)/Homer	1e-4	-9.650e+00	0.0021	304.0	16.46%	6436.4	13.32%
12		RORgt(NR)/EL4-RORgt.Flag-ChIP-Seq(GSE56019)/Homer	1e-3	-9.002e+00	0.0037	37.0	2.00%	492.1	1.02%
13		ETS:RUNX(ETS,Runt)/Jurkat-RUNX1-ChIP-Seq(GSE17954)/Homer	1e-3	-8.425e+00	0.0061	26.0	1.41%	308.9	0.64%
14		PU.1:IRF8(ETS:IRF)/pDC-Irf8-ChIP-Seq(GSE66899)/Homer	1e-3	-7.442e+00	0.0152	58.0	3.14%	958.9	1.98%
15		Gata4(Zf)/Heart-Gata4-ChIP-Seq(GSE35151)/Homer	1e-3	-7.113e+00	0.0198	228.0	12.34%	4857.6	10.05%
16		RAR:RXR(NR),DR5/ES-RAR-ChIP-Seq(GSE56893)/Homer	1e-2	-6.441e+00	0.0363	26.0	1.41%	354.1	0.73%
17		Barx1(Homeobox)/Stomach-Barx1.3xFlag-ChIP-Seq(GSE69483)/Homer	1e-2	-6.097e+00	0.0482	128.0	6.93%	2590.5	5.36%
18		IRF1(IRF)/PBMC-IRF1-ChIP-Seq(GSE43036)/Homer	1e-2	-6.002e+00	0.0500	45.0	2.44%	747.6	1.55%
19		Nkx2.5(Homeobox)/HL1-Nkx2.5.biotin-ChIP-Seq(GSE21529)/Homer	1e-2	-5.988e+00	0.0500	437.0	23.66%	10120.0	20.94%
20		Arnt:Ahr(bHLH)/MCF7-Arnt-ChIP-Seq(Lo_et_al.)/Homer	1e-2	-5.769e+00	0.0569	79.0	4.28%	1496.3	3.10%
21		Gfi1b(Zf)/HPC7-Gfi1b-ChIP-Seq(GSE22178)/Homer	1e-2	-5.755e+00	0.0569	128.0	6.93%	2617.6	5.42%
22		IRF3(IRF)/BMDM-Irf3-ChIP-Seq(GSE67343)/Homer	1e-2	-5.646e+00	0.0584	91.0	4.93%	1774.2	3.67%
23		NF1:FOXA1(CTF,Forkhead)/LNCAP-FOXA1-ChIP-Seq(GSE27824)/Homer	1e-2	-5.448e+00	0.0681	18.0	0.97%	231.1	0.48%
24		Tgif1(Homeobox)/mES-Tgif1-ChIP-Seq(GSE55404)/Homer	1e-2	-5.052e+00	0.0970	578.0	31.29%	13835.4	28.63%
25		PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-2	-4.688e+00	0.1340	105.0	5.68%	2168.4	4.49%

Table 1: NESPR binding motifs found by HOMER.

-LOG(P-value)	Difference	Gene names	Peptides	Unique peptides	Sequence coverage [%]	Unique sequence coverage [%]	Mol. weight [kDa]	Q-value	Score	Intensity	MS/MS count
4.95306693	-4.298970222	DYNLT1	2	2	33.6	33.6	12.452	0	29.513	124360000	5
6.837375906	-2.549619198	HNRNPC	19	19	55.6	55.6	33.67	0	323.31	4023100000	176
6.033170893	-1.892391205	HNRNPA2B1	14	13	52.7	50.1	37.429	0	323.31	2313600000	236
4.423670926	-1.882473946	RALY	8	8	26.5	26.5	32.463	0	52.096	111120000	32
5.513619675	-1.706149578	RPA1	14	14	27.8	27.8	68.137	0	123.52	226270000	45
2.849364644	-1.573633671	PRDX1	3	2	13.6	10.1	22.11	0	20.782	20206000	4
6.273055324	-1.355454922	U2AF1	4	4	21.2	21.2	27.872	0	47.878	132130000	25
2.970004452	-1.338625908	SSBP1	3	3	27.7	27.7	17.259	0	21.595	41663000	10
4.157471653	-1.103886127	HNRNPD	8	8	25.4	25.4	38.434	0	151.3	564650000	61
4.810413646	-1.037278652	HNRNPL	15	15	45.7	45.7	64.132	0	198.02	906140000	113
4.590511473	-0.977371216	TOP1	12	12	16.1	16.1	90.725	0	93.475	217250000	44
3.965004136	-0.874911308	HNRNPA1;HNRNPA1L2	10	9	31.5	29	38.746	0	323.31	972350000	100
4.857059421	-0.692703247	KHSRP	8	6	11.7	10.4	73.114	0	76.01	126120000	41
3.123992951	-0.689872742	ELAVL4;ELAVL2	7	2	18.2	6.8	41.769	0	43.776	78490000	19
5.11045099	-0.673391342	RBMX;RBMXL1	15	15	35.3	35.3	42.331	0	128.84	314830000	75
2.919569733	-0.586668491	HNRNPA3	7	7	24.3	24.3	39.594	0	163.2	402750000	53
2.892759648	-0.585682392	LMNB2	11	9	18.9	16.3	69.948	0	61.693	74184000	30
3.328331311	-0.551114559	NCL	14	14	20.4	20.4	76.613	0	127.16	173420000	48
3.831097894	-0.531153202	HNRNPK	13	13	37.1	37.1	50.976	0	238.1	1299300000	130

Table 2: Proteins bound to NESPR, identified using ChIRP-MS.



PART 4: Discussion and future perspectives

4.1 Discussion and future perspectives

Neuroblastoma is a rare childhood malignancy, arising from neural crest precursor cells. As an embryonal tumor, neuroblastoma is categorized as a DNA copy number disease, with a low mutational rate. One of the few genes to be mutated and implicated in neuroblastoma pathogenesis is ALK[1,2], with mutations occurring in 8 to 10% of patients. However, recurrent copy number variations (CNVs) have been described in a greater fraction of patients. MYCN amplification[3], TERT rearrangements[4] and segmental copy number alterations[5,6] are associated with poor prognosis and outcome. Although several driver oncogenes have already been identified, a lot remains to be discovered about their modes of action and their target genes. With this thesis, my aim was to contribute to our understanding of the long non-coding RNA components of these key signaling networks (figure 1).

4.1.1 LncRNAs as targets, modulators or regulators of key neuroblastoma genes

Advances in the genomic characterization of neuroblastoma have enabled the identification of multiple neuroblastoma oncogenes, including *MYCN*[3] and *ALK*[7] and neuroblastoma cell identity genes such as *HAND2* and *PHOX2B*[8]. These genes regulate a multitude of protein coding genes, however, insights in the role of lncRNAs up- and downstream of these genes is lacking.

We aimed to provide a core set of lincRNAs that regulate, are regulated by or have a modulating effect on proteins with important roles in neuroblastoma tumorigenesis. While our findings are mostly based on correlative analysis of gene expression data in neuroblastoma tumor samples, we included various model systems to further fine tune the selection of lincRNAs downstream of these key neuroblastoma genes. Nevertheless, further studies are required to pinpoint which of these candidate lincRNAs are actually functional and drive part of the phenotypes know to be associated to these genes.

One of the prioritized lincRNAs is *MEG3*, a lincRNA defined as a tumor suppressor in breast cancer[9], chronic myeloid leukemia[10], liver cancer[11], and others cancer types[12,13]. A well-

described function of *MEG3* is the direct inhibition of the p53 inhibitor MDM2, which will result in apoptotic activity[14]. Other mechanisms that have been determined include the sponging of miR-21[10,15] – a miRNA promoting proliferation and invasion – or recruitment of PRC2[16,17] to promoter regions of target genes resulting in repressive chromatin marks. Interestingly, in our study *MEG3* is one of the most abundant lincRNA genes in neuroblastoma with predicted modulatory effects on MYCN activity. This could point to the fact that *MEG3* may play an oncogenic role in neuroblastoma. Nevertheless, in 25% of neuroblastoma tumors gain of hypermethylation in the *MEG3* promoter region and loss of expression has already been described in a subset of patients[18]. Perturbation of *MEG3* expression in neuroblastoma cells will be key to better understand its mode of action.

While single lncRNA perturbation studies are key to unravel lncRNA functionality, they are also labor intensive. The high number of candidate lncRNAs emerging from expression based studies has exposed the pressing need for high-throughput functional screening approaches. With the recent emergence of CRISPR technology, opportunities to generate screening platforms to modulate RNA expression levels are ubiquitous. As described in the introduction, CRISPR interference (CRISPRi) allows reversible gene expression repression, enabling inspection of the effect on various phenotypes, including proliferation, survival and differentiation[19,20]. This type of screening assigns functional roles to lncRNAs, after which the most promising genes can be studied individually to elucidate the underlying regulatory mechanisms.

A first large scale effort to determine lncRNA functions with CRISPRi technology investigated 16,401 lncRNA loci in 7 different cell lines[20]. Using 10 sgRNAs per transcription start site, 499 lncRNA loci were identified to play a robust role in cell growth modulation. Importantly, not one single lncRNA was involved in cell growth regulation across all 7 cell types, and 89% of lncRNAs affected cell growth in just one cell type. This reveals the necessity to perform cell- and cancer-type specific screenings, as the cellular context is vital in assessing lncRNA functions. Currently, we are working on a neuroblastoma CRISPRi screening platform to validate the functional relevance of the prioritized lncRNAs. In addition, we performed a highly rigorous selection to prioritize the most promising candidate lncRNA for detailed functional evaluation. This lncRNA, later coined *NESPR*, was identified to play a central role in one of the

most important regulatory networks in neuroblastoma tumors, regulating the noradrenergic cell identity.

4.1.2 Unravelling the role of *NESPR* in neuroblastoma

NESPR is located in *cis* to *PHOX2B*, correlates with *PHOX2B* expression and, similar to *PHOX2B*, shows a highly neuroblastoma-specific expression pattern. This led us to hypothesize that *NESPR* could be involved in the regulation of *PHOX2B* expression and, consequently, the establishment of the noradrenergic cell identity.

Through *in vitro* perturbation of *NESPR* using LNA-modified gapmer ASOs, we were able to reveal a link between *NESPR* and *PHOX2B* expression. We chose an antisense strategy based on ASOs with an LNA modification, since *NESPR* transcripts were located in both the cytoplasm and the nucleus. RNAi-based strategies such as siRNAs have great potency to reduce the expression of cytoplasmic transcripts while effects on nuclear transcripts are less pronounced and more variable[21,22]. A pool of 30 siRNAs combined into a siPOOL was used as an alternative for the ASO-mediated knockdown of *NESPR*. However, the siPOOL only provided knockdown of the cytoplasmic fraction of the lncRNA and had no effect on *PHOX2B* expression, revealing that the function of *NESPR* is mediated by the nuclear fraction. As ASOs possess the ability to efficiently reduce expression in both cellular compartments, further experiments were performed using the ASO approach. The LNA modified gapmer configuration allows the use of lower concentrations and demonstrates a higher stability. Five ASOs were designed against *NESPR*, and the two most potent ones were selected for further evaluation. One of the ASOs binds the last exon of the longest and most abundant transcript, whereas the other binds an intronic region. Using two independent ASOs for functional studies is crucial in order to distinguish on-target from off-target effects. From a research perspective, further reduction of off-target effects and false-positive perturbed target genes can be achieved through the inclusion of more ASOs targeting *NESPR.* In a therapeutic point of view, the ASO needs to be fine-tuned further to minimize off-target effects and a maximize knockdown potency.

Neuroblastoma cells are notoriously hard to transfect, making it difficult to achieve stable and significant knockdown in each treatment, despite optimization. Tests with several types of

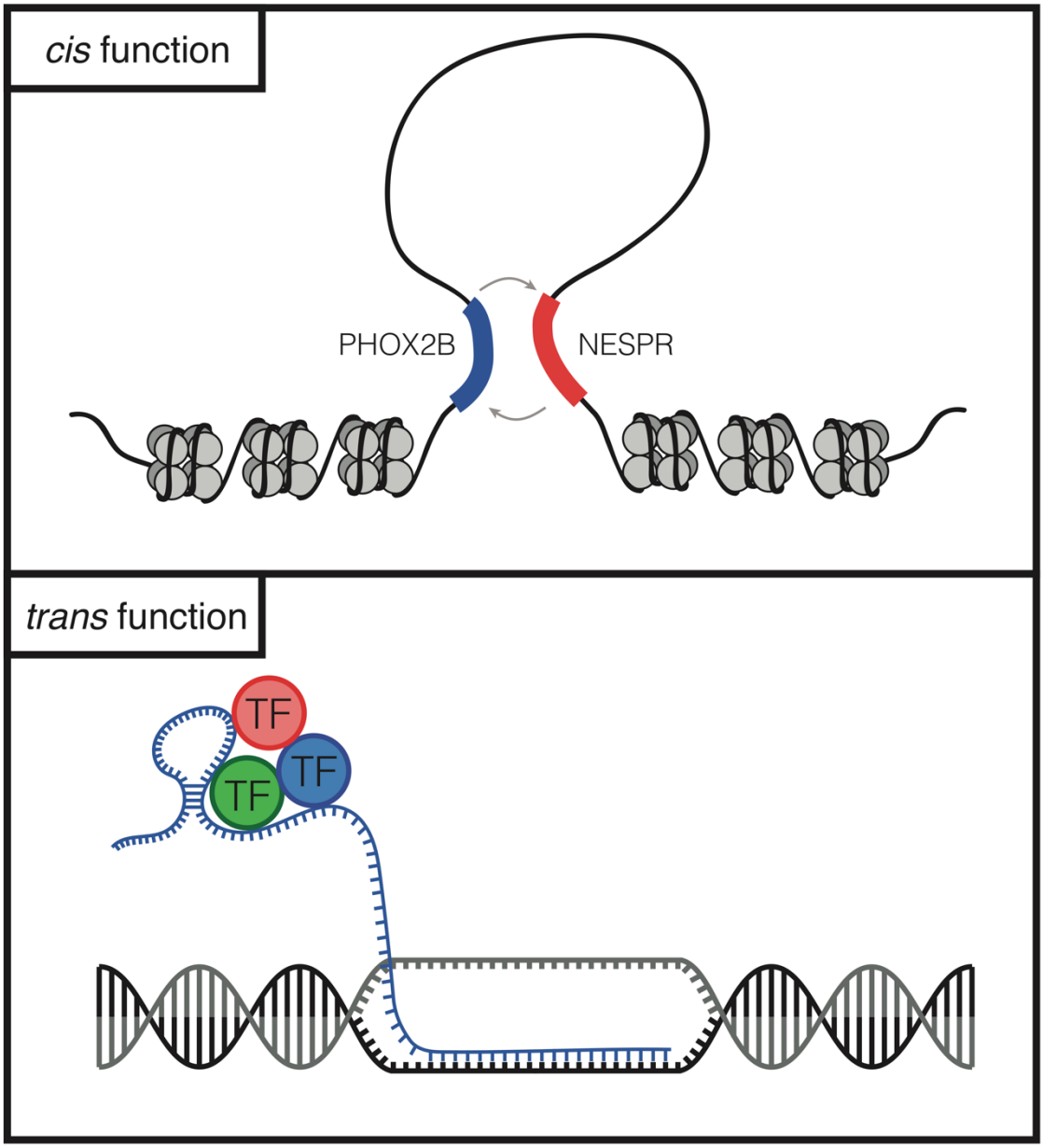


Figure 1: *NESPR* can function in *cis*, regulating *PHOX2B* in a feedforward loop. The regulatory function can be implicated in chromatin loop formation, by changing chromatin structure to bring the enhancer elements into the neighborhood of *PHOX2B*. Results also indicate a *trans* function for *NESPR*, as it shows binding sites throughout the genome. In this configuration, *NESPR* can act as a scaffold for transcription factors, which will regulate gene expression in that genomic locus.

transfection reagents to reach higher knockdown percentages, still resulted in a variable outcome. Consequently, we started looking for other ways reduce *NESPR* expression, such as CRISPRi. The CRISPRi platform confirmed the previous ASO-mediated results.

NESPR is an unusual lncRNA as it is highly conserved on sequence level in 46 other vertebrate species. This high level of conservation gives us the opportunity to investigate its role in animal models, including a hypothetic function of *NESPR* in neuronal development because of the established link between *NESPR* and *PHOX2B*, with *PHOX2B* being an important developmental regulator of the peripheral nervous system[23]. *NESPR* not only has a high sequence conservation between humans and 46 other vertebrate species, but also shows positional conservation and preservation of the number of exons between human and zebrafish. Zebrafish underwent partial genome duplication, causing more than 20% of the genome to be present in two copies[24]. Two zebrafish orthologues of the human *PHOX2B* gene, *Phox2ba* and *Phox2bb*, are present in the zebrafish genome. Surprisingly, in neuroblastoma tumors produced by zebrafish only one orthologue, *Phox2bb*, is expressed. Coincidentally, in that active region downstream of the expressed *Phox2bb* gene, a two-exon *Nespr* transcript can be detected. However, in contrast to the murine *NESPR*, the zebrafish *NESPR* transcript shows no sequence similarity with its human counterpart. This provides us with a unique opportunity to study the impact of sequence conservation versus positional conservation on *NESPR* function.

A recent study published by the group of Alexander Shier[25] aimed to assess the functionality of multiple highly conserved lncRNAs in zebrafish, including *Nespr*. They produced, amongst others, zebrafish containing complete and partial *Nespr* knock-outs through CRISPR/Cas9. They discovered that deletion of the whole gene resulted in jaw deformation and the inability to inflate its swim bladder. However, deletion of the TSS did not result in any malformation or impaired development, producing fertile adults. This could indicate that the deletion of the enhancer elements and the related chromatin marks is potentially the main driver of the abnormal phenotype, whereas actual expression of *Nespr* is not required[25]. However, *Phox2bb* expression levels were not measured, leaving the regulatory impact of *Nespr* on *Phox2bb* expression in zebrafish unresolved. To assess the importance of *Nespr* on *Phox2bb* expression, these knock-out (KO) zebrafish have been procured from the group of Alexander Schier. We are currently trying to replicate the phenotypic results in addition to determining the effect of *Nespr* loss on *Phox2bb* abundance. Simultaneously, we are interested in the oncogenic capacity of *NESPR* in neuroblastoma pathogenesis and wish to unravel its functional role through backcrossing of *MYCN* amplified and *Nespr* KO zebrafish.

Zebrafish are a useful model system because of its transparent and fast developing embryos, high fertility and conservation of biological processes. However, it also has disadvantages such as genome duplications and the large evolutionary distance to humans[26]. In that aspect, rodents are sometimes better suited as a model system for human genetic studies. In collaboration with the VIB, we are currently working on a KO mouse model to study the effects of *NESPR* on mammalian development. As aforementioned, deletion of the entire gene may result in loss of enhancer elements that overlap with *NESPR* exons. This will impede our understanding of the functional role of *NESPR* as we will not be able to distinguish the effect of the lncRNA deletion and the loss of the enhancers. However, deletion of the TSS is also not without risks, due to the presence of a CTCF binding site in the promoter region. An alternative approach is a polyA tail insertion after the TSS, leading to a halted transcription and a truncated RNA molecule. In this manner, the enhancer elements will not be affected and development effects can solely be ascribed to the absence of *NESPR* RNA or *NESPR* transcription.

4.1.2.1 NESPR as a regulator of transcriptional regulation through chromatin looping or transcription factor recruitment?

Our data extensively shows the regulatory connection between *PHOX2B* and *NESPR*, with knockdown of either gene resulting in downregulation of the other. Although the exact mechanism still needs to be elucidated, we can put forward several hypotheses. The first mechanism explaining this proposed feedforward loop is *NESPR* acting as a functional unit of a chromatin remodeling complex. Through recruitment of chromatin modifiers, it can have an initiating role in the formation of the insulated neighborhood. Furthermore, lncRNAs have been described to interact with CTCF[27], a regulator of chromatin structure, which could be an alternative approach towards chromatin looping. These assumptions have to be substantiated *in vitro*, using 4C-seq experiments upon *NESPR* knockdown (figure 2). If *NESPR* is implicated in chromatin loop formation, reduced expression levels should lead to a loss of chromatin interactions. However, quantitative comparison of two different 4C libraries is not straightforward and differences in the level of interaction need to be sufficiently large to allow robust conclusions. To further support the hypothesis of the involvement of *NESPR* in chromatin remodeling, we investigated the interaction of *NESPR* with chromatin modifying

complexes through ChIRP-MS. A second mechanism of regulating *PHOX2B* transcription is by recruiting transcription factors to the enhancer elements located in the *NESPR* locus. Examples of interactions between transcription factors and RNA molecules are still growing in number, suggesting that the RNA is implicated in the regulation of the transcription factor[28]. For example, binding of *RMST* to SOX2 influences transcription of SOX2 targets by acting as a transcriptional coregulator and binding the promoter regions of several SOX2 targets[29]. Techniques including ChIRP-MS and ChIRP-seq help us to further unravel the actual role of *NESPR* in the transcriptional regulation of *PHOX2B* (figure 2).

ChIRP-MS allows us to specifically pull down *NESPR* and its associated RNA binding proteins, by tiling the entire *NESPR* transcript with complementary biotinylated probes. After detachment of the associated proteins, the *NESPR* binding partners are identified through mass spectrometry. In preliminary data, TOP1 and RBMX were found to be binding partners of NESPR. TOP1[30], an antioxidant enzyme, is a critical component in normal development and instrumental in relaxation of supercoiled DNA. Removal of supercoiling alleviates hinderance of the replication fork and might limit R-loop formation, transcription stalling evasion and double strand break formation. Sequestration of TOP1 through binding with *NESPR* could lead to inhibition of the protein. In neuroblastoma, TOP1 is overexpressed in *MYCN* amplified tumors, shows a strong correlation with survival and is positively correlated with the noradrenergic gene expression program. RBMX[31,32] is an RNA-binding motif protein crucial in chromatid segregation and suggested to be necessary for brain development. Loss of the protein causes premature detachment of sister chromatids, leading to aberrant cell division. It has also been described to form a topoisomerase complex together with TOP1 and *NORAD*, which is a lncRNA essential for genomic stability[33]. Similar to TOP1, RBMX is more abundant in *MYCN* amplified tumors, and is correlated with both survival and the noradrenergic network. Other proteins such as NCL and RPA1 that bind *NESPR*, also play a functional role in chromatid separation.

To validate the NESPR-associated proteins found by ChIRP-MS, we will perform RIP-seq[34] (RNA Immunoprecipitation, followed by high-throughput sequencing), a technique used to identify RNA binding proteins (figure 2). By immunoprecipitating the rubonucleoproteins (RNPs) of interest from cell lysates, while maintaining RNA integrity, the associated RNAs can be isolated

and identified through sequencing. CLIP-seq[35,36] is a similar method, but crosslinks the individual subunits of the complex followed by protein digest to elucidate RNA components of RNPs.

In addition to binding to proteins, lncRNAs are also known to bind to DNA, forming RNA:DNA triplexes[37]. They are able to recruit chromatin modifying complexes, methyltransferases or transcription factors to specific genomic locations to perform their function. As mentioned in the introduction, *HOTAIR*, *KCNQ1OT1* and *XIST* are several known lncRNAs with a role in recruitment of protein complexes to their binding sites. Using ChIRP-seq[38], we assessed DNA binding sites of *NESPR* genome-wide, helping us to further unravel its functionality. Motif analysis revealed enrichment of two transcription factors of the noradrenergic core regulatory circuit, i.e. GATA3 and ISL1. GATA3[39] is a transcription factor and critical regulator of normal developmental processes of for instance the nervous system. In addition, multiple cancer types express this protein coding gene, including neuroblastoma and urothelial[40] and breast cancer. Because of the presence of GATA3 in various tissue types and its importance in developmental pathways, its regulation – which is partly mediated by other interacting proteins – must be tightly controlled. *ISL1* is a transcription factor present in sympathetic neurons and plays an important role in their development[41]. The gene is active in progression of several cancer types, among which breast[42] and gastric cancer[43] and neuroblastoma[44]. Several key neuroblastoma genes, such as *ALK, LIN28B, CCND1* and *GATA3* are modulated by ISL1[44,45]. Recently, ISL1's involvement in multiple oncogenic pathways in neuroblastoma was uncovered, together with the physical interaction of the protein with GATA3[44]. This complex regulates several genes, such as *LMO1* and *CTBP2,* crucial for neuroblastoma differentiation and proliferation. However, direct interactions of ISL1 or GATA3 and lncRNAs have not yet been described. Even though in T-helper-2 cells GATA3 is regulated by GATA3-AS1, an RNA molecule divergently transcribed from the same promoter as GATA3[46], direct interactions between the latter or other lncRNAs and the GATA3 protein have not yet been documented. However, as described in the beginning of this subsection, lncRNA-transcription factor interactions exist. In several genomic regions, we found *NESPR* binding sites coinciding with GATA3 and ISL1 binding sites, suggesting a mode of action in which *NESPR* guides the proteins to their respective target sequence or is a coregulator of the transcriptional activity of GATA3 and ISL1.

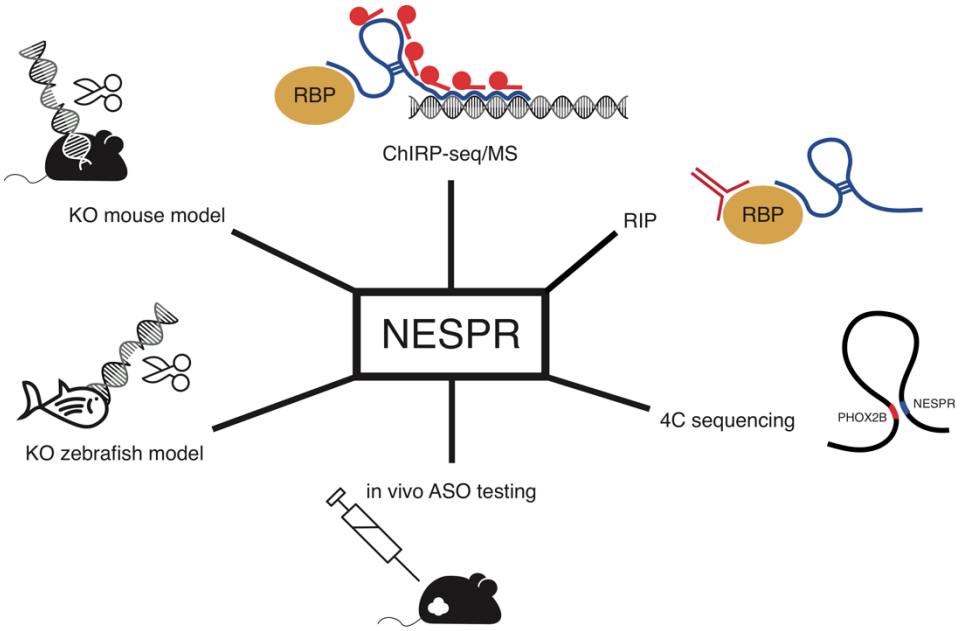


Figure 2: Future perspectives to further unravel the functionality of *NESPR* in chromatin remodeling, transcriptional regulation, development and neuroblastoma initiation and progression.

4.1.2.2 NESPR as a new therapeutic target in neuroblastoma?

In vitro perturbation of *NESPR* results in decreased proliferation and increased caspase 3/7 activity. These are interesting properties to investigate further in an *in vivo* setting by means of xenograft mouse models or genetic mouse models (figure 2). As free-uptake of ASOs is impossible in the majority of neuroblastoma cell lines, it might be that this remains a problem in an *in vivo* setting, making verifying whether a tumor formed through orthotopic injection of NGP neuroblastoma cells is able to take up free ASOs necessary. Furthermore, off-target effects have to be reduced. Unintended binding and subsequent degradation and alteration of gene expression levels through imperfect complementary sequence binding when mismatches are tolerated, has to be avoided. Not surprisingly, the best way to minimize off-target effects is to design ASOs that have high sequence complementarity to the intended target, but lack complementary regions with any other RNA molecule[47]. Furthermore, lowering the binding affinity also leads to a higher sequence-specificity of the ASO. Moreover, lengthening the gapmer increases the chances of introduced mismatches when the gapmer binds to the unintended targets, while maintaining a constant binding affinity by adjusting the number of LNA modifications of the gapmer[47]. As *NESPR* is expressed in certain developing neuronal cell types, this could also lead to on-target toxicity. A therapy targeting *NESPR* where

there are normal developing cells present that depend on the expression of *NESPR*, should be avoided to prevent abnormal development.

4.1.3 Biomarker potential of lncRNAs

A possibility that has not been addressed throughout this scripture is the use of lncRNAs as a neuroblastoma biomarker. A biomarker allows clinicians to personalize and optimize patient treatments, based on quantitative or qualitative analysis, preferably in the least invasive manner. As these molecules are required to be specific and sensitive to be clinically relevant, several of the prioritized neuroblastoma lncRNAs could be potential biomarkers.

Tumor cells often excrete RNAs in circulation where they show a rather high stability by evading nuclease activity[48–51]. They can occur in several configurations in the extracellular environment, dependent on their method of excretion. Excreted lncRNAs can be naked or encapsulated in extracellular vehicles such as exosomes and high-density lipoprotein or apoptotic bodies[52]. LncRNA levels in body fluids often reflect their deregulation in the tumor, making them non-invasive biomarker candidates[52]. Of course, such lncRNAs should circulate in sufficient concentrations to allow robust detection and quantification. The most well-known example of a lncRNA that is used in the clinic is *PCA3.* Expression of this lncRNA is found in urine of prostate cancer patients and should allow a more precise and sensitive diagnosis of prostate cancer. However, recent studies do not confirm a clinical benefit of using this assay in combination with existing tests and show that the implementation of this test would not be cost-effective[53,54]. An example of a potential biomarker is *H19*[55], a lncRNA with elevated expression in plasma samples of gastric cancer patients, because of its rather high sensitivity and specificity in detection of gastric cancer. Furthermore, it shows higher efficacy in detecting early stage gastric cancer than the regularly used biomarkers CEA and CA199 used today. Other promising lncRNAs to be used as biomarkers are *HULC*[51] (plasma – liver cancer) and *HOTAIR*[52,56] (saliva – oral squamous cell carcinoma).

NESPR is a rather highly expressed neuroblastoma-specific lncRNA, making it a potential biomarker in this pediatric malignancy. However, RNA-sequencing of platelet free plasma samples of neuroblastoma patients did not reveal *NESPR* transcripts suggesting it's not

excreted or excreted in too low concentrations to be detectable. Therefore, *NESPR* is likely not applicable as a biomarker in plasma, but other lncRNAs from our study could.

4.2 Conclusions

Only a handful of lncRNAs have been described as potential oncogenes or tumor suppressors in neuroblastoma. By exploring the lncRNome of this malignancy, we provide a core set of lncRNAs with a potential implication on neuroblastoma pathogenesis. However, these genes have been selected through an *in-silico* approach and await experimental validation of their suspected functions. We investigated the functional role and mechanism of our top candidate *NESPR*, highlighting its role in *PHOX2B* regulation and potential involvement in cell growth and survival. Further experimental work focusing on the other lncRNAs associated with neuroblastoma driver genes, will unravel their importance in neuroblastoma and could lead to new therapeutic and biomarker targets.

References

1. Trigg, R. M. & Turner, S. D. ALK in Neuroblastoma: Biological and Therapeutic Implications. *Cancers (Basel).* **10**, 113 (2018).
2. Chen, Y. *et al.* Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* **455**, 971 (2008).
3. Huang, M. & Weiss, W. A. Neuroblastoma and MYCN. *Cold Spring Harb. Perspect. Med.* **3**, a014415–a014415 (2013).
4. Peifer, M. *et al.* Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **526**, 700 (2015).
5. Bown, N. Neuroblastoma tumour genetics: clinical and biological aspects. *J. Clin. Pathol.* **54**, 897–910 (2001).
6. Bown, N. *et al.* Gain of Chromosome Arm 17q and Adverse Outcome in Patients with Neuroblastoma. *N. Engl. J. Med.* **340**, 1954–1961 (1999).
7. De Brouwer, S. *et al.* Meta-analysis of Neuroblastomas Reveals a Skewed ALK Mutation Spectrum in Tumors with MYCN Amplification. *Clin. Cancer Res.* **16**, 4353 LP – 4362 (2010).
8. Boeva, V. *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* **49**, 1408 (2017).
9. Sun, L., Li, Y. & Yang, B. Downregulated long non-coding RNA MEG3 in breast cancer regulates proliferation, migration and invasion by depending on p53's transcriptional activity. *Biochem. Biophys. Res. Commun.* **478**, 323–329 (2016).
10. Li, Z., Yang, L., Liu, X., Nie, Z. & Luo, J. Long noncoding RNA MEG3 inhibits proliferation of chronic myeloid leukemia cells by sponging microRNA21. *Biomed. Pharmacother.* **104**, 181–192 (2018).
11. Zheng, Q. *et al.* Long noncoding RNA MEG3 suppresses liver cancer cells growth through inhibiting β-catenin by activating PKM2 and inactivating PTEN. *Cell Death Dis.* **9**, 253 (2018).
12. Wang, P., Chen, D., Ma, H. & Li, Y. Long non-coding RNA MEG3 regulates proliferation and apoptosis in non-small cell lung cancer via the miR-205-5p/LRP1 pathway. *RSC Adv.* **7**, 49710–49719 (2017).
13. Luo, G. *et al.* Long Non-Coding RNA MEG3 Inhibits Cell Proliferation and Induces Apoptosis in Prostate Cancer. *Cell. Physiol. Biochem.* **37**, 2209–2220 (2015).
14. Zhou, Y. *et al.* Activation of p53 by MEG3 Non-coding RNA. *J. Biol. Chem.* **282**, 24731–24742 (2007).
15. Zhang, J. *et al.* Long noncoding RNA MEG3 is downregulated in cervical cancer and affects cell proliferation and apoptosis by regulating miR-21. *Cancer Biol. Ther.* **17**, 104–113 (2015).
16. Terashima, M., Tange, S., Ishimura, A. & Suzuki, T. MEG3 Long Noncoding RNA Contributes to the Epigenetic Regulation of Epithelial-Mesenchymal Transition in Lung Cancer Cell Lines. *J. Biol. Chem.* **292**, 82–99 (2017).
17. Mondal, T. *et al.* MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA–DNA triplex structures. *Nat. Commun.* **6**, 7743 (2015).
18. Astuti, D. *et al.* Epigenetic alteration at the DLK1-GTL2 imprinted domain in human neoplasia: analysis of neuroblastoma, phaeochromocytoma and Wilms' tumour. *Br. J. Cancer* **92**, 1574–1580 (2005).
19. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of

gene expression. *Nat. Protoc.* **8**, 2180–2196 (2013).
20. Liu, S. J. *et al.* CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **355**, aah7111 (2017).
21. Chiu, Y.-L., Ali, A., Chu, C., Cao, H. & Rana, T. M. Visualizing a Correlation between siRNA Localization, Cellular Uptake, and RNAi in Living Cells. *Chem. Biol.* **11**, 1165–1175 (2004).
22. Lennox, K. A. & Behlke, M. A. Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides. *Nucleic Acids Res.* **44**, 863–877 (2016).
23. Pattyn, A., Morin, X., Cremer, H., Goridis, C. & Brunet, J.-F. The homeobox gene Phox2b is essential for the development of autonomic neural crest derivatives. *Nature* **399**, 366 (1999).
24. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
25. Goudarzi, M., Berg, K., Pieper, L. M. & Schier, A. F. Long non-coding RNAs are largely dispensable for zebrafish embryogenesis, viability and fertility. *bioRxiv* 374702 (2018). doi:10.1101/374702
26. Lieschke, G. J. & Currie, P. D. Animal models of human disease: zebrafish swim into view. *Nat. Rev. Genet.* **8**, 353 (2007).
27. Saldaña-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* **28**, 723–734 (2014).
28. Long, Y., Wang, X., Youmans, D. T. & Cech, T. R. How do lncRNAs regulate transcription? *Sci. Adv.* **3**, eaao2110–eaao2110 (2017).
29. Ng, S.-Y., Bogu, G. K., Soh, B. S. & Stanton, L. W. The Long Noncoding RNA RMST Interacts with SOX2 to Regulate Neurogenesis. *Mol. Cell* **51**, 349–359 (2013).
30. Li, M. & Liu, Y. Topoisomerase I in Human Disease Pathogenesis and Treatments. *Genomics. Proteomics Bioinformatics* **14**, 166–171 (2016).
31. Cho, Y., Ideue, T., Nagayama, M., Araki, N. & Tani, T. RBMX is a component of the centromere noncoding RNP complex involved in cohesion regulation. *Genes to Cells* **23**, 172–184 (2018).
32. Matsunaga, S. *et al.* RBMX: A Regulator for Maintenance and Centromeric Protection of Sister Chromatid Cohesion. *Cell Rep.* **1**, 299–308 (2012).
33. Munschauer, M. *et al.* The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature* **561**, 132–136 (2018).
34. Lu, Z., Guan, X., Schmidt, C. A. & Matera, A. G. RIP-seq analysis of eukaryotic Sm proteins identifies three major categories of Sm-containing ribonucleoproteins. *Genome Biol.* **15**, R7 (2014).
35. Ule, J. *et al.* CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science (80-.).* **302**, 1212 LP – 1215 (2003).
36. Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: A method for identifying protein–RNA interaction sites in living cells. *Methods* **37**, 376–386 (2005).
37. Li, Y., Syed, J. & Sugiyama, H. RNA-DNA Triplex Formation by Long Noncoding RNAs. *Cell Chem. Biol.* **23**, 1325–1333 (2016).
38. Chu, C., Quinn, J. & Chang, H. Y. Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.* 3912 (2012). doi:10.3791/3912

39. Chou, J., Provot, S. & Werb, Z. GATA3 in development and cancer differentiation: cells GATA have it! *J. Cell. Physiol.* **222**, 42–49 (2010).
40. Liu, H., Wilkerson, M. L., Lin, F. & Shi, J. Immunohistochemical Evaluation of GATA3 Expression in Tumors and Normal Tissues: A Useful Immunomarker for Breast and Urothelial Carcinomas. *Am. J. Clin. Pathol.* **138**, 57–64 (2012).
41. Sun, Y. *et al.* A central role for Islet1 in sensory neuron development linking sensory and spinal gene regulatory programs. *Nat. Neurosci.* **11**, 1283–1293 (2008).
42. Zhang, Y. *et al.* ISL1 promotes cancer progression and inhibits cisplatin sensitivity in triple-negative breast cancer cells. *Int. J. Mol. Med.* **42**, 2343–2352 (2018).
43. Guo, T. *et al.* ISL1 predicts poor outcomes for patients with gastric cancer and drives tumor progression through binding to the ZEB1 promoter together with SETD7. *Cell Death Dis.* **10**, 33 (2019).
44. Zhang, Q. *et al.* Collaborative ISL1/GATA3 interaction in controlling neuroblastoma oncogenic pathways overlapping with but distinct from MYCN. *Theranostics* **9**, 986–1000 (2019).
45. Zhang, Q. *et al.* Temporal requirements for ISL1 in sympathetic neuron proliferation, differentiation, and diversification. *Cell Death Dis.* **9**, 247 (2018).
46. Gibbons, H. R. *et al.* Divergent lncRNA GATA3-AS1 Regulates GATA3 Transcription in T-Helper 2 Cells . *Frontiers in Immunology* **9**, 2512 (2018).
47. Hudlebusch, H. *et al.* Identifying and avoiding off-target effects of RNase H-dependent antisense oligonucleotides in mice. *Nucleic Acids Res.* **46**, 5366–5380 (2018).
48. Tang, J. *et al.* Circulation long non-coding RNAs act as biomarkers for predicting tumorigenesis and metastasis in hepatocellular carcinoma. *Oncotarget* **6**, 4505–4515 (2015).
49. Reis, E. M. & Verjovski-Almeida, S. Perspectives of Long Non-Coding RNAs in Cancer Diagnostics. *Front. Genet.* **3**, 32 (2012).
50. ARITA, T. *et al.* Circulating Long Non-coding RNAs in Plasma of Patients with Gastric Cancer. *Anticancer Res.* **33**, 3185–3193 (2013).
51. Xie, H., Ma, H. & Zhou, D. Plasma HULC as a promising novel biomarker for the detection of hepatocellular carcinoma. *Biomed Res. Int.* **2013**, 136106 (2013).
52. Bolha, L., Ravnik-Glavač, M. & Glavač, D. Long Noncoding RNAs as Biomarkers in Cancer. *Dis. Markers* **2017**, 7243968 (2017).
53. Nicholson, A. *et al.* The clinical effectiveness and cost-effectiveness of the PROGENSA® prostate cancer antigen 3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and economic evaluation. *Heal. Technol Assess* **19**, (2015).
54. Helsmoortel, H., Everaert, C., Lumen, N., Ost, P. & Vandesompele, J. Detecting long non-coding RNA biomarkers in prostate cancer liquid biopsies: Hype or hope? *Non-coding RNA Res.* **3**, 64–74 (2018).
55. Zhou, X., Yin, C., Dang, Y., Ye, F. & Zhang, G. Identification of the long non-coding RNA H19 in plasma as a novel biomarker for diagnosis of gastric cancer. *Sci. Rep.* **5**, 11516 (2015).
56. Tang, H., Wu, Z., Zhang, J. & Su, B. Salivary lncRNA as a potential marker for oral squamous cell carcinoma diagnosis. *Mol. Med. Rep.* 761–766 (2013). doi:https://doi.org/10.3892/mmr.2012.1254

Samenvatting

Neuroblastoom, een pediatrische extracraniale tumor, ontstaat uit progenitor cellen uit de neurale kam. Onder normale omstandigheden ontwikkelen deze cellen zich verder tot het sympathische zenuwstelsel. Deze raadselachtige ziekte vertoont een gevarieerd ziekteverloop, gaande van spontaan regresserende tumoren tot agressieve tumoren die in een hoog aantal patiënten tot sterfte leiden. Ondanks multimodale behandeling van hoog-risicopatiënten blijft de overlevingskans van deze kinderen met neuroblastoom nog steeds zeer laag. In een poging de kansen op overleving te vergroten, probeert men gerichtere behandelingen aan te bieden. Maar hiervoor zijn, naast de gekende genomische aberraties, verdere inzichten nodig in de genetische achtergrond van deze kanker. Omwille van de frequente aanwezigheid van *copy number* variaties, wordt neuroblastoom omschreven als een *copy number* aandoening. Deleties van chromosoomarmen 1p en 11q, en amplificatie van 17q worden het vaakst waargenomen. In tegenstelling tot structurele veranderingen, zijn mutaties in een veel kleiner percentage vertegenwoordigt in neuroblastoom. *ALK* is de meest voorkomende mutatie in deze pediatrische kanker en vertoont slechts een incidentie van 8-10%. Andere belangrijke genen zijn *MYCN*, *PHOX2B* en *TERT*. *MYCN* is in ongeveer 25% van de gevallen geamplificeerd, *PHOX2B* is belangrijk in normale ontwikkeling en *TERT* vertoont hoge expressieniveaus in hoog-risico tumoren en onderhoudt de telomeer lengte. Deze oncogene *driver* genen zijn echter proteïne-coderende genen, het niet-coderende transcriptoom in neuroblastoom werd tot nu toe vrijwel nog niet onderzocht.

Om lange intergenische niet-coderende RNAs (lincRNAs) met een oncogeen of tumor-suppressief potentieel te identificeren, concentreerden we ons op lincRNA's die een associatie vertoonden met *MYCN*, *ALK* of *PHOX2B*. Door de combinatie van RNA-seq data van primaire tumoren met modelsystemen voor *MYCN*, *ALK* en *PHOX2B*, toonden we aan dat elk van deze *driver* genen een selecte set lincRNA's reguleert. Verschillende van deze lincRNA's zijn geassocieerd met overlevingskans en het ziektestadium van de patiënt. De lincRNA's werden verder geprioriteerd op basis van hun locatie in super-enhancers en neuroblastoom specifieke expressie. Dit resulteerde in een lijst van genen met een potentiële impact op neuroblastoom tumorgenese. Daarnaast identificeerden we lincRNA's die mogelijks *MYCN* of *PHOX2B*

activiteit en/of expressie reguleren. Verschillende *c-MYC* geassocieerde genen, waaronder *MEG3* en *TSIX*, werden in deze studie als modulators van *MYCN*-activiteit in neuroblastoom teruggevonden. Genen die via deze computationele analyse werden geïdentificeerd, zijn aangerijkt in verschillende *hallmark* gen sets. Een voorbeeld hiervan is de IL6-JAK-STAT3 gen set, een netwerk dat vaak actief is in kankers met een slechtere prognose.

De resultaten van deze studie zijn echter predicties en dienen nog experimenteel gevalideerd worden. Bij het verklaren van de functie van de geselecteerde lncRNA's in neuroblastoom pathogenese, hebben we ons eerst gefocust op onze top-kandidaat *NESPR*. *NESPR*, gelegen in de *PHOX2B* locus, is het meest neuroblastoom-specifieke lincRNA en is geassocieerd met zowel de overlevingskans van de patiënt als *MYCN* amplificatie. Perturbatie van het gen door middel van antisense oligonucleotides (ASO) leidde tot een afname in *PHOX2B* expressie. Ook 780 andere genen vertoonden een verlaagde expressie na *NESPR* knockdown. Het verlies van *NESPR* expressie heeft een negatief effect op celgroei en kolonie-vorming, terwijl we een inductie van apoptosis waarnemen door een verhoging van de caspase activiteit. Ondanks het feit dat het exacte mechanisme van de *NESPR* werking nog niet gekend is, vermoeden we dat het een rol speelt in de vorming van secundaire chromatine structuren, waarmee het *PHOX2B* expressie kan reguleren. Daarnaast zijn er indicaties dat *NESPR* ook een *PHOX2B*-onafhankelijke functie heeft. Er werd immers een aanrijking aan GATA3 en ISL1 sequentie motieven gevonden in *NESPR*-bindende regio's. GATA3 en ISL1 zijn twee super-enhancer geassocieerde transcriptiefactoren in het noradrenergische netwerk die instaan voor de bepaling van cel identiteit.

Samengevat leidde deze studie uiteindelijk tot een lijst van lincRNA's geassocieerd met cruciale genen in neuroblastoom. Uit deze lijst werd één van de top-kandidaten experimenteel gevalideerd teneinde zijn functionele activiteit in deze pediatrische kanker te verduidelijken.

Summary

Neuroblastoma is a pediatric extracranial tumor, derived from neural crest progenitor cells normally giving rise to the sympathetic nervous system. This enigmatic malignancy shows clinical outcomes ranging from spontaneous regression and long-term survival, to aggressive high-risk tumors resulting in death in a high percentage of cases. Although patients having the worst prognosis are subjected to multimodal therapy, the survival rate still remains poor. To cater to the specific needs of the patients, a more targeted approach is preferred. However, targeted strategies require insight into the genetic composition of the cancer in order to identify new therapeutic targets. Neuroblastoma is classified as a copy number disease as copy number variations are frequently observed and contribute to the disease progression and prognosis. The most common copy number variations include 1p deletions, 11q deletions and 17q gains. Although these structural alterations show a high prevalence in neuroblastoma, the mutational incidence is considerably smaller. *ALK* is the most frequently mutated gene, with mutations observed in 8-10% of neuroblastoma patients. Other key driver genes that have been identified are *MYCN*, *PHOX2B* and *TERT*. *MYCN* is amplified in around 25% of cases, *PHOX2B* is a gene instrumental in normal development and *TERT* has been shown to be upregulated in high-risk neuroblastoma tumors leading to telomere elongation. However, these oncogenic drivers are protein coding genes, the non-coding part of the genome remains rather unexplored in neuroblastoma.

To identify long intergenic non-coding RNAs (lincRNAs) with an oncogenic or tumor suppressive potential, we focused on lincRNAs having a link with *MYCN*, *ALK* or *PHOX2B*. By combining RNA-seq data from primary tumor samples with model systems for *MYCN*, *ALK* and *PHOX2B*, we could demonstrate that each of these driver genes regulates a distinct lincRNA set. Several of these lincRNAs were found to be associated with patient survival and disease stage. LincRNAs located in a super-enhancer or expressed specifically in neuroblastoma cells were further prioritized, leading to a core set of genes with potential implications in neuroblastoma tumorigenesis. In addition, we examined the data to find lincRNAs in control of *MYCN* or *PHOX2B* activity or expression, resulting in the identification of several lincRNAs that are known to be associated with c-MYC. Examples include *TSIX* and *MEG3,* that we

classified as modulators of *MYCN* activity in neuroblastoma. Targets established by means of this computational workflow were enriched in several cancer hallmark gene sets. An example of such a hallmark is the IL6-JAK-STAT3 pathway, a network often hyperactivated in several cancer types leading to a poor outcome.

However, these results are predictions and still need to be experimentally validated. To this end, we aimed to unravel the function of our top candidate in neuroblastoma pathogenesis. *NESPR*, located in the vicinity of *PHOX2B*, is the most specific lincRNA in this pediatric malignancy, associated with survival and MYCN amplification. Antisense oligonucleotide (ASO) mediated gene perturbation resulted in a downregulation of *PHOX2B* expression and altered expression patterns for 780 other genes. Loss of *NESPR* expression reduced cell growth and colony formation capacity, while an increase in caspase activity suggests an induction of apoptosis. Although the precise mechanism of *NESPR* is not yet determined, we suspect a role in *PHOX2B* gene expression regulation through chromatin structure modification. However, analysis of *NESPR* binding sites also suggests a PHOX2B independent function. *NESPR* binding sites were shown to be enriched for GATA3 and ISL1 motif sequences, two super-enhancer associated transcription factors contained in the noradrenergic core regulatory circuit in charge of neuroblastoma cell identity.

In conclusion, by investigating the non-coding transcriptome of neuroblastoma we have provided a prioritized core set of lincRNAs associated with key driver genes in neuroblastoma. We experimentally validated one of the top candidate genes, verifying its functional activity in this pediatric malignancy.

Curriculum Vitae

Name:	Dries Rombaut
Gender:	Male
Date of Birth:	02-11-1990
Place of Birth:	Dendermonde, Belgium
Nationality:	Belgian

Contact: dries.rombaut@ugent.be

Work experience

2013-present PhD student in biomedical sciences
Ghent University, Faculty of Medicine and Health Sciences, Center for Medical Genetics Ghent, Vandesompele Lab
'Identification of neuroblastoma specific lncRNAs'

Education

2011-2013 Master of Science: Bioscience Engineering: cell and gene biotechnology, Ghent University
Master thesis: 'Circulating miRNAs in serum and plasma as non-invasive biomarkers in neuroblastoma', Faculty of Bioengineering Science/Faculty of Medical and Health Sciences, Ghent University

2012 Interuniversity student at the Free University of Brussels
Course in Biomedical Engineering techniques (Drug Design, Pharmacokinetics, Cellular Engineering) and Biomedical Technologies and Parasitology

2008-2011 Bachelor of Bioscience Engineering, Ghent University

2002-2008 General secondary education
Latin – Mathematics
OLVP Bornem, Bornem, Belgium

Courses and certificates

2016 Effective Graphical Displays, Jean-Luc Doumont, Ghent University

2016 Effective Scientific Communication, Jean-Luc Doumont, Ghent University

2015 HPC Unix command line, shell scripting and HPC basics, Ghent University

2015 Laboratory animal science – Faculty of Medicine and Health Sciences, Ghent University.
The course has been recognized by the Belgian Federal Public Service of Public Health, Food Chain Sagety and Environment as being in compliance with Annex VIII (education of persons responsible for directing animal experiments) of the Royal Decree of the 14th of November 1993 covering the protection of experimental animals.

Awards and grants

Best oral presentation, Oncopoint, 2019
The neuroblastoma-specific lncRNA NESPR controls noradrenergic cell identity

Best stormsession oral presentation, Oncopoint, 2017
lnc-PHOX2B-2: a highly specific lncRNA in neuroblastoma

Best stormsession oral presentation, Oncopoint, 2015
Long non-coding RNAs as upstream regulators and downstream effectors of TP53 activity in neuroblastoma

IWT PhD fellowship renewal grant, Brussels, Belgium, 2016
Identification of candidate therapeutic lncRNAs in the TP53 pathway

IWT PhD fellowship grant, Brussels, Belgium, 2014
Identification of candidate therapeutic lncRNAs in the TP53 pathway

Peer-reviewed publications (A1)

Wallaert, A., Durinck, K., Van Loocke, W., Van de Walle, I., Matthijssens, F., Volders, P.-J., Avila Cobos, F., Rombaut, D., Rondou, P., Mestdagh, P., Vandesompele, J., Poppe, B., Taghon, T., Soulier, J., Van Vlierberghe, P., Speleman, F. (2016).
Long noncoding RNA signatures define oncogenic subtypes in T-cell acute lymphoblastic leukemia, LEUKEMIA, 30, 1927-1930

Avila Cobos, F., Anckaert J., Volders, PJ., Everaert C., Rombaut D., et al. (2017).
Zipper plot: visualizing transcriptional activity of genomic regions. BMC Bioinformatics 18, 231.

Bieke Decaesteker, Geertrui Denecker, Christophe Van Neste, Emmy M. Dolman, Wouter Van Loocke, Moritz Gartlgruber, Carolina Nunes, Fanny De Vloed, Pauline Depuydt, Karen Verboom, Dries Rombaut, Siebe Loontiens, Jolien De Wyn, et al. (2018).
TBX2 is a neuroblastoma core regulatory circuitry component enhancing MYCN/FOXM1 reactivation of DREAM targets, Nature Communications, 9, 4866

Karen Verboom, Celine Everaert, Nathalie Bolduc, Kenneth J Livak, Nurten Yigit, Dries Rombaut, et al.
SMARTer single cell total RNA sequencing, revision submitted for publication in NAR methods

Dries Rombaut, Hua-Sheng Chiu, Bieke Decaesteker, Celine Everaert, Nurten Yigit, Agathe Peltier, Isabelle Janoueix, Christoph Bartenhagen, Matthias Fischer, Stephen Roberts, Nicky D'Haene, Katleen De Preter, Frank Speleman, Geertrui Denecker, Pavel Sumazin, Jo Vandesompele, Steve Lefever, Pieter Mestdagh
Integrative analysis identifies lincRNAs up- and downstream of neuroblastoma driver genes, Scientific Reports, 2019

Oral presentations

Rombaut Dries, Lefever Steve, Koster Jan, Schulte Johannes, Versteeg Rogier, van Sluis Peter, Vandesompele Jo, Mestdagh Pieter
Long non-coding RNAs as upstream regulators and downstream effectors of TP53 activity in neuroblastoma, Stormsession Oncopoint, February 11, 2015, Ghent, Belgium

Dries Rombaut, Karen Verboom, Robrecht Cannoodt, Nurten Yigit, Celine Everaert, Katleen De Preter, Jo Vandesompele
TP53 regulated genes identified through Single Cell activation analysis, Single Cell workshop, January 15, 2016, Ghent, Belgium

Rombaut Dries, Lefever Steve, Koster Jan, Schulte Johannes, Versteeg Rogier, van Sluis Peter, Vandesompele Jo, Mestdagh Pieter
lncRNAs as novel components in the TP53 pathway, Stormsession Oncopoint, March 2, 2016, Ghent, Belgium

Dries Rombaut, Steve Lefever, Nurten Yigit, Celine Everaert, Jo Vandesompele, Pieter Mestdagh
lnc-PHOX2B-2: a highly specific lncRNA in neuroblastoma, Stormsession Oncopoint, March 15, 2017, Ghent, Belgium

Dries Rombaut, Louis Delhaye, Tiago França Brazao, Eva D'haene, Giorgio Milazzo, Roberto Ciaccio, Nurten Yigit, Agathe Peltier, Isabelle Janoueix, Celine Everaert, Matthias Fischer, Thorsten Simon, Johan Van Nes, Rogier Versteeg, Björn Menten, Giovanni Perrini, Jo Vandesompele, Sven Eyckerman, Steve Lefever, Pieter Mestdagh
The neuroblastoma-specific lncRNA NESPR is a component of the noradrenergic core regulatory circuit and is essential for neuroblastoma cell survival, Advances in Neuroblastoma Research (ANR), May 9th-15th, 2018, San Francisco, USA

Eric de Bony, Dries Rombaut, Louis Delhaye, Tiago França Brazao, Eva D'haene, Giorgio Milazzo, Roberto Ciaccio, Nurten Yigit, Agathe Peltier, Isabelle Janoueix, Celine Everaert, Matthias Fischer, Thorsten Simon, Johan Van Nes, Rogier Versteeg, Björn Menten, Giovanni Perrini, Jo Vandesompele, Sven Eyckerman, Steve Lefever, Pieter Mestdagh
The neuroblastoma-specific lincRNA NESPR controls noradrenergic cell identity, Oncopoint, March 20, 2019, Ghent, Belgium

Conference posters

Rombaut Dries, Lefever Steve, Koster Jan, Schulte Johannes, Versteeg Rogier, van Sluis Peter, Vandesompele Jo, Mestdagh Pieter
lncRNAs as upstream regulators and downstream effectors of TP53 in neuroblastoma, Keystone meeting, lncRNAs: From evolution to function, March 15th-21st, 2015, Keystone, Colorado, USA

Rombaut Dries, Lefever Steve, Koster Jan, Schulte Johannes, Versteeg Rogier, van Sluis Peter, Vandesompele Jo, Mestdagh Pieter
lncRNAs as novel components in the TP53 pathway, IUAP meeting, June 8th, 2016, Brussels, Belgium

Rombaut Dries, Lefever Steve, Koster Jan, Schulte Johannes, Versteeg Rogier, van Sluis Peter, Vandesompele Jo, Mestdagh Pieter
lncRNAs as novel components in the TP53 pathway, Advances in Neuroblastoma Research, 2016, Cairns, Australia

Other abstracts

Karen Verboom, Celine Everaert, Nathalie Bolduc, Kenneth Livak, Nurten Yigit, Dries Rombaut, Jasper Anckaert, Morten Venø, Jørgen Kjems, Franki Speleman, et al., *SMARTer single cell total RNA sequencing*, Single Cell Biology, 2019, Breckenridge, Colorado, USA

Bieke Decaesteker, Geertrui Denecker, Christophe Van Neste, Fanny De Vloed, Moritz Gartlgruber, Pauline Depuydt, Wouter Van Loocke, Shana Claeys, Dries Rombaut, Siebe Loontiens, et al., *The core regulatory circuit component TBX2 is implicated in cell cycle control and proliferation in neuroblastoma*, Advances in Neuroblastoma Research, 2018, San Francisco, USA

Francisco Avila Cobos, Jasper Anckaert, Pieter-Jan Volders, Celine Everaert, Dries Rombaut, Jo Vandesompele, Katleen De Preterand Pieter Mestdagh, *Zipper plot: visualizing transcriptional activity of genomic regions*, Belgian Society for Human Genetics, 2018, Ghent, Belgium

Bieke Decaesteker, Geertrui Denecker, Christophe Van Neste, Fanny De Vloed, Moritz Gartlgruber, Pauline Depuydt, Wouter Van Loocke, Shana Claeys, Dries Rombaut, Siebe Loontiens, et al., *TBX2 is implicated in cell cycle control and proliferation in neuroblastoma*, Belgian Society for Human Genetics, 2018, Ghent, Belgium

Bieke Decaesteker, Geertrui, Wouter Van Loocke, Moritz Gartlgruber, Pauline Depuydt, Fanny De Vloed, Shana Claeys, Dries Rombaut, Siebe Loontiens, Anton Henssen, et al., *The core regulatory circuit component TBX2 is implicated in cell cycle control and proliferation in neuroblastoma*, OncoPoint, 2018, Ghent, Belgium

Bieke Decaesteker, Geertrui Denecker, Wouter Van Loocke, Moritz Gartlgruber, Pauline Depuydt, Fanny De Vloed, Shana Claeys, Dries Rombaut, Siebe Loontiens, Anton Henssen, et al., *The core regulatory circuit component TBX2 is implicated in cell cycle control and proliferation in neuroblastoma*, Villa Joep, 7th Scientific progress meeting, 2017, Utrecht, The Netherlands

Bieke Decaesteker, Fanny Devloed, Siebe Loontiens, Geertrui Denecker, Wouter Van Loocke, Moritz Gartlgruber, Dries Rombaut, Anneleen Beckers, Shana Claeys, Anton Henssen, Suzanne Vanhauwaert, Frank Westermann, Johannes Schulte, Katleen De Preter, Frank Speleman
The TBX2 super-enhancer marked transcription factor on 17q is overexpressed in neuroblastoma and infers poor prognosis, f-tales, 2016, Ghent, Belgium

Bieke Decaesteker, Sara De Brouwer, Robrecht Cannoodt, Dries Rombaut, Wouter Van Loocke, Annelies Fieuw, Fanny De Vloed, Pieter-Jan Volders, Björn Menten, Rogier Versteeg, et al., *The genomic region encompassing the overlapping SOX11 and Lnc-SOX11-1:1 loci exhibits a complex DNA copy number and mRNA expression pattern*, OncoPoint, 2016, Ghent, Belgium

Bieke Decaesteker, Sara De Brouwer, Robrecht Cannoodt, Dries Rombaut, Wouter Van Loocke, Annelies Fieuw, Fanny Devloed, Pieter-Jan Volders, Bjorn menten, Rogier Versteeg, Patrick Reynolds, Michael Hogarty, Jan Koster, Frank Speleman, Katleen De Preter
The genomic region encompassing the overlapping SOX11 and lnc-SOX11-1:1 loci exhibits a complex DNA copy number and mRNA expression pattern, OncoPoint, 2016, Ghent, Belgium

Coaching

Thesis supervision

Yannick Van de Vyver, Bachelor in de Biomedische Laboratoriumtechnologie
Identificatie van lncRNA's in de TP53-pathway, 2015
Promotor: Jo Vandesompele

Nienke Heireman, Master of Science in Biomedical Sciences
Therapeutische lncRNA's in de TP53-pathway in neuroblastoom, 2016
Promotor: Steve Lefever

CURRICULUM VITAE

Ayse Bassez, Master of Science in Biomedical Sciences
Identificatie van neuroblastoom-specifieke lncRNA's, 2017
Promotor: Steve Lefever

Personal note

To be honest, this is – in my opinion – the hardest part of my dissertation. Because, how do you thank all the people who helped shape your final work and, at the same time, your personality. I've tried to pull this extraordinary feat off by starting at the beginning: my interview with Jo. The first time we talked, one of the questions he asked me was: 'Do you think you are able to handle stressful situations and setbacks?'. At that time, the question rather seemed irrelevant and I thought I already swam in troubled waters before, but after a while, I started to realize that the question couldn't be more relevant. The fact that I am writing this personal note right now, means that I've gotten a lot of help along the way.

First of all, Jo. Thank you for giving me the opportunity to start this amazing journey, the chance to do high-level research in a supporting environment with amazing colleagues. Your support in the beginning of my PhD meant everything to me. Steve, I will never forget our talks, that might have looked like arguments to other people, but was our way of interacting and, actually, the best way to get your attention. I am grateful for the help you have given me throughout the years and the expertise that you shared with me. Pieter, you are one of the most intelligent scientists I know and I cannot thank you enough for the mentorship you have provided for the last five years. There were many times I've doubted myself because of an abundance of setbacks, but you always had my back and assisted me in getting the research back on track. Your insight and vast knowledge have been indispensable in completing this thesis.

Of course, I cannot only give credit to these three people, as many colleagues have contributed significantly to this work. To the lab technicians, thanks for taking over some of the experiments when it was needed. Nurten, it wasn't always an easy ride to follow my (sometimes chaotic) train of thought, but thank you for sticking with me and powering through. Together we made huge strides forward, that wouldn't have been possible without your input and skill. Bieke and Jolien, I was lucky enough to make one of the most mindboggling trips ever to New Zealand and I am happy you two were a part of it. And Bieke, thank you for not killing us while driving. Eva, Shanna and Jasper, I will never forget our time together in the office and the wonderful memories that those times produced. To the rest of

my colleagues, I have had an amazing time because of you and I hope I will find another work environment in which I will find such joy and pleasure to work.

Aan mijn vrienden, bedankt voor de steun de voorbije jaren. Hoewel ik jullie waarschijnlijk vaak verveelde met verhalen over het werk, zijn jullie er steeds geweest. Ik hoop dat we nog vele mijlpalen samen kunnen beleven. Moeke en vake, bedankt om mij steeds aan te moedigen en mij de mogelijkheid te geven om van jongs af aan mijn interesse in de wetenschap te laten botvieren. Door jullie steun heb ik meer kunnen verwezelijken dan ik ooit voor mogelijk hield en schrijf ik nu dit dankwoord voor mijn thesis.

Als laatste, maar zeker niet minst belangrijke persoon, Céline. De laatste jaren was je er steeds voor mij als ik het nodig had, wanneer het niet liep zoals ik het wou. Na een mindere dag kon ik thuiskomen en mijn frustraties uiten, op een goede dag deelde je mee in de vreugde. Ik hoop dat ik jou even hard kan helpen bij het schrijven van jouw thesis, als jij mij hebt geholpen. Ik kan niet wachten om de rest van ons leven samen door te brengen.

Bedankt.

Dries