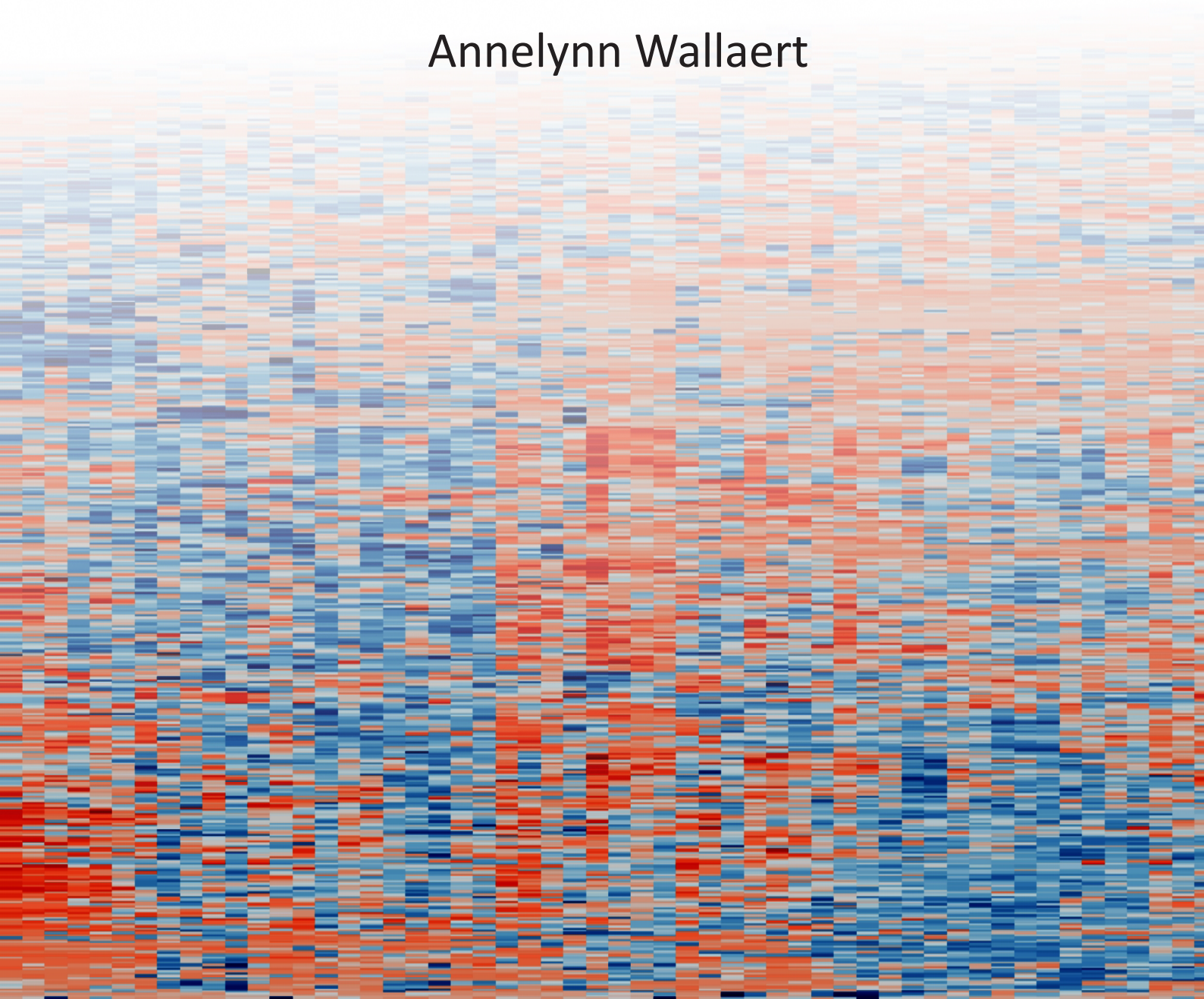
GHENT UNIVERSITY

FACULTY OF MEDICINE AND HEALTH SCIENCES

Deciphering the untranslated message in T-cell acute lymphoblastic leukemia

Annelynn Wallaert



FACULTY OF MEDICINE AND HEALTH SCIENCES

DECIPHERING THE UNTRANSLATED MESSAGE IN T-CELL ACUTE LYMPHOBLASTIC LEUKEMIA

Annelynn Wallaert

Promotor: Prof. dr. Frank Speleman
Co-promotors: Prof. dr. Pieter Van Vlierberghe and dr. Pieter Rondou

This thesis is submitted as fulfillment of the requirements for the degree of Doctor in Medical Sciences, 2017

Center for Medical Genetics Ghent
Ghent University Hospital, Medical Research Building 1
De Pintelaan 185, 9000 Gent, Belgium
annelynn.wallaert@ugent.be

GHENT UNIVERSITY

Thesis submitted to fulfill the requirements for the degree of Doctor in Medical Sciences

Promotor

Prof. dr. Frank Speleman (Ghent University, Belgium)

Co-promotor

Prof. dr. Pieter Van Vlierberghe (Ghent University, Belgium)

dr. Pieter Rondou (Ghent University, Belgium)

Members of the examination committee

dr. Joost Kluiver (University of Groningen, The Netherlands)

dr. Charles De Bock (KU Leuven, Belgium)

Prof. dr. Jan Philippé (Ghent University, Belgium)

Prof. dr. Katleen De Preter (Ghent University, Belgium)

dr. Filip Matthijssens (Ghent University, Belgium)

dr. Christophe Van Neste (Ghent University, Belgium)

Prof. dr. Elfride De Baere, chairman (Ghent University, Belgium)

The research described in this thesis was conducted at the Center for Medical Genetics Ghent, Ghent University, Ghent, Belgium.

This work was supported by the Fund for Scientific Research Flanders (FWO; PhD grant to Annelynn Wallaert and project grants), the GOA-UGent, the Children Cancer Fund Ghent, the Belgian Program of Interuniversity Poles of Attraction (IUAP) and the Belgian Foundation Against Cancer.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

3C	Chromosome Conformation Capture
AF10	ALL1-Fused Gene From Chromosome 10
ALL	Acute Lymphoblastic Leukemia
AML	Acute Myeloid Leukemia
ANRIL	Antisense Noncoding RNA in the INK4 Locus
ASO	Antisense Oligonucleotide
ATAC	Assay for Transposase-Accessible Chromatin
BALR-2	B-ALL Associated lncRNA-2
BCL11B	B-cell Lymphoma/Leukemia 11B
BCL2	B-Cell Lymphoma 2
bHLH	Class B Helix-Loop-Helix
BRCA1/2	Breast Cancer 1/2
C. Elegans	Caenorhabditis elegans
C/EBPα	CCAAT/enhancer-binding protein α
CALM	Phospatidylinositol Binding Clathrin Assembly Protein
cAMP	Cyclic Adenosinemonofosfaat
CBP	CREB Binding Protein
CCAT1	Colon Cancer Associated Transcript 1
CD	Cluster of Differentiation
CDK4/6	Cyclin-dependent Kinase 4/6
CDKN2A/B	Cyclin-dependent Kinase inhibitor 2A/B
ceRNA	Competitive Endogenous RNA
CHART	Capture Hybridization Analysis of RNA Targets
ChIP	Chromatin Immunoprecipitation
ChIRP	Chromatin Isolation by RNA Purification
circRNA	Circular RNA
CLL	Chronic Lymphocytic Leukemia
COLDAIR	Cold Assisted Intronic noncoding RNA
CREB	cAMP Response Element Binding
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRNDE	Colorectal Neoplasia Differentially Expressed
CRNDEP	CRNDE-protein
CYLD	CYLD Lysine 63 Deubiquitinase
DLL	Delta-like Ligand
DNA	Deoxyribonucleic Acid
DNMT3A	DNA (cytosine-5)-Methyltransferase 3A
DTX1	Deltex 1
DUSP5/6	Dual Specificity Phosphatase 5/6
EED	Embryonic Ectoderm Development
EGR1	Early Growth Response protein 1
eRNA	Enhancer RNA
ETP	Early T-cell Progenitor
ETS1	ETS Proto-Oncogene 1
ETV6	ETS Variant 6

EZH2	Enhancer Of Zeste 2
FBXW7	F-box and WD Repeat Domain containing 7
FDA	US Food and Drug Administration
Fenderr	FOXF1 adjacent Non-coding Developmental Regulatory RNA
FLC	Flowering Locus C
GAS5	Growth Arrest Specific 5
GATA3	GATA binding protein 3
GEO	Gene Expression Omnibus
gRNA	Guide RNA
GRα	Glucocorticoid Receptor α
GSI	γ-Secretase Inhibitor
H3K27ac	Histon 3 Lysine 27 acetylation
H3K4me1	Histon 3 Lysine 4 monomethylation
H3K4me3	Histon 3 Lysine 4 trimethylation
HBP1	HMG-Box Transcription Factor 1
HDAC	Histone Deacetylase
HES1	Hes Family BHLH Transcription Factor 1
HHEX	Hematopoietically-expressed Homeobox protein
HOTAIR	HOX Transcript Antisense RNA
HOTTIP	HOXA Transcript at the distal Tip
HOXA	Homeobox A
HULC	Highly Upregulatd in Liver Cancer
ICN1	Intra-cellular NOTCH1
IDH1/2	Isocitrate Dehydrogenase (NADP(+)) 1/2
IGF1R	Insulin-like Growth Factor 1 Receptor
IGN	Imprinted Gene Network
IKZF1	IKAROS Family Zinc Finger 1
IL3	Interleukin 3
IL7R	Interleukin 7 Receptor
JAK1	Janus Kinase 1
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
LMO2	LIM domain Only 2
LNA	Locked Nucleic Acid
lncRNA	Long noncoding RNA
LSD1	Lysine Specific Demethylase 1A
LUNAR1	Leukemia Induced Noncoding Activator RNA 1
MALAT1	Metastasis Associated Lung Adenocarcinoma Transcript 1
MAML1	Mastermind Like Transcriptional Coactivator 1
MBD1	Methyl-CpG-binding Domain protein 1
MDM2	Mouse Double Minute 2 homolog
MEF2C	Myocyte-specific Enhancer Factor 2C
MEG3	Maternally Expressed 3
MHC	Major Histocompatibility Complex
miRNA	Micro RNA
MITF	Microphthalmia-associated Transcription Factor
MLL	Myeloid/Lymphoid or Mixed Lineage Leukemia

mRNA	Messenger RNA
mTOR	Mechanistic Target Of Rapamycin
NALT	NOTCH1 Associated lncRNA in T-ALL
NBAT1	Neuroblastoma-Associated Transcript 1
ncRNA	Noncoding RNA
NEAT1	Nuclear Enriched Abundant Transcript 1
NF-κB	Nuclear Factor κ of activated B cells
NF1	Neurofibromatose type 1
NK cells	Natural Killer cells
N-Me	NOTCH-bound MYC enhancer
NRARP	NOTCH-Regulated Ankyrin Repeat Protein
NRAS	Neuroblastoma RAS viral oncogene homolog
NUP214	Nucleoporin 214
PCAT1	Prostate Cancer Associated Transcript 1
PCGEM	Prostate-Specific Transcript
PDCD4	Programmed Cell Death protein 4
PHF6	Plant Homeodomain Finger protein 6
PI3K	Phosphatidylinositol-4,5-biphospate 3-kinase
Pint	p53 Induced Noncoding Transcript
piRNA	Piwi-interacting RNA
PIWI	P-element Induced Wimpy testis
PKA	Protein Kinase A
PKRAA1	Protein Kinase AMP-Activated Catalytic Subunit Alpha 1
POLII	RNA Polymerase 2
PP2A	Protein Phosphatase 2
PRC2	Polycomb Repressive Complex 2
pre-TCR	Premature T-cell Receptor
PTCRA	Pre T-cell Antigen Receptor α
PTEN	Phosphatase and Tensin Homolog
PTENP1	PTEN Pseudogene 1
PTPN22	Protein Tyrosine Phosphatase, Non-receptor type 22
RACE	Rapid Amplification of cDNA Ends
RAG	Recombination Activated Gene
RAP	RNA Antisense Purification
RB1	Retinoblastoma 1
RBPJ	Recombination Signal Binding Protein for Immunoglobulin Kappa J Region
REST	RE1 Silencing Transcription Factor
RIP	RNA Immunoprecipitation
RISC	RNA-Induced Silencing Complex
RNA	Ribonucleic Acid
rRNA	Ribosomal RNA
RUNX1/3	Runt-related Transcription Factor 1/3
SCHLAP1	SWI/SNF Complex Antagonist Associated With Prostate Cancer 1
SET	SET Nuclear Proto-Oncogene
shRNA	Short Hairpin RNA
siRNA	Short Interfering RNA

snoRNA	Small Nucleolar RNA
SNP	Single Nucleotide Polymorphism
snRNA	Small Nuclear RNA
SRSF2	Splicing Factor, arginine/serine-rich 2
STAT	Signal Transducer And Activator of Transcription
STIL	SCL/TAL1 Interrupting Locus
SUZ12	Suppressor Of Zeste 12 Protein Homolog
T-ALL	T-cell Acute Lymphoblastic Leukemia
TAL-R	TAL-rearranged
TAL1	T-cell Acute Lymphocytic leukemia 1
TCGA	The Cancer Genome Atlas
TCR	T-cell Receptor
TERC	Telomerase RNA component
TLX1/3	T-cell Leukemia Homeobox 1/3
TP53	Tumor Protein P53
TPM	Transcripts Per Kilobase Million
TRIB2	Tribbles Pseudokinase 2
tRNA	Transfer RNA
UTR	Untranslated Region
WDR5	WD Repeat-containing protein 5
WT1	Wilms Tumor 1
XIST	X-Inactive Specific Transcript
ZEB2	Zinc finger E-Box Binding homeobox 2

CHAPTER 1
INTRODUCTION

1 T-CELL ACUTE LYMPHOBLASTIC LEUKEMIA

"Several cases exist with a great excess of white blood cells (…) In fact I believe the excess of white blood cells is due to an arrest of maturation of blood. From my theory on the origin of blood cells, the overabundance of white blood cells should be the result of an arrest of the development of intermediate cells."
-Alfred Donné, 1844-

This text is the first description of leukemia by the hand of a French physician Alfred Donné in 1844[1, 2] and it is quite remarkable that this already states the main features of leukemia and its subtype T-cell acute lymphoblastic leukemia (T-ALL). T-ALL is a malignant disorder in which the accumulation of several oncogenic events leads to a differentiation arrest during T-cell development in the thymus and to an overpopulation of immature T-cells in the blood stream and infiltration of lymphoblasts in the bone marrow. These immature T-cell lymphoblasts do not have the potential to fight infections as normal T-cells have and since they overpopulate other cell types in the blood, these cells cannot function either. Clinical consequences of this perturbed blood cell maturation are fatigue, shortness of breath, multiple infections, bruising and easy bleeding and the presence of mediastinal thymic masses.

In pediatric cancers, acute lymphoblastic leukemia is the most prevalent malignancy. T-cell acute lymphoblastic leukemia (T-ALL) represents 10 to 15% of acute lymphoblastic leukemia (ALL) patients. The cure rates in T-ALL patients have risen tremendously in the last decades due to intensified chemotherapy protocols. The overall survival is nowadays over 80% for pediatric T-ALL and around 60% for adult T-ALL[3, 4]. Despite the quite optimistic survival rates, relapse patients and patients with primary resistant T-ALL have a very bad prognosis and also the long-term side effects of chemotherapy are not to be neglected[4].

1.1 T-cell development

T-cells are essential components of the immune system, which protects the body from infectious agents and the damage they can cause. Mature T-cells in the bloodstream can be subdivided in three main categories with specific functions: (1) cytotoxic $CD8^+$ T-cells detect and kill cells that are infected by viruses or intracellular pathogens; (2) T-helper cells ($CD4^+$) aid in the function of other white blood cells by, for example, activating B-cells to produce antibodies and activating macrophages infected by pathogens; (3) and regulatory T-cells ($CD4^+$) are involved in the suppression of lymphocytes to strictly control the immune response.

Before T-cells can perform their function in the blood stream, they have to go through a very orchestrated maturation process in the thymus, which includes several stages of proliferation, differentiation and selection (Figure 1). First, multipotent hematopoietic progenitor cells have to migrate from the bone marrow to the thymus. These uncommitted early T-cell progenitors

(ETPs) are characterized as $CD34^{+}CD1a^{-}$ and some of these cells can still differentiate into several hematological lineages (lymphoid, myeoloid and in some instances erythroid cells)[5]. Due to the absence of either the CD4 or the CD8 receptor at the membrane, these first stages of T-cell development are also known as "double negative" thymocytes. To induce T-cell specification and proliferation, strong NOTCH-signaling is required, which is introduced by several NOTCH-ligands (especially Delta-like ligand 4, DLL4) present on the membrane of the cortical thymic epithelial cells[6]. The activation of NOTCH1-signaling leads to a subset of NOTCH1-primed ETPs that have high levels of CD7, but still the possibility to develop into natural killer (NK) cells[6, 7]. A recent study by Van de Walle et al. further elucidated the transcriptional program leading to T-cell commitment, with the transcription factor 'GATA binding protein 3' (GATA3) as a main player. GATA3 represses some of the NOTCH1 target genes (ex. *IGF1R*, *DTX1* and *RUNX3*) leading to a partial reduction in NOTCH1 signaling[8]. Next to that, GATA3 represses several stem cell genes (ex. *LMO2* and *MEF2C*) and genes necessary for NK-lineage development and is also involved in the upregulation of transcription factors necessary for the next stages in T-cell development (for example *BCL11B*). Finally, the completion of T-cell commitment is marked by the expression of CD1a[9].

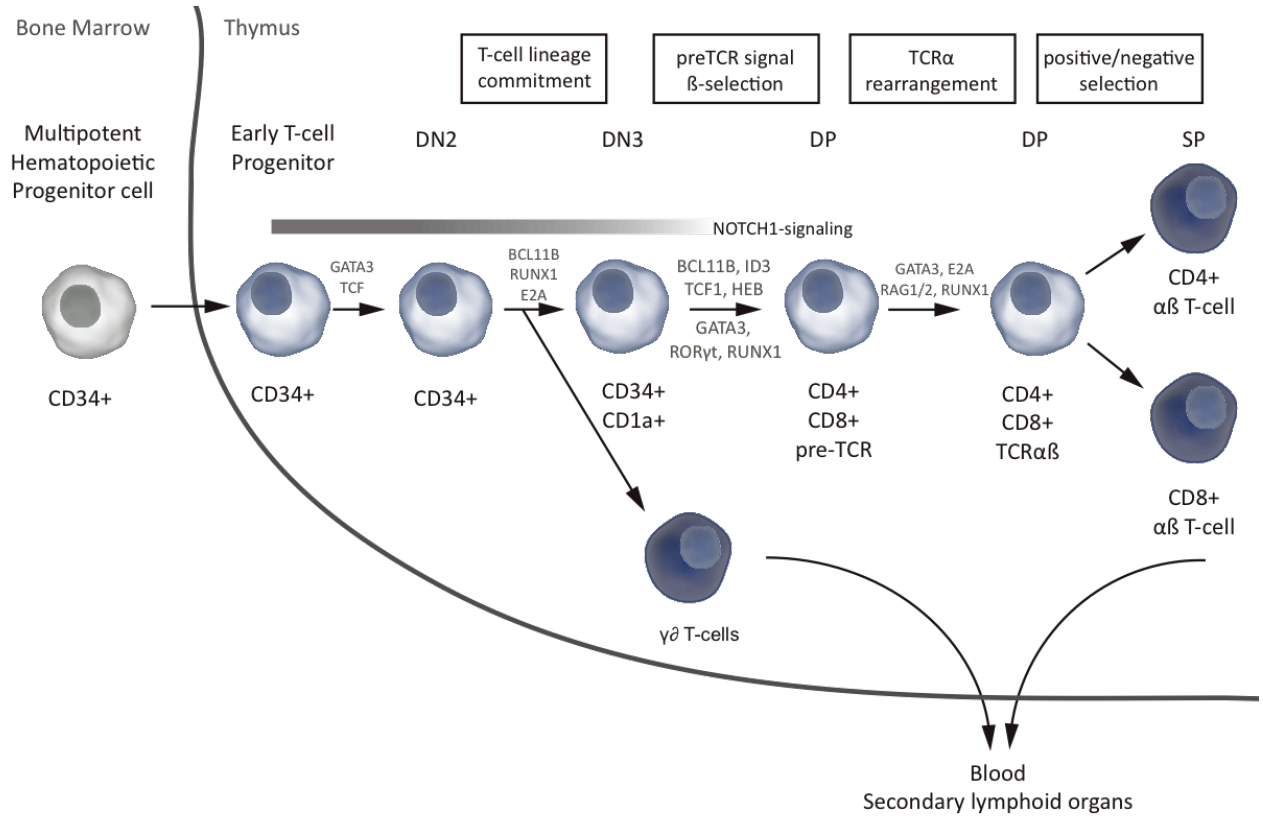


Figure 1. Schedule of normal T-cell development in the thymus. $CD34^{+}$ early T-cell progenitors migrate from the bone marrow to the thymus. In the thymus, several environmental cues and transcriptional regulators will guide the T-cell through the different stages of T-cell development. Mature $CD4^{+}$ αβ, $CD8^{+}$ αβ and γδ T-cells will leave the thymus towards the blood stream and secondary lymphoid organs. (Based on figures from [12, 13], illustrations from www.somersault1824.be)

After T-cell commitment, the next step in T-cell development is called β–selection which depends upon successful V(D)J recombination of the T-cell receptor genes *TCRγ* and *TCRδ* or *TCRβ*. These rearrangements are orchestrated by the 'Recombination activated gene' (RAG) recombinases and the 'Interleukin 7 receptor' (IL7R) and result in the joining of several DNA segments through recombination, resulting in unique T-cell receptor loci. This eventually leads to a great diversity of antigens that can be detected by T-cells. The RAG mediated double strand breaks place the cells undergoing V(D)J recombination in a vulnerable position, exposing them to unwanted DNA damage and illegitimate recombination events. To avoid this, recombining cells go into G1 arrest until correctly assembled premature T-cell receptors (pre-TCRs) are present at the membranes. The pre-TCR consists of a mature TCRβ-chain and a pre-Tα chain, as *TCRα* rearrangements only happen in later stages of T-cell development. This pre-TCRα chain is encoded by the NOTCH1-dependent 'Pre T-cell Antigen Receptor α' (*PTCRA*) gene and signaling through this pre-TCR receptor will shift the cells into rapid proliferation[6]. Furthermore, this will lead to the expression of CD4 and CD8 membrane receptors, bringing T-cell development in the 'double positive' stage. In addition, under the impulse of a strong NOTCH1 signal, a small subset of immature thymocytes will undergo rearrangements of the *TCRγ* and *TCRδ* loci producing *γδ* T-cells with functions in the stress response in epithelial cells[6, 10].

As indicated, the double positive αβ T-cells will undergo TCRα gene rearrangement. Given that cells underwent a proliferative burst immediately after β-selection, thymocytes are produced with the same β-chain, but with a large diversity of α-chains, again leading to an increase in antigen detection. A mature TCRαβ receptor and the markers CD3, CD4 and CD8 mark these thymocytes, activating positive and negative selection of the T-cells. During this stage, the interaction of the T-cell receptor with an antigen in the context of the major histocompatibility complex (MHC) will be evaluated to eliminate autoreactive T-cells and T-cells without a functional TCR. Under impulse of the downregulation of *E2A* and *HEB*, this will eventually lead to mature single positive ($CD4^+$ or $CD8^+$) TCRαβ T-cells that will leave the thymus and will perform a plethora of functions in the bloodstream and peripheral immune organs (see above)[11].

As can be appreciated, thymic T-cell development is a tightly regulated process in which several transcription factors play a key role[12-14]. Another remarkable event during T-cell development is the rearrangement of the T-cell receptor loci, which makes this process prone to DNA damage. It is therefore not surprising that this process can go astray, leading ultimately to malignant transformation of immature T-cells towards full blown T-cell acute lymphoblastic leukemia.

1.2 T-ALL oncogenic driver genes and T-ALL subtypes

T-ALL occurs from a multistep cancer process in which a sequence of multiple consecutive oncogenic events leads to clonal expansion of developmentally arrested malignant T-cells. T-ALL patients can be subdivided into several genetic subtypes based on their transcriptional profiles and presence of mutually exclusive driver oncogenes[15-18].

These genomic pictures of heterogeneous T-ALLs emerged from several studies using transcriptome profiling. Ferrando et al. was able to couple the T-ALL oncogenes 'T-cell leukemia Home-

obox 1' (*TLX1*/*HOX11*), 'T-cell Acute Lymphocytic leukemia 1' (*TAL1*) and 'LYL1 basic helix-loop-helix family member' (*LYL1*) to specific gene expression profiles by microarray expression profiling. Next to that, hierarchical clustering of expression profiles allowed the identification of a new subgroup that clustered close to the *TLX1* group and was marked by high expression levels of 'T-cell leukemia Homeobox 3' (*TLX3*/*HOX11L2*)[15]. A second important study by the Soulier team detected a subset of patients with an unknown disease mechanism, which clustered in a separate group, but near the *TLX1* and *TLX3* samples. All patient samples in this cluster had markedly elevated expression levels of *HOXA* genes, resulting from a recurrent cryptic chromosomal rearrangement[16]. In addition, T-ALL samples with an immature immunophenotype (strong CD34 expression) also marked a separate cluster.

These two papers thus detected new oncogenic T-ALL transcription factors by means of gene expression profiling and they discovered the subtype specific gene expression pattern in the different T-ALL subtypes. In the next paragraphs, these subtypes will be discussed in more detail and an overview can be found in Figure 2 and Table 1.

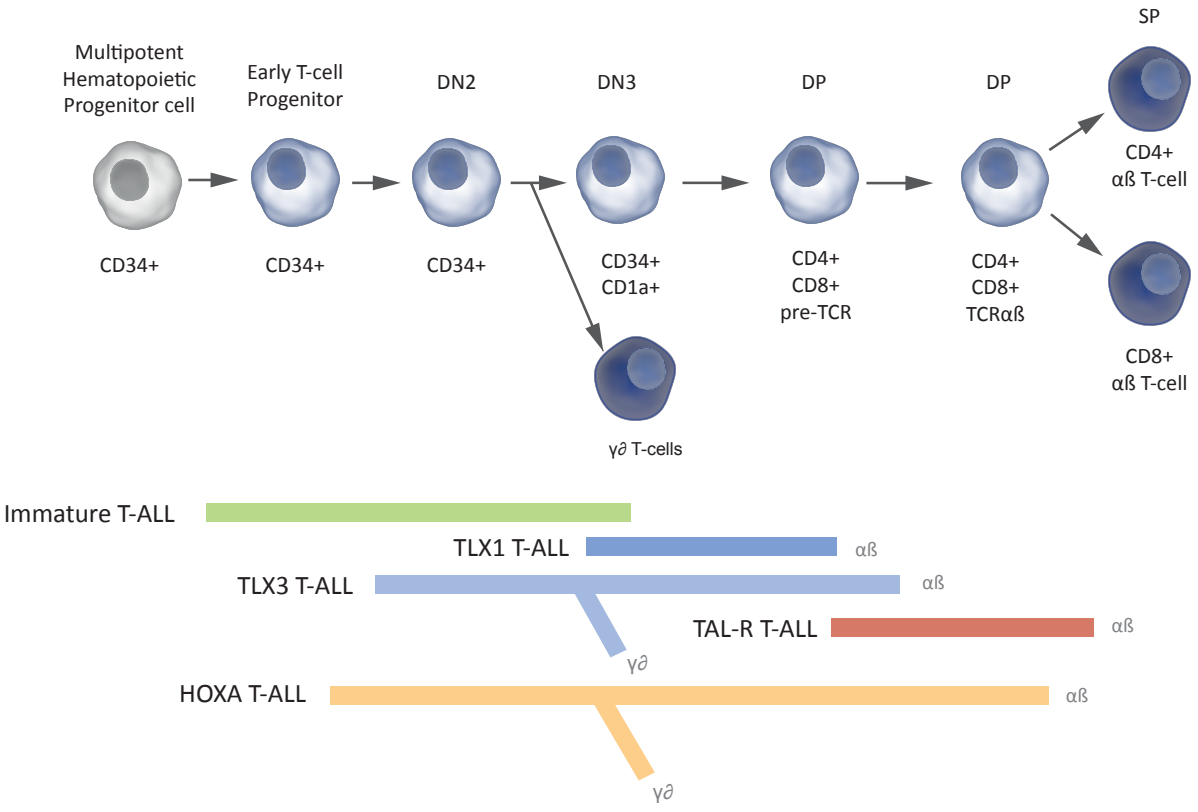


Figure 2. Schematic overview of T-ALL genetic subtypes in relation to their T-cell developmental stage. T-ALL occurs from malignant development of immature T-cells. The genetic subtypes of T-ALL can be linked to a specific stage of differentiation arrest. (Based on a figure from [18], illustrations from www.somersault1824.be)

1.2.1 Immature T-ALL

The immature T-ALL subtype is marked by a differentiation arrest during the very early, double negative, stages of T-cell development. Remarkably, thus far no recurrent specific driver oncogene could be identified. Immature T-ALLs exhibit high expression of *CD34*, *LMO2, MEF2C* and *LYL1,* together with myeloid markers CD13 and CD33[15, 16, 19]. In comparison to other T-ALL subtypes, immature T-ALL patients appear to have fewer deletions of the *CDKN2A* and *CDKN2B* locus and less *NOTCH1* activating mutations[20]. The prevalence of immature T-ALL is higher in adult patients compared to pediatric cases and, remarkably, some of the oncogenic mutations involved in immature T-ALL development (ex. *DNMT3A*) are only present in adult T-ALL samples[21].

A part of the immature T-ALL cases, with a differentiation arrest at the ETP-stage of T-cell development, can be further subclassified into early T-cell precursor ALL (ETP-ALL). ETP-ALL is characterized by the absence of expression of CD1a and CD8, weak expression of CD5 and the expression of one or more stem cell or myeloid markers (CD117, CD34, CD13, CD33, CD11b, CD65 and/or HLA-DR), resembling a bi-phenotypic nature[22]. Using whole genome sequencing, Zhang et al. established a comprehensive genomic picture of the mutational landscape of ETP-ALL characterized by inactivating mutations in genes involved in hematopoietic development (*RUNX1*, *GATA3*, *BCL11B*, *ETV6*…), activating mutations of mediators of cytokine receptor signaling and RAS signaling (*NRAS*, *KRAS*, *IL7R*, *JAK1* …) and various genetic lesions affecting epigenetic regulators, most notably the 'Polycomb Repressive Complex 2'(PRC2) components *EZH2*, *SUZ12* and *EED*[23]. Of further interest, some adult ETP-ALL patients also show high levels of *DNMT3A*, *IDH1* and *IDH2*, which are known oncogenes in acute myeloid leukemia (AML) in keeping with the bi-phenotypic nature of ETP-ALL[24].

In addition to the early studies describing high expression levels for presumed oncogenes in immature T-ALL, more recently yet another driver gene was added to the list. Unexpectedly, Goossens et al. observed T-cell malignancies with immature characteristics in a mouse model with Zeb2 gain-of-function in a p53 knock out genetic background. Together with the finding of rare translocations juxtaposing *ZEB2* in the vicinity of the *BCL11B* locus, leading to *ZEB2* overexpression and *BCL11B* downregulation, this provided strong support for an oncogenic role of *ZEB2* in immature T-ALL development. Further functional studies illustrated that ZEB2 caused increased IL7R/JAK/STAT signaling driving thymocyte proliferation and survival[25].

1.2.2 TLX1 and TLX3 driven T-ALL

Five to ten percent of pediatric and 30% of adult T-ALL cases has an ectopic expression of the *TLX1* oncogene. *TLX1* (*HOX11*) is a homeobox gene involved in the development of the spleen during embryogenesis and is not expressed in thymocytes. Translocations of the *TLX1* gene to the regulatory regions of the T-cell receptor genes TCRα/TCRδ or TCRβ have been described (t(7;10)(q34;q24) t(10;14)(q24;q11)), leading to the overexpression of *TLX1*[26]. The lymphoblasts of these patients showed an arrest at the early cortical $CD1a^+$ stage during T-cell development, prior to the $CD4^+CD8^+$ double positive stage and were all committed to the αβ–lineage[15, 27].

The oncogenic role of TLX1 in T-ALL was studied in more detail by De Keersmaecker et al., using a *TLX1* overexpressing mouse model. After a long latency, these mice developed T-cell neoplasms with a high degree of aneuploidy and defects in the mitotic checkpoints. Furthermore, it was shown that *TLX1* overexpression caused inhibition of T-cell differentiation and required additional genetic events for further T-ALL oncogenesis such as *Notch1* activating and *Pten*, *Tp53* and *Bcl11b* inactivating mutations[28]. A remarkable finding was done by Durinck et al. who studied the genome-wide binding of TLX1 in T-ALL and observed an unexpected attenuation of NOTCH1-signaling through TLX1 binding at *NOTCH1* regulatory sequences as well as several NOTCH1 targets, including *NOTCH3*[29]. This could explain the high frequency of *NOTCH1* activating mutations in TLX1 T-ALL as NOTCH1 signaling is critical for cell survival.

Another homeobox gene, *TLX3* (*HOX11L2*), is ectopically expressed in approximately 20-25% of pediatric T-ALL patients and 5% of adult T-ALL cases, defining yet another T-ALL subtype[30, 31]. The overexpression of *TLX3* is mostly caused by a cryptic t(5;14)(q35;q32) translocation juxtaposing *TLX3* nearby the *BCL11B* locus[31]. Furthermore, also translocations of *TLX3* to TCRα/TCRδ (t(5;14)(q32;q11)) and *CDK6* (t(5;7)(q35;q21)) have been described in T-ALL patients[32, 33]. The *TLX3* patients show a broader range of differentiation arrest as immature and *TLX1* T-ALL. Several cases have a more immature immunophenotype as compared to *TLX1* T-ALL, but others are then linked to more mature αβ or γδ lineages[27, 34, 35].

An elegant study discovered that *TLX1* and *TLX3* positive T-ALLs showed a low frequency of TCRα rearrangements and high repressive H3K27me3 histone marks at the TCRα locus. This led to the discovery that TLX1 and TLX3 interact with ETS1, resulting in the epigenetic silencing of the TCRα locus and the specific tumor differentiation stage[36].

1.2.3 TAL-rearranged T-ALL

The TAL-rearranged (TAL-R) subtype represents T-ALLs arising through overexpression of E2A/HEB transcription factors due to unscheduled constitutive *TAL1*, *LYL1* or *LMO2* activation during thymopoiesis. *TAL1* was first discovered through the rare translocations t(1;14)(p32;q11) and t(1;7) (p32;q35))[37], leading to the juxtaposition of *TAL1* near strong enhancers of the TCRα/TCRδ or TCRβ loci. Subsequently, a highly recurrent cryptic interstitial 1p32 deletion was detected in 25% of T-ALL cases, placing *TAL1* under the control of the *STIL* promotor[37]. The class B helix-loop-helix (bHLH) protein TAL1 binds LIM-domain only proteins (LMO1 or LMO2) and overexpression of these *LMO* genes themselves is a second mechanism linked to this T-ALL subtype. Typically, *LMO1/2* activation is detected in 10% of T-ALL cases and results from TCR mediated illegitimate recombination events producing various translocations (t(11;14)(p13;q11), t(7;11)(q35;p13), t(11;14)(p15;q11) or t(7;11) (q35;p15))[38, 39]. In addition, interstitial 11p13 deletions have also been detected, removing a negative regulatory element of *LMO2*[40]. *TAL-R* T-ALL cases are arrested at a relatively late stage during T-cell development and therefore present a more mature $CD3^+CD4^+CD8^+$ double positive phenotype from the αβ T-cell lineage[27].

Further studies have unraveled the mode of action of the TAL1 transcription factor. TAL1 was shown to be part of a large complex, called the TAL1-complex, containing also LMO2, RUNX1, E2A and HEB[41-43]. Members of this complex could interact with the histone acetylase p300 lead-

ing to the activation of target genes including *MYB* and *TRIB2* [44, 45]. Furthermore, this complex activates miRNA-223, which inhibits the expression of *FBXW7*, a known tumor suppressor gene in T-ALL that acts through the inhibition of NOTCH1 signaling[46].

1.2.4 HOXA driven T-ALL

The *HOXA* subgroup is typically marked by cases with illegitimate TCR rearrangement driven HOXA gene cluster activation. The prototypical rearrangement is a pericentric inversion of chromosome 7 (inv(7)(p15q34)) which was shown to activate several HOXA cluster genes, most notably *HOXA9*, *HOXA10* and *HOXA11*, and presents with a mature immunophenotype of either the αβ- or the γδ-lineage[16, 47]. In addition to HOXA cluster rearranged cases, HOXA gene up-regulation could also result from other oncogenic events through e.g. formation of the CALM-AF10 (t(10;11)(p13;q14) or the SET-NUP214 (del(9q34)) fusion protein[48, 49]. The CALM-AF10 protein interacts with DOT1L, an H3K79 methyltransferase that is required for *HOXA* activation, while SET-NUP214 was shown to act as a transcriptional cofactor for *HOXA* genes[49, 50]. Furthermore, a small subset of T-ALL cases that cluster in this subgroup carries *MLL*-rearrangements[51, 52]. The MLL-ENL and MLL-AF10 fusion proteins also interact with DOT1L leading to the activation of the *HOXA* genes[53]. These indirect *HOXA* activated cases have, in comparison to HOXA cluster rearranged cases, an immature or a γδ immunophenotype.

1.2.5 A quest to link T-ALL subtypes to prognosis

In the last decade, several studies have been published trying to link T-ALL patients from a specific subgroup to a good or bad prognosis. There was first of all a specific focus on the ETP-ALL cases as they were linked to a very poor prognosis and a high risk of relapse in multiple studies[19, 21, 22 56]. However, in recent years, these findings couldn't be validated, as in new clinical studies there was a similar outcome for both ETP-ALL and non ETP-ALL patients[57-59]. One of the possible reasons could be the intensified treatment protocols in these recent studies, which already led to a very good survival rate for all T-ALL patients. The TLX1 patients were linked to a favorable outcome, as patients with $CD1a^+$ lymphoblasts (most of *TLX1* T-ALL patients) had a good survival prognosis due to the low expression of anti-apoptotic genes during the early cortical stages of T-cell development[15, 54]. On the other hand, TLX3 patients were first linked to a poor outcome, but also this could not be confirmed by others[15, 27, 30, 60-62]. The difficulty to identify specific patients with a bad prognosis in the last studies might be due to the increase in overall survival of T-ALL cases due to intensified treatment protocols. However, the prognosis for relapse patients in T-ALL remains extremely poor, it might be crucial to identify the specific oncogenic mechanism that leads to these relapses.

Table 1. Summary of the most frequent oncogenic driver events for T-ALL subtypes.

	Mechanism of activation	Frequency	Ref
Immature T-ALL			
No specific oncogenic driver event			
TLX1/3 T-ALL			
TLX1	Translocations to TCRα/TCRδ or TCRβ	5-10 % of pediatric T-ALL 30 % of adult T-ALL	15, 26, 54
TLX3	Translocation to BCL11B locus	20-25 % of pediatric T-ALL 5 % of adult T-ALL	31-33
TAL-rearranged T-ALL			
TAL1	Interstitial deletion placing TAL1 under the control of the STIL promoter	25 %	37
LMO2	Translocations to TCRα/TCRδ or TCRβ Deletion to remove negative regulatory element	6 % 3 %	38, 39
HOXA T-ALL			
HOXA	inv(7)(p15q34)	3 %	16, 47
CALM-AF10	t(10;11)(p13;q14) leads to a novel fusion protein	5-10 %	48, 55
SET-NUP214	del(9q34) leads to a novel fusion protein	3 %	49

1.3 Cooperative genetic defects in T-ALL formation

As indicated above, T-ALL oncogenesis is a multistep process. In addition to the inappropriate activation of the above-mentioned proto-oncogenes as initiating event, further genetic alterations are required to breach other tumor suppressing signaling pathways in order for the cells to become fully malignant. Decades of research, including most notably, recent whole genome sequencing efforts, have provided a detailed view on the plethora of oncogenes and tumor suppressors implicated in this complex process[63]. Of these, one of the most frequently perturbed pathways is controlled by NOTCH1, known to play multiple crucial roles in controlling normal T-cell development[10, 64-70] with overactivation of NOTCH1-signaling being observed in over half of T-ALL patients[71]. In addition, deletions of the cell cycle regulators *CDKN2A* (*p16/INK4A*) or *CDKN2B* (*p14/ARF*) are also highly recurrent genetic defects detected in up to 70% of the cases[72]. Like in most other cancer entities, sequencing studies have revealed loss of function mutations in epigenetic regulators including *PHF6*, *UTX*, *SUZ12* and *EZH2*[23, 73, 74]. Further important oncogenic events are the inactivation of *WT1*[75] and *PTEN*[76, 77] or the overexpression of *MYB*[78, 79] and *IL7R*[80, 81]. An overview of the most frequent oncogenic events, next to the T-ALL subtype driving events, can be found in Table 2.

Table 2. Cooperative genetic defects in T-ALL. The most frequent oncogenic events detected in T-ALL irrespective of the T-ALL subtype driver events. (Based on[63])

	Oncogenic lesion	Frequency	Ref
Oncogenes			
NOTCH1	Activating mutation	> 60 %	[71]
IL7R	Activating mutation	10 %	[80, 81]
JAK1/3	Activating mutation	10-25 %	[23, 82]
MYB	t(6;7)(q23;q24) Duplication	3 % 8 %	[78, 79]
Tumorsuppresor genes			
CDKN2A/2B	9p21 deletion	> 70 %	[72]
PHF6	Inactivating mutation or deletion	20-40 %	[74]
FBXW7	Inactivating mutation	10-30 %	[83, 84]
RUNX1	Inactivating mutation or deletion	10-20%	[85, 86]
EZH2	Inactivating mutation or deletion	10-15%	[23]
PTEN	10q23 deletion Inactivating mutation	10-15 % 10-15 %	[76, 77]
BCL11B	Inactivating mutation or deletion	10 %	[87]
SUZ12	Inactivating mutation or deletion	10 %	[23]
WT1	Inactivating mutation or deletion	10 %	[75]

The NOTCH1 oncogene

The NOTCH1-signaling pathway is highly conserved across species, transmitting signals from the cellular environment directly into changes in gene expression. NOTCH1 is a heterodimeric transmembrane receptor that is activated through Serrate-like (JAG1 or JAG2) or Delta-like (DLL1, DLL4) ligands. Upon interaction of the receptor with one of its ligands, a conformational change occurs, leading to the proteolytic cleavage of the receptor by an ADAM metalloprotease and a γ-secretase complex. This cleavage releases the intracellular part of the receptor (ICN1), containing a nuclear localization signal (NLS). In the nucleus, ICN1 functions as a transcription factor in complex with RBPJκ, aiding in the DNA binding, and with the co-activating ligand of the Mastermind-like family (MAML)[88]. This complex activates a plethora of genes, including *c-MYC*, *HES1*, *DTX1* and *IL7R*, through the recruitment of chromatin remodeling complexes and histone acetyltransferases (*CBP*/*p300*)[89]. FBXW7, an E3-ubiquitin ligase, terminates NOTCH1-signaling due to the recognition of the PEST-domain in ICN1, which leads to its proteasomal degradation[84]. In the absence of ICN1, RBPJκ associates with co-repressors that recruit histone deacetylases (HDACs), repressing the activation of NOTCH1 target genes[90]. (Figure 3)

Notch signaling is involved in development and cellular processes in several tissues, including epithelial, endothelial, neuronal, bone, blood and muscle cells[91] and as already mentioned,

NOTCH1-signaling is an important regulator of T-cell development. Strong NOTCH1 signaling is first required for T-cell lineage specification, while subsequent signaling needs to be downregulated to inhibit natural killer cell formation[8]. After T-cell commitment, NOTCH1 is again necessary in the steps leading to the β-selection checkpoint and the γδ T-cell development.

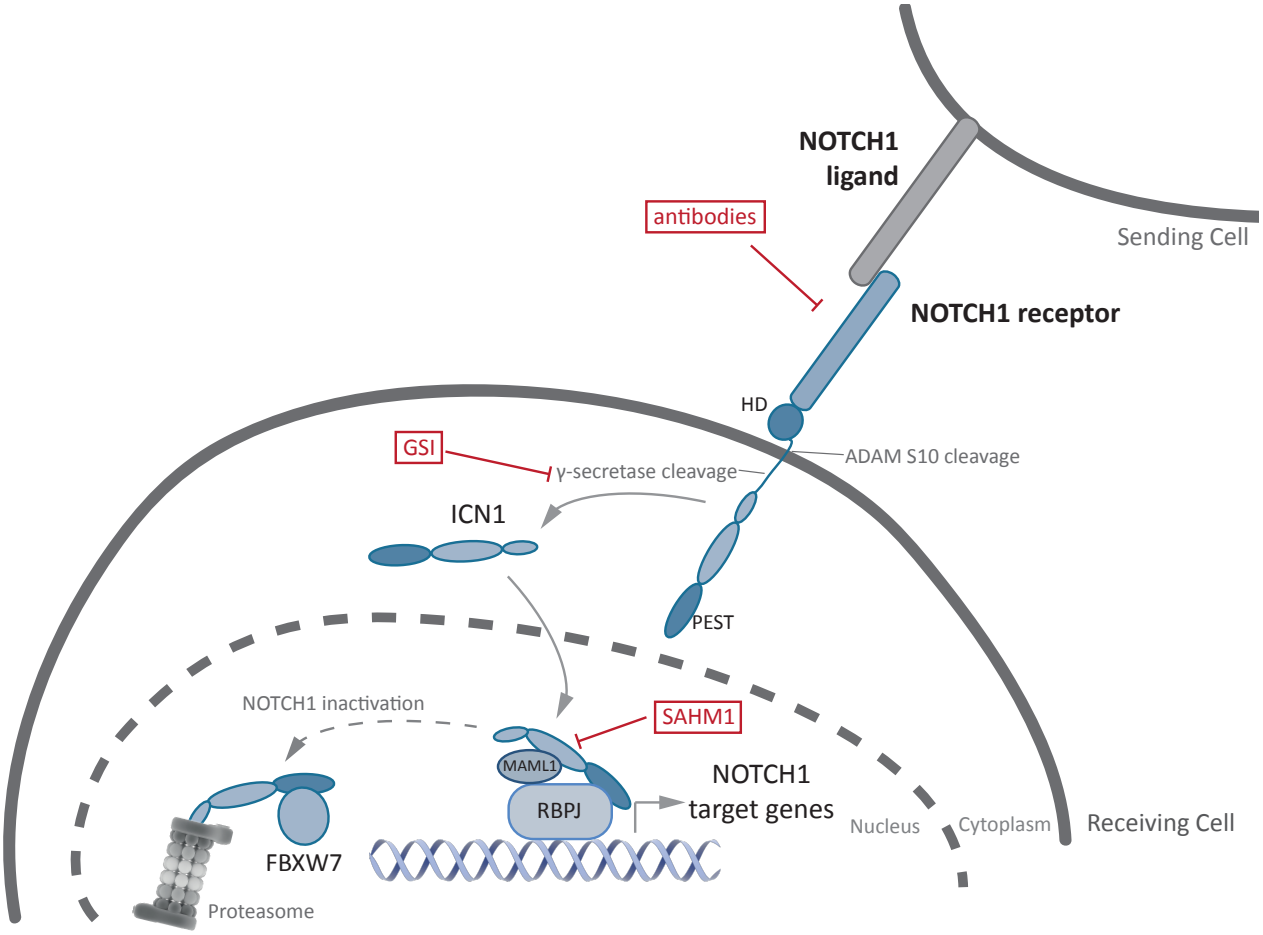


Figure 3. NOTCH1-signaling pathway. The interaction of the NOTCH1 receptor on the surface of the early T-cell progenitors (Receiving Cell) with the DLL4 ligand expressed on the surface of thymic stroma cells (Sending Cell) results in proteolytic cleavage of the receptor. The intracellular domain of NOTCH1 (ICN1) is released from the membrane and translocates to the nucleus, where it interacts with RBPJ and MAML1. This complex activates the expression of NOTCH1 target genes, leading to cell differentiation, T-cell lineage commitment, proliferation and survival. The termination of NOTCH1 signaling is mediated by FBXW7. The NOTCH1 signaling pathway can be therapeutically targeted through GSI (γ-secretase inhibitor) based blocking of the γ-secretase cleavage, antibodies blocking the NOTCH1-receptor or the small peptide SAHM1 inhibiting the ICN1-transcriptional complex. (Based on[63], illustration from www.somersault1824.be)

In over half of the T-ALL patients, an overactivation of the NOTCH1-signaling pathway is driving the immature T-cells to an oncogenic proliferation. The role of NOTCH1 in T-ALL was first described by a rare t(7;9)(q34;q34.3) translocation, leading to the constitutive activation of *NOTCH1*[92]. Much later, the high frequency of *NOTCH1* mutations was detected through the discovery of several hot spot activating mutations. First of all, there are the missense mutations or small indels in the heterodimerization domain (HD). Secondly, nonsense mutations, insertion or deletions can lead to a partial or complete removal of the PEST domain, which inhibits the degradation of ICN1 by FBXW7 in the nucleus. Several patients were also identified with a combi-

nation of mutations in both the HD and the PEST domain[71]. Inactivating mutations or deletions of *FBXW7* also increase the NOTCH1-signalling and frequently occur in T-ALL. These mutations interfere with the interaction of FBXW7 and ICN1, decreasing ICN1 degradation, but additionally also blocking degradation of c-MYC, one of the major NOTCH1 downstream targets[83, 84]. Also several translocations leading to the overexpression of *c-MYC* have been detected in T-ALL patients[93, 94] and recently the focal duplication of a *c-MYC* enhancer was proven oncogenic in 5 % of T-ALL cases[95] (see Box 1).

The effect of oncogenic NOTCH1 signaling can be linked to several direct NOTCH1 target genes involved in anabolic pathways and cell growth, with *MYC*, *IL7R*, *PTCRA* and *IGF1R* as major effectors[96-99]. Activated NOTCH1 also affects cell growth by interacting with the PI3K/AKT/mTOR pathway[100]. Next to the effects on cell growth, NOTCH1 promotes G1/S cell cycle progression through the upregulation of *CDK4* and *CDK6*[101] and the downregulation of *p27*/*KIP* and *p18*/*INK4C*[102]. By activating NF-κB, NOTCH1 might also be involved in increased cell survival[103]. Furthermore, NOTCH1 can control several epigenetic modulators of gene expression[104].

Box 1. A NOTCH1-controlled long range *MYC* enhancer[95].

In 2014, the Ferrando team detected a focal duplication located 1.4 Mb downstream of the MYC oncogene in 5 % of T-ALL cases. Interestingly, this region, which is devoid of protein coding genes, appeared to be bound by the intracellular NOTCH1 complex in T-ALL cell lines. Further analysis of regulatory factors and epigenetic histone marks, revealed the active enhancer features of this region (ex. P300 occupancy, high H3K27ac and high H3K4me1/H3K4me3 ratio). Next to that, 3C-analysis (Chromosome Conformation Capture) confirmed the association of the *MYC* promoter and this enhancer and reporter assays revealed an enhancer dose dependent activation of the promoter. This enhancer was named the *NOTCH*-bound *MYC* enhancer (N-Me).

In vivo N-Me knockout (N-Me$^{-/-}$) mice were viable, but with a reduced thymus size and cellularity. In the thymus, there appeared an accumulation of double negative T-cells and a reduction in double positive and CD4^{+} and CD8^{+} single positive cells. Furthermore, heterozygous N-Me in a NOTCH1-induced tumor model could delay tumor development, whereas N-Me knockout could fully inhibit tumor initiation. This shows a strong dose dependent effect of the N-Me enhancer on the NOTCH1-leukemia initiation, with the N-Me regulation of *MYC* expression as a major cause.

With this study, the oncogenic effect of a focal duplication in a region devoid of protein coding genes was proven. This was the first enhancer that showed to have major impact on T-ALL development and marks the importance of non-coding regions in gene regulation.

All of the above clearly indicates that NOTCH1 is an important effector in T-ALL oncogenesis in over half of T-ALL patients thus putting NOTCH1 as a major target for precision drugging. Gamma-secretase inhibitors (GSIs) were the first compounds of choice, leading to a decrease in in-

tracellular NOTCH1[105]. The first clinical trials of these GSIs showed at first an anti-leukemic effect, but soon resistance to this compound was encountered[106]. Another disadvantage of this treatment was the dose-limiting gastrointestinal toxicity, resulting from the inhibition of NOTCH1 and NOTCH2 in the intestinal progenitor cells[107]. Various drug combinations with GSIs were tested to look for drugs that could alleviate the side effects and glucocorticoids were identified as a candidate. Combination regimens with chemotherapy or drugs targeting other pathways involved in T-ALL (ex. NF-κB signaling[103], PI3K-AKT-mTOR inhibitors[108] and HDAC inhibitors[109]) could also be used. Other therapeutic options would be the use of an antibody that specifically blocks the NOTCH1 receptor[110], the inhibition of the ADAM metalloprotease[111] or the inhibition of the ICN1-RBPJκ-MAML1 transcriptional complex by a peptide called SAHM1[112]. Further research to the downstream effectors of NOTCH1 oncogenesis might lead to the discovery of new drugging targets.

1.4 References

1. Karrman K, Johansson B. Pediatric T-Cell Acute Lymphoblastic Leukemia. Genes Chromosomes Cancer. 2016.
2. Degos L. John Hughes Bennett, Rudolph Virchow... and Alfred Donne: the first description of leukemia. Hematol J. 2001;2(1):1.
3. Pui CH, Robison LL, Look AT. Acute lymphoblastic leukaemia. Lancet. 2008;371(9617):1030-43.
4. Sutton R, Shaw PJ, Venn NC, Law T, Dissanayake A, Kilo T, et al. Persistent MRD before and after allogeneic BMT predicts relapse in children with acute lymphoblastic leukaemia. Br J Haematol. 2015;168(3):395-404.
5. Hao QL, George AA, Zhu J, Barsky L, Zielinska E, Wang X, et al. Human intrathymic lineage commitment is marked by differential CD7 expression: identification of CD7- lympho-myeloid thymic progenitors. Blood. 2008;111(3):1318-26.
6. Van de Walle I, De Smet G, De Smedt M, Vandekerckhove B, Leclercq G, Plum J, et al. An early decrease in Notch activation is required for human TCR-alphabeta lineage differentiation at the expense of TCR-gammadelta T cells. Blood. 2009;113(13):2988-98.
7. Weerkamp F, Luis TC, Naber BA, Koster EE, Jeannotte L, van Dongen JJ, et al. Identification of Notch target genes in uncommitted T-cell progenitors: No direct induction of a T-cell specific gene program. Leukemia. 2006;20(11):1967-77.
8. Van de Walle I, Dolens AC, Durinck K, De Mulder K, Van Loocke W, Damle S, et al. GATA3 induces human T-cell commitment by restraining Notch activity and repressing NK-cell fate. Nat Commun. 2016;7:11171.
9. Blom B, Spits H. Development of human lymphoid cells. Annual review of immunology. 2006;24:287-320.
10. Garcia-Peydro M, de Yebenes VG, Toribio ML. Sustained Notch1 signaling instructs the earliest human intrathymic precursors to adopt a gammadelta T-cell fate in fetal thymus organ culture. Blood. 2003;102(7):2444-51.
11. Jones ME, Zhuang Y. Acquisition of a functional T cell receptor during T lymphocyte development is enforced by HEB and E2A transcription factors. Immunity. 2007;27(6):860-70.
12. Yui MA, Rothenberg EV. Developmental gene networks: a triathlon on the course to T cell identity. Nature reviews Immunology. 2014;14(8):529-45.
13. Rothenberg EV, Ungerback J, Champhekar A. Forging T-Lymphocyte Identity: Intersecting Networks of Transcriptional Control. Adv Immunol. 2016;129:109-74.
14. Rothenberg EV, Kueh HY, Yui MA, Zhang JA. Hematopoiesis and T-cell specification as a model developmental system. Immunol Rev. 2016;271(1):72-97.
15. Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC, et al. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer Cell. 2002;1(1):75-87.
16. Soulier J, Clappier E, Cayuela JM, Regnault A, Garcia-Peydro M, Dombret H, et al. HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). Blood. 2005;106(1):274-86.
17. Van Vlierberghe P, Pieters R, Beverloo HB, Meijerink JP. Molecular-genetic insights in paediatric T-cell acute lymphoblastic leukaemia. Br J Haematol. 2008;143(2):153-68.
18. Meijerink JP. Genetic rearrangements in relation to immunophenotype and outcome in T-cell acute lymphoblastic leukaemia. Best Pract Res Clin Haematol. 2010;23(3):307-18.
19. Homminga I, Pieters R, Langerak AW, de Rooi JJ, Stubbs A, Verstegen M, et al. Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. Cancer Cell. 2011;19(4):484-97.
20. Volanakis EJ, Williams RT, Sherr CJ. Stage-specific Arf tumor suppression in Notch1-induced T-cell acute lymphoblastic leukemia. Blood. 2009;114(20):4451-9.
21. Van Vlierberghe P, Ambesi-Impiombato A, Perez-Garcia A, Haydu JE, Rigo I, Hadler M, et al. ETV6 mutations in early immature human T cell leukemias. J Exp Med. 2011;208(13):2571-9.
22. Coustan-Smith E, Mullighan CG, Onciu M, Behm FG, Raimondi SC, Pei D, et al. Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. The Lancet Oncology. 2009;10(2):147-56.
23. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012;481(7380):157-63.
24. Van Vlierberghe P, Ambesi-Impiombato A, De Keersmaecker K, Hadler M, Paietta E, Tallman MS, et al. Prognostic relevance of integrated genetic profiling in adult T-cell acute lymphoblastic leukemia. Blood. 2013;122(1):74-82.
25. Goossens S, Radaelli E, Blanchet O, Durinck K, Van der Meulen J, Peirs S, et al. ZEB2 drives immature T-cell lymphoblastic leukaemia development via enhanced tumour-initiating potential and IL-7 receptor signalling. Nat Commun. 2015;6:5794.

26. Hatano M, Roberts CW, Minden M, Crist WM, Korsmeyer SJ. Deregulation of a homeobox gene, HOX11, by the t(10;14) in T cell leukemia. Science. 1991;253(5015):79-82.
27. van Grotel M, Meijerink JP, Beverloo HB, Langerak AW, Buys-Gladdines JG, Schneider P, et al. The outcome of molecular-cytogenetic subgroups in pediatric T-cell acute lymphoblastic leukemia: a retrospective study of patients treated according to DCOG or COALL protocols. Haematologica. 2006;91(9):1212-21.
28. De Keersmaecker K, Real PJ, Gatta GD, Palomero T, Sulis ML, Tosello V, et al. The TLX1 oncogene drives aneuploidy in T cell transformation. Nat Med. 2010;16(11):1321-7.
29. Durinck K, Van Loocke W, Van der Meulen J, Van de Walle I, Ongenaert M, Rondou P, et al. Characterization of the genome-wide TLX1 binding profile in T-cell acute lymphoblastic leukemia. Leukemia. 2015;29(12):2317-27.
30. Mauvieux L, Leymarie V, Helias C, Perrusson N, Falkenrodt A, Lioure B, et al. High incidence of Hox11L2 expression in children with T-ALL. Leukemia. 2002;16(12):2417-22.
31. Bernard OA, Busson-LeConiat M, Ballerini P, Mauchauffe M, Della Valle V, Monni R, et al. A new recurrent and specific cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. Leukemia. 2001;15(10):1495-504.
32. Su XY, Busson M, Della Valle V, Ballerini P, Dastugue N, Talmant P, et al. Various types of rearrangements target TLX3 locus in T-cell acute lymphoblastic leukemia. Genes Chromosomes Cancer. 2004;41(3):243-9.
33. Hansen-Hagge TE, Schafer M, Kiyoi H, Morris SW, Whitlock JA, Koch P, et al. Disruption of the RanBP17/Hox11L2 region by recombination with the TCRdelta locus in acute lymphoblastic leukemias with t(5;14)(q34;q11). Leukemia. 2002;16(11):2205-12.
34. van Grotel M, Meijerink JP, van Wering ER, Langerak AW, Beverloo HB, Buijs-Gladdines JG, et al. Prognostic significance of molecular-cytogenetic abnormalities in pediatric T-ALL is not explained by immunophenotypic differences. Leukemia. 2008;22(1):124-31.
35. Asnafi V, Beldjord K, Garand R, Millien C, Bahloul M, LeTutour P, et al. IgH DJ rearrangements within T-ALL correlate with cCD79a expression, an immature/TCRgammadelta phenotype and absence of IL7Ralpha/CD127 expression. Leukemia. 2004;18(12):1997-2001.
36. Dadi S, Le Noir S, Payet-Bornet D, Lhermitte L, Zacarias-Cabeza J, Bergeron J, et al. TLX homeodomain oncogenes mediate T cell maturation arrest in T-ALL via interaction with ETS1 and suppression of TCRalpha gene expression. Cancer Cell. 2012;21(4):563-76.
37. Baer R. TAL1, TAL2 and LYL1: a family of basic helix-loop-helix proteins implicated in T cell acute leukaemia. Semin Cancer Biol. 1993;4(6):341-7.
38. Sanchez-Garcia I, Rabbitts TH. LIM domain proteins in leukaemia and development. Semin Cancer Biol. 1993;4(6):349-58.
39. Royer-Pokora B, Loos U, Ludwig WD. TTG-2, a new gene encoding a cysteine-rich protein with the LIM motif, is overexpressed in acute T-cell leukaemia with the t(11;14)(p13;q11). Oncogene. 1991;6(10):1887-93.
40. Van Vlierberghe P, van Grotel M, Beverloo HB, Lee C, Helgason T, Buijs-Gladdines J, et al. The cryptic chromosomal deletion del(11)(p12p13) as a new activation mechanism of LMO2 in pediatric T-cell acute lymphoblastic leukemia. Blood. 2006;108(10):3520-9.
41. Hsu HL, Huang L, Tsan JT, Funk W, Wright WE, Hu JS, et al. Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. Mol Cell Biol. 1994;14(2):1256-65.
42. Wadman IA, Osada H, Grutz GG, Agulnick AD, Westphal H, Forster A, et al. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. EMBO J. 1997;16(11):3145-57.
43. Palii CG, Perez-Iratxeta C, Yao Z, Cao Y, Dai F, Davison J, et al. Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. EMBO J. 2011;30(3):494-509.
44. Huang S, Qiu Y, Stein RW, Brandt SJ. p300 functions as a transcriptional coactivator for the TAL1/SCL oncoprotein. Oncogene. 1999;18(35):4958-67.
45. Sanda T, Lawton LN, Barrasa MI, Fan ZP, Kohlhammer H, Gutierrez A, et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. Cancer Cell. 2012;22(2):209-21.
46. Mansour MR, Sanda T, Lawton LN, Li X, Kreslavsky T, Novina CD, et al. The TAL1 complex targets the FBXW7 tumor suppressor by activating miR-223 in human T cell acute lymphoblastic leukemia. J Exp Med. 2013;210(8):1545-57.
47. Speleman F, Cauwelier B, Dastugue N, Cools J, Verhasselt B, Poppe B, et al. A new recurrent inversion, inv(7)(p15q34), leads to transcriptional activation of HOXA10 and HOXA11 in a subset of T-cell acute lymphoblastic leukemias. Leukemia. 2005;19(3):358-66.

48. Bohlander SK, Muschinsky V, Schrader K, Siebert R, Schlegelberger B, Harder L, et al. Molecular analysis of the CALM/AF10 fusion: identical rearrangements in acute myeloid leukemia, acute lymphoblastic leukemia and malignant lymphoma patients. Leukemia. 2000;14(1):93-9.
49. Van Vlierberghe P, van Grotel M, Tchinda J, Lee C, Beverloo HB, van der Spek PJ, et al. The recurrent SET-NUP214 fusion as a new HOXA activation mechanism in pediatric T-cell acute lymphoblastic leukemia. Blood. 2008;111(9):4668-80.
50. Okada Y, Jiang Q, Lemieux M, Jeannotte L, Su L, Zhang Y. Leukaemic transformation by CALM-AF10 involves upregulation of Hoxa5 by hDOT1L. Nat Cell Biol. 2006;8(9):1017-24.
51. Rubnitz JE, Camitta BM, Mahmoud H, Raimondi SC, Carroll AJ, Borowitz MJ, et al. Childhood acute lymphoblastic leukemia with the MLL-ENL fusion and t(11;19)(q23;p13.3) translocation. J Clin Oncol. 1999;17(1):191-6.
52. Ferrando AA, Armstrong SA, Neuberg DS, Sallan SE, Silverman LB, Korsmeyer SJ, et al. Gene expression signatures in MLL-rearranged T-lineage and B-precursor acute leukemias: dominance of HOX dysregulation. Blood. 2003;102(1):262-8.
53. Okada Y, Feng Q, Lin Y, Jiang Q, Li Y, Coffield VM, et al. hDOT1L links histone methylation to leukemogenesis. Cell. 2005;121(2):167-78.
54. Ferrando AA, Neuberg DS, Dodge RK, Paietta E, Larson RA, Wiernik PH, et al. Prognostic importance of TLX1 (HOX11) oncogene expression in adults with T-cell acute lymphoblastic leukaemia. Lancet. 2004;363(9408):535-6.
55. Asnafi V, Radford-Weiss I, Dastugue N, Bayle C, Leboeuf D, Charrin C, et al. CALM-AF10 is a common fusion transcript in T-ALL and is specific to the TCRgammadelta lineage. Blood. 2003;102(3):1000-6.
56. Inukai T, Kiyokawa N, Campana D, Coustan-Smith E, Kikuchi A, Kobayashi M, et al. Clinical significance of early T-cell precursor acute lymphoblastic leukaemia: results of the Tokyo Children's Cancer Study Group Study L99-15. Br J Haematol. 2012;156(3):358-65.
57. Zuurbier L, Gutierrez A, Mullighan CG, Cante-Barrett K, Gevaert AO, de Rooi J, et al. Immature MEF2C-dysregulated T-cell leukemia patients have an early T-cell precursor acute lymphoblastic leukemia gene signature and typically have non-rearranged T-cell receptors. Haematologica. 2014;99(1):94-102.
58. Homminga I, Vuerhard MJ, Langerak AW, Buijs-Gladdines J, Pieters R, Meijerink JP. Characterization of a pediatric T-cell acute lymphoblastic leukemia patient with simultaneous LYL1 and LMO2 rearrangements. Haematologica. 2012;97(2):258-61.
59. Patrick K, Wade R, Goulden N, Mitchell C, Moorman AV, Rowntree C, et al. Outcome for children and young people with Early T-cell precursor acute lymphoblastic leukaemia treated on a contemporary protocol, UKALL 2003. Br J Haematol. 2014;166(3):421-4.
60. Baak U, Gokbuget N, Orawa H, Schwartz S, Hoelzer D, Thiel E, et al. Thymic adult T-cell acute lymphoblastic leukemia stratified in standard- and high-risk group by aberrant HOX11L2 expression: experience of the German multicenter ALL study group. Leukemia. 2008;22(6):1154-60.
61. Ballerini P, Blaise A, Busson-Le Coniat M, Su XY, Zucman-Rossi J, Adam M, et al. HOX11L2 expression defines a clinical subtype of pediatric T-ALL associated with poor prognosis. Blood. 2002;100(3):991-7.
62. Cave H, Suciu S, Preudhomme C, Poppe B, Robert A, Uyttebroeck A, et al. Clinical significance of HOX11L2 expression linked to t(5;14)(q35;q32), of HOX11 expression, and of SIL-TAL fusion in childhood T-cell malignancies: results of EORTC studies 58881 and 58951. Blood. 2004;103(2):442-50.
63. Belver L, Ferrando A. The genetics and mechanisms of T cell acute lymphoblastic leukaemia. Nat Rev Cancer. 2016;16(8):494-507.
64. Radtke F, Wilson A, Stark G, Bauer M, van Meerwijk J, MacDonald HR, et al. Deficient T cell fate specification in mice with an induced inactivation of Notch1. Immunity. 1999;10(5):547-58.
65. Wolfer A, Wilson A, Nemir M, MacDonald HR, Radtke F. Inactivation of Notch1 impairs VDJbeta rearrangement and allows pre-TCR-independent survival of early alpha beta Lineage Thymocytes. Immunity. 2002;16(6):869-79.
66. Kumano K, Chiba S, Kunisato A, Sata M, Saito T, Nakagami-Yamaguchi E, et al. Notch1 but not Notch2 is essential for generating hematopoietic stem cells from endothelial cells. Immunity. 2003;18(5):699-711.
67. Ciofani M, Zuniga-Pflucker JC. Notch promotes survival of pre-T cells at the beta-selection checkpoint by regulating cellular metabolism. Nat Immunol. 2005;6(9):881-8.
68. Taghon T, Yui MA, Pant R, Diamond RA, Rothenberg EV. Developmental and molecular characterization of emerging beta- and gammadelta-selected pre-T cells in the adult mouse thymus. Immunity. 2006;24(1):53-64.
69. Taghon T, Van de Walle I, De Smet G, De Smedt M, Leclercq G, Vandekerckhove B, et al. Notch signaling is required for proliferation but not for differentiation at a well-defined beta-selection checkpoint during human T-cell development. Blood. 2009;113(14):3254-63.

70. Taghon T, Waegemans E, Van de Walle I. Notch signaling during human T cell development. Curr Top Microbiol Immunol. 2012;360:75-97.
71. Weng AP, Ferrando AA, Lee W, Morris JPt, Silverman LB, Sanchez-Irizarry C, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science. 2004;306(5694):269-71.
72. Hebert J, Cayuela JM, Berkeley J, Sigaux F. Candidate tumor-suppressor genes MTS1 (p16INK4A) and MTS2 (p15INK4B) display frequent homozygous deletions in primary cells from T- but not from B-cell lineage acute lymphoblastic leukemias. Blood. 1994;84(12):4038-44.
73. Van der Meulen J, Sanghvi V, Mavrakis K, Durinck K, Fang F, Matthijssens F, et al. The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. Blood. 2015;125(1):13-21.
74. Van Vlierberghe P, Palomero T, Khiabanian H, Van der Meulen J, Castillo M, Van Roy N, et al. PHF6 mutations in T-cell acute lymphoblastic leukemia. Nat Genet. 2010;42(4):338-42.
75. Tosello V, Mansour MR, Barnes K, Paganin M, Sulis ML, Jenkinson S, et al. WT1 mutations in T-ALL. Blood. 2009;114(5):1038-45.
76. Mendes RD, Sarmento LM, Cante-Barrett K, Zuurbier L, Buijs-Gladdines JG, Povoa V, et al. PTEN microdeletions in T-cell acute lymphoblastic leukemia are caused by illegitimate RAG-mediated recombination events. Blood. 2014;124(4):567-78.
77. Palomero T, Sulis ML, Cortina M, Real PJ, Barnes K, Ciofani M, et al. Mutational loss of PTEN induces resistance to NOTCH1 inhibition in T-cell leukemia. Nat Med. 2007;13(10):1203-10.
78. Clappier E, Cuccuini W, Kalota A, Crinquette A, Cayuela JM, Dik WA, et al. The C-MYB locus is involved in chromosomal translocation and genomic duplications in human T-cell acute leukemia (T-ALL), the translocation defining a new T-ALL subtype in very young children. Blood. 2007;110(4):1251-61.
79. Lahortiga I, De Keersmaecker K, Van Vlierberghe P, Graux C, Cauwelier B, Lambert F, et al. Duplication of the MYB oncogene in T cell acute lymphoblastic leukemia. Nat Genet. 2007;39(5):593-5.
80. Shochat C, Tal N, Bandapalli OR, Palmi C, Ganmore I, te Kronnie G, et al. Gain-of-function mutations in interleukin-7 receptor-alpha (IL7R) in childhood acute lymphoblastic leukemias. J Exp Med. 2011;208(5):901-8.
81. Zenatti PP, Ribeiro D, Li W, Zuurbier L, Silva MC, Paganin M, et al. Oncogenic IL7R gain-of-function mutations in childhood T-cell acute lymphoblastic leukemia. Nat Genet. 2011;43(10):932-9.
82. Flex E, Petrangeli V, Stella L, Chiaretti S, Hornakova T, Knoops L, et al. Somatically acquired JAK1 mutations in adult acute lymphoblastic leukemia. J Exp Med. 2008;205(4):751-8.
83. O'Neil J, Grim J, Strack P, Rao S, Tibbitts D, Winter C, et al. FBW7 mutations in leukemic cells mediate NOTCH pathway activation and resistance to gamma-secretase inhibitors. J Exp Med. 2007;204(8):1813-24.
84. Thompson BJ, Buonamici S, Sulis ML, Palomero T, Vilimas T, Basso G, et al. The SCFFBW7 ubiquitin ligase complex as a tumor suppressor in T cell leukemia. J Exp Med. 2007;204(8):1825-35.
85. Della Gatta G, Palomero T, Perez-Garcia A, Ambesi-Impiombato A, Bansal M, Carpenter ZW, et al. Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. Nat Med. 2012;18(3):436-40.
86. Grossmann V, Haferlach C, Weissmann S, Roller A, Schindela S, Poetzinger F, et al. The molecular profile of adult T-cell acute lymphoblastic leukemia: mutations in RUNX1 and DNMT3A are associated with poor prognosis in T-ALL. Genes Chromosomes Cancer. 2013;52(4):410-22.
87. Gutierrez A, Kentsis A, Sanda T, Holmfeldt L, Chen SC, Zhang J, et al. The BCL11B tumor suppressor is mutated across the major molecular subtypes of T-cell acute lymphoblastic leukemia. Blood. 2011;118(15):4169-73.
88. Aster JC. In brief: Notch signalling in health and disease. The Journal of pathology. 2014;232(1):1-3.
89. Wang H, Zang C, Taing L, Arnett KL, Wong YJ, Pear WS, et al. NOTCH1-RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. Proc Natl Acad Sci U S A. 2014;111(2):705-10.
90. Nagel AC, Krejci A, Tenin G, Bravo-Patino A, Bray S, Maier D, et al. Hairless-mediated repression of notch target genes requires the combined activity of Groucho and CtBP corepressors. Mol Cell Biol. 2005;25(23):10433-41.
91. Hori K, Sen A, Artavanis-Tsakonas S. Notch signaling at a glance. Journal of cell science. 2013;126(Pt 10):2135-40.
92. Ellisen LW, Bird J, West DC, Soreng AL, Reynolds TC, Smith SD, et al. TAN-1, the human homolog of the Drosophila notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. Cell. 1991;66(4):649-61.
93. La Starza R, Borga C, Barba G, Pierini V, Schwab C, Matteucci C, et al. Genetic profile of T-cell acute lymphoblastic leukemias with MYC translocations. Blood. 2014;124(24):3577-82.

94. Erikson J, Finger L, Sun L, ar-Rushdi A, Nishikura K, Minowada J, et al. Deregulation of c-myc by translocation of the alpha-locus of the T-cell receptor in T-cell leukemias. Science. 1986;232(4752):884-6.
95. Herranz D, Ambesi-Impiombato A, Palomero T, Schnell SA, Belver L, Wendorff AA, et al. A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. Nat Med. 2014;20(10):1130-7.
96. Palomero T, Lim WK, Odom DT, Sulis ML, Real PJ, Margolin A, et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. Proc Natl Acad Sci U S A. 2006;103(48):18261-6.
97. Medyouf H, Gusscott S, Wang H, Tseng JC, Wai C, Nemirovsky O, et al. High-level IGF1R expression is required for leukemia-initiating cell activity in T-ALL and is supported by Notch signaling. J Exp Med. 2011;208(9):1809-22.
98. Reizis B, Leder P. Direct induction of T lymphocyte-specific gene expression by the mammalian Notch signaling pathway. Genes Dev. 2002;16(3):295-300.
99. Gonzalez-Garcia S, Garcia-Peydro M, Martin-Gayo E, Ballestar E, Esteller M, Bornstein R, et al. CSL-MAML-dependent Notch1 signaling controls T lineage-specific IL-7R{alpha} gene expression in early human thymopoiesis and leukemia. J Exp Med. 2009;206(4):779-91.
100. Chan SM, Weng AP, Tibshirani R, Aster JC, Utz PJ. Notch signals positively regulate activity of the mTOR pathway in T-cell acute lymphoblastic leukemia. Blood. 2007;110(1):278-86.
101. Joshi I, Minter LM, Telfer J, Demarest RM, Capobianco AJ, Aster JC, et al. Notch signaling mediates G1/S cell-cycle progression in T cells via cyclin D3 and its dependent kinases. Blood. 2009;113(8):1689-98.
102. Dohda T, Maljukova A, Liu L, Heyman M, Grander D, Brodin D, et al. Notch signaling induces SKP2 expression and promotes reduction of p27Kip1 in T-cell acute lymphoblastic leukemia cell lines. Experimental cell research. 2007;313(14):3141-52.
103. Vilimas T, Mascarenhas J, Palomero T, Mandal M, Buonamici S, Meng F, et al. Targeting the NF-kappaB signaling pathway in Notch1-induced T-cell leukemia. Nat Med. 2007;13(1):70-7.
104. Ntziachristos P, Tsirigos A, Van Vlierberghe P, Nedjic J, Trimarchi T, Flaherty MS, et al. Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. Nat Med. 2012;18(2):298-301.
105. Palomero T, Barnes KC, Real PJ, Glade Bender JL, Sulis ML, Murty VV, et al. CUTLL1, a novel human T-cell lymphoma cell line with t(7;9) rearrangement, aberrant NOTCH1 activation and high sensitivity to gamma-secretase inhibitors. Leukemia. 2006;20(7):1279-87.
106. Deangelo DJ, Stone RM, Silverman LB, Stock W, Attar EC, Fearen I, et al. A phase I clinical trial of the notch inhibitor MK-0752 in patients with T-cell acute lymphoblastic leukemia/lymphoma (T-ALL) and other leukemias. J Clin Oncol. 2006;24(18_suppl):6585.
107. Riccio O, van Gijn ME, Bezdek AC, Pellegrinet L, van Es JH, Zimber-Strobl U, et al. Loss of intestinal crypt progenitor cells owing to inactivation of both Notch1 and Notch2 is accompanied by derepression of CDK inhibitors p27Kip1 and p57Kip2. EMBO Rep. 2008;9(4):377-83.
108. Hales EC, Taub JW, Matherly LH. New insights into Notch1 regulation of the PI3K-AKT-mTOR1 signaling axis: targeted therapy of gamma-secretase inhibitor resistant T-cell acute lymphoblastic leukemia. Cell Signal. 2014;26(1):149-61.
109. Sanda T, Li X, Gutierrez A, Ahn Y, Neuberg DS, O'Neil J, et al. Interconnecting molecular pathways in the pathogenesis and drug sensitivity of T-cell acute lymphoblastic leukemia. Blood. 2010;115(9):1735-45.
110. Wu Y, Cain-Hom C, Choy L, Hagenbeek TJ, de Leon GP, Chen Y, et al. Therapeutic antibody targeting of individual Notch receptors. Nature. 2010;464(7291):1052-7.
111. Sulis ML, Saftig P, Ferrando AA. Redundancy and specificity of the metalloprotease system mediating oncogenic NOTCH1 activation in T-ALL. Leukemia. 2011;25(10):1564-9.
112. Moellering RE, Cornejo M, Davis TN, Del Bianco C, Aster JC, Blacklow SC, et al. Direct inhibition of the NOTCH transcription factor complex. Nature. 2009;462(7270):182-8.

2 NONCODING RNAS

"Only a small proportion of the RNA made in the nucleus of animal and higher plant cells serves as a template for the synthesis of protein. This RNA is characterized by its ability to assume a form which protects it from intracellular degradation. Most of the nuclear RNA, however, is made on parts of the DNA which do not contain information for the synthesis of specific proteins."

-Henry Harris, 1965-

In 1964, Henry Harris discovered that a large amount of RNA molecules in the nucleus were not transferred to the cytoplasm where translation takes places[1]. This led to the comment one year later in a book chapter in "Evolving Genes and Proteins"[2], describing for the first time the existence of noncoding RNAs. H. Harris saw these noncoding transcripts as a part of evolution and assumed these molecules would undergo mutations and evolve into messenger RNAs (mRNAs). However, science only needed a few more years to appreciate the existence and function of noncoding RNAs as R.W. Holley was awarded with the Nobel Prize for his research that led to the identification of the structure and sequence of the transfer RNA (tRNA) for alanine[3-5].

2.1 Discovery and diversity of noncoding RNAs

In the late 1950s, Francis Crick hypothesized about the flow of genetic information and came with the idea he called "The central dogma". This theory states that DNA is transcribed into RNA, which is then translated into protein, the molecule with the active function, and that the information cannot be transferred back from protein to nucleic acid[6]. It was just a few years later that the first RNA molecule with a function by itself was described by R.W. Holley, the tRNA[3, 4]. A tRNA is the adaptor between the mRNA and the amino acid sequence of proteins, following the genetic code. Other important functional noncoding RNAs, the ribosomal RNAs (rRNAs), were discovered in the 1980s and it appeared that they were also involved in protein synthesis as important components of the ribosomal ribonucleoprotein complex. This led to the discovery that RNA molecules can have a catalytic function, calling them ribozymes[7-9].

Since the discovery of DNA, RNA and proteins, molecular biology research has boomed and more functional noncoding RNA molecules were discovered to be involved in several crucial pathways in the cell. Most of these RNAs are now characterized as small noncoding RNAs, as they are not longer than 200 nucleotides. DNA replication makes use of the small noncoding Y RNAs for their interaction with chromatin and initiation proteins[10] and these Y RNAs are required for cell proliferation in tumors[11]. Furthermore, small nuclear RNAs (snRNAs) are part of the spliceosome, having a role in pre-mRNA splicing[12] and small nucleolar RNAs (snoRNAs) are involved in the RNA-editing process[13, 14]. Piwi-interacting RNAs (piRNAs) are necessary for

spermatogenesis and function by guiding the PIWI proteins to transposons to induce their silencing[15]. Another example of small ncRNAs are microRNAs which are involved in post-transcriptional gene regulation, this type of ncRNAs will be discussed in more detail in section 2.2. Very recently, the involvement of a short guide RNA (gRNA) was discovered in the CRISPR/Cas system in the prokaryotic immune defense. This system uses a Cas endonuclease to cleave viral DNA and the gRNA is necessary for the guidance of the endonuclease complex to the viral DNA[16-18].

Another important noncoding RNA is the Telomerase RNA Component (*TERC*). The telomerase ribonucleoprotein is responsible for the maintenance of the telomere length at the end of the chromosomes. The TERC RNA of this complex is over 500 nucleotides in size and has a dual function. It acts as a scaffold for several proteins of the complex and it also contains the template RNA sequence for DNA synthesis by the Telomerase Reverse Transcriptase (TERT)[19]. In the last decade, much work has been done on the so-called long noncoding RNAs (lncRNAs), noncoding RNA molecules with a length of at least 200 nucleotides, and several of these lncRNAs have been discovered with a broad diversity of functions. One a example is the *XIST* lncRNA that is involved in dosage compensation by recruiting the PRC2-complex that enhances epigenetic silencing of one X-chromosome[20, 21]. The role of lncRNAs in the cell and in cancer is discussed in more detail in section 2.3.

With the discovery of several functional noncoding RNAs it became clear that the genome was more than just DNA coding for proteins (2%) and 'junk' DNA (98%). The group of John Mattick further established the importance of the noncoding DNA. They showed that the ratio of noncoding DNA to the full genomic sequence could explain the complexity of species, with low complex species having relatively less noncoding DNA and with *Homo sapiens* on top of the list[22]. This was a response to the G-value paradox, which states that there is no relationship between the amount of coding genes and the species complexity[23]. With this, it is hypothesized that the key to the complexity of organisms lies to a large extent in the size of the genome from which noncoding RNAs are transcribed. One of the recent estimates is that less than 2% of the genome is coding for proteins and at least 75% is actively transcribed into noncoding RNAs[24]. In the next parts, the role of miRNAs and lncRNAs in the normal cellular processes and in cancer development will be discussed in more detail.

2.2 MicroRNAs

2.2.1 MicroRNA biogenesis and function

In 1993, the Ambros team described microRNAs (miRNAs) for the first time after the identification of *lin-4* in *C. elegans*[25]. From the year 2000 on, the process of miRNA biogenesis was studied in depth and it was shown that miRNAs play an important role in development and disease. MicroRNAs are defined as short noncoding RNAs with a length of approximately 21 nucleotides that function as post-transcriptional repressors of their target genes[26-28]. They bind with their seed sequence (mostly nucleotides 2 to 8 starting from the 5' end of the miRNA) to a comple mentary sequence in the 3' untranslated region (3'UTR) of the target mRNA[29]. The importance

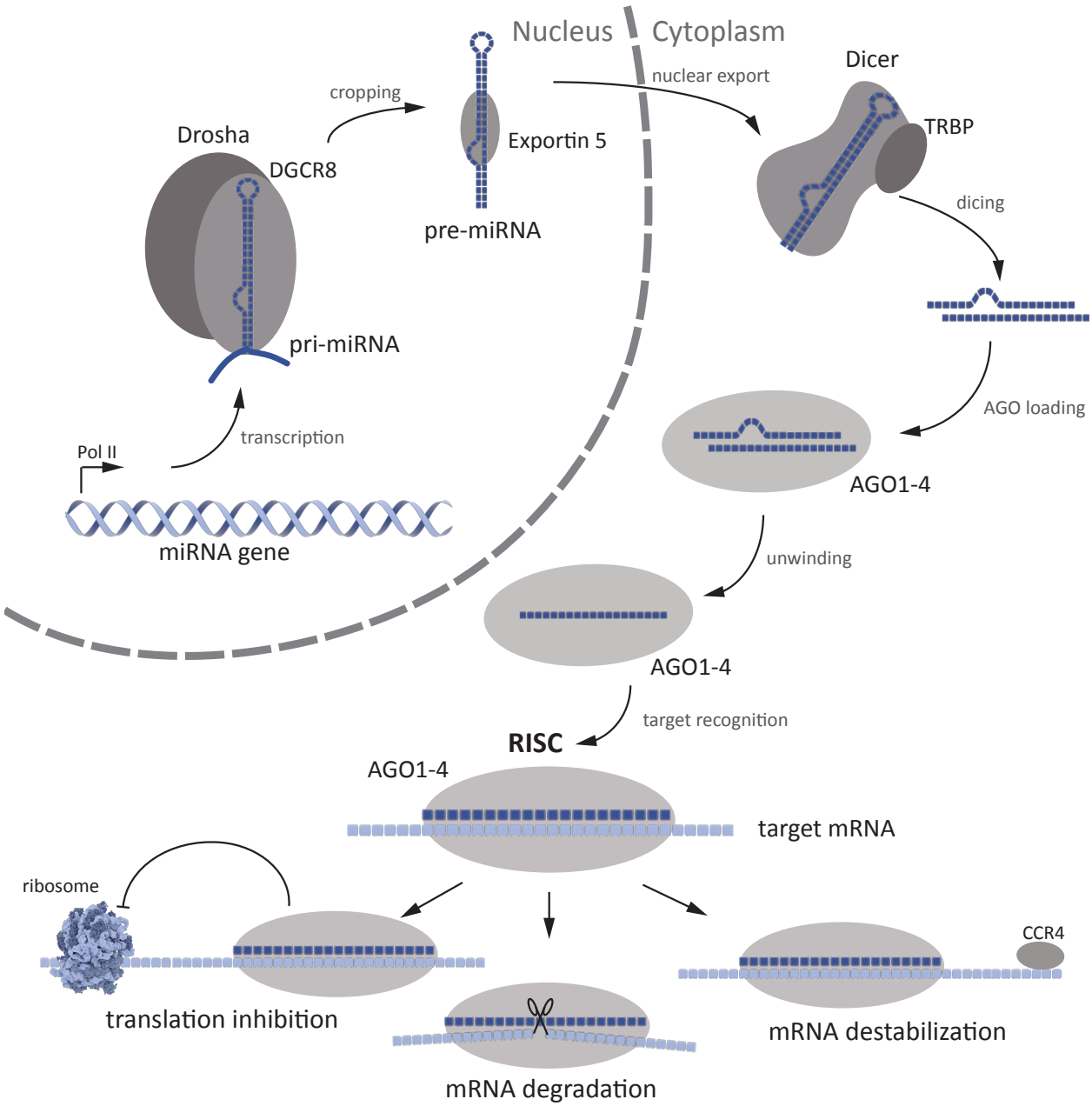


Figure 4. MicroRNA biogenesis. MicroRNAs are transcribed by RNA Pol II in the nucleus into pri-miRNAs. The pri-miRNAs are processed by the RNase III enzyme DROSHA into pre-miRNA that is exported to the cytoplasm by Exportin-5. In the cytoplasm the pre-miRNA is processed by another RNase III enzyme DICER forming a miRNA-duplex. One strand is loaded into the RISC complex together with the Argonaute proteins and the miRNA guides the RISC complex to its target mRNA(s). The RISC complex inhibits the mRNA function through translation inhibition, mRNA degradation and/or mRNA destabilization. (Figure adapted from [31, 32])

of miRNAs in gene regulation is suggested by the large amount of possible target genes particular miRNAs have and the variety of different miRNAs that can bind the 3'UTR of specific mRNAs[30].

MicroRNAs are transcribed from intra- or intergenic regions, leading to large primary miRNA (pri-miRNA) transcripts consisting of several hairpin structures. These transcripts are processed in the nucleus by the ribonuclease DROSHA to single hairpin structures, called precursor miRNAs (pre-miRNA). Pre-miRNAs are transported to the cytoplasm and processed by another ribonuclease, DICER, forming a mature double stranded miRNA. One strand, the guide strand, is incorporated into the 'RNA-induced silencing complex' (RISC) and guides RISC to the 3'UTR of its target mRNA. The RISC complex negatively regulates the mRNA through either the inhibition of translation, mRNA destabilization and/or the degradation of the mRNA[31]. The effect of one miRNA on the expression of a particular mRNA can be modest, but due to the multiple seed sequences for different miRNAs in the 3'UTR of the mRNA, the combined effect of several miRNAs can have a significant result on the overall mRNA expression (see Figure 4).

The post-transcriptional regulation by several miRNAs of one target mRNA allows strict regulation in time and context. It is thus not a surprise that miRNAs have been detected as key players in several fundamental processes in the cell, such as development, differentiation, proliferation and cell death, but this also implies that deregulation of miRNA expression can lead to disease and cancer development.

2.2.2 MicroRNAs in cancer

MicroRNAs are involved in several processes in normal cellular development, in which they can target tumor suppressor genes (as oncomiRs) or proto-oncogenes (as tumor suppressor miRNAs). If this strict process of miRNA-controlled target gene expression is disturbed, cancer development can occur. Several different oncogenic events in this process have been described in a plethora of cancer types (Figure 5).

The downregulation of tumor suppressor miRNA function can be caused by the genetic loss, epigenetic silencing or transcriptional repression of the miRNA gene or through a general downregulation of the miRNA biogenesis pathway (ex. reduced DICER levels[33], or a mutation in Exportin-5 leading to a reduced nuclear export of the pre-miRNA[34]). Furthermore, the targeted mRNAs can escape repression from a specific miRNA as a result of a mutation in the seed sequence or through expression of an isoform with a reduced 3'UTR. Recent discoveries also detected that other noncoding RNAs can influence the miRNA action. These ncRNAs, called "competing endogenous RNA" (ceRNA), compete for the miRNA binding, leading to an increase in expression of the target gene[35].

The miRNA cluster *miR-15a/16-1* was the first miRNA cluster described to have a tumor suppressive role. This locus is deleted in more than half of chronic lymphocytic leukemia (CLL) patients (del(13q14)), with one of its primary targets being the *BCL2* oncogene[36, 37]. The *let-7* cluster is another example of which the tumor suppressor miRNA members show reduced expression levels in different cancer types[38]. Known targets of this cluster are the *RAS* oncogenes,

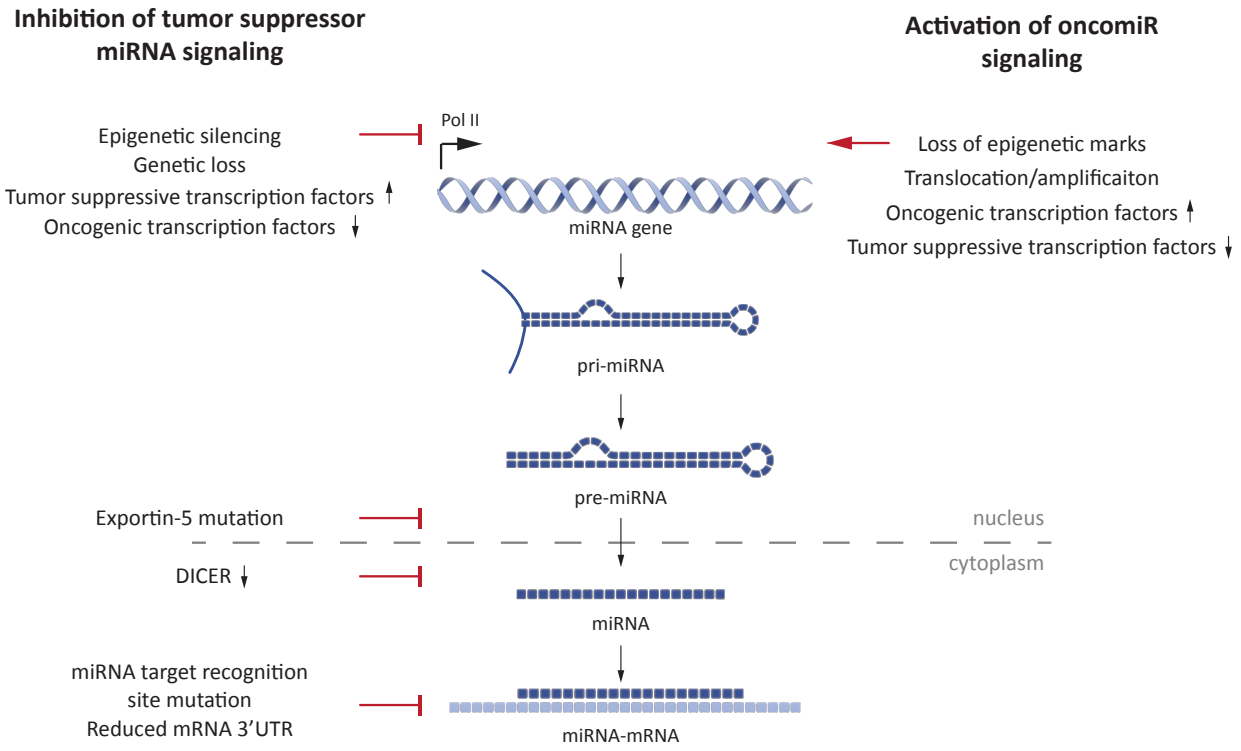


Figure 5. Deregulation of miRNAs in cancer development. The reduction of a miRNA that functions as a tumor suppressor can lead to tumor formation. This reduction can occur due to changes at the miRNA gene locus or in the miRNA transcriptional regulation. Inhibition of the general miRNA biogenesis machinery also reduces miRNA functions. The overexpression of oncomiRs can occur from genetic changes (translocation or amplification of the region) or increased transcription due to oncogenic transcription factors. (Figure adapted from [32])

MYC and *HMGA*[39-41]. It was also shown that both tumor suppressor miRNA clusters described above are repressed by the MYC oncogene[42].

The overexpression of oncogenic miRNAs or oncomiRs can be caused by the loss of repressive epigenetic marks, activation resulting from translocation and increased expression due to activated oncogenic transcription factors. The *miR-17~92* cluster (*oncomiR-1*) is a prototypical oncogenic miRNA cluster[43, 44]. This cluster contains 6 different miRNAs and has two paralogue clusters that are deregulated in several cancer types. The overexpression of this cluster is due to gain or amplification of the miRNA cluster locus, the direct activation of the *MYC* oncogene or the decreased repression by the tumor suppressor gene *TP53*[45-48]. This cluster plays an important regulatory role in different pathways in the cell, such as in cell cycle and proliferation, development, angiogenesis etc. Some important targets of the *miR-17~92* cluster are the tumor suppressor genes *PTEN*, *BIM* and *p21* (*CDKN2A*)[48, 49].

Several comprehensive reviews have been published on the role of miRNAs in cancer development[32, 49-52]. Not surprisingly, several oncomirs and tumor suppressor miRNAs have been identified to play an important role in T-ALL development as well. A detailed overview of this can be found in the review in the next part of this introduction.

2.3 Long noncoding RNAs

2.3.1 Long noncoding RNAs emerging from the dark side of the genome

It was already known for a long time that functional RNAs were part of the cellular machinery (see section 2.1). Technological developments driven by the human genome project at the beginning of the twenty-first century[53, 54], including tiling microarrays for gene expression analysis and novel sequencing technologies[55, 56], fueled the discovery of novel long noncoding RNA (lncRNA) genes. Since then, several microarray hybridization and deep sequencing analyses provided further in depth studies and now it is estimated that at least roughly 75% of the human genome is transcribed[24]. While this finding is overwhelming, most lncRNAs remain functionally unannotated and it remains to be determined how many lncRNAs effectively execute specific cellular functions.

More evidence supporting the transcription of ncRNAs came from the epigenetics field that profiled specific histone markers across the genome and discovered methods to detect open (or active) chromatin. Several, but not all, of the previously detected ncRNAs were characterized by DNase1 hypersensitivity, active chromatine marks (ex. H3K4me3 marks at the promoter and H3K36me3 marks along the transcriptional region) and the binding of transcription factors[57-61]. These chromatin signatures led to the discovery that there is some species conservation of chromatin marks and the promoter regions of lncRNAs, although sequence conservation is very low[57, 62].

Even before these large scale studies, some lncRNAs were already discovered. In the early 1990s *H19* and *XIST* were detected as lncRNAs with a role in epigenetic regulation[20, 21, 63] and also the *TERC* RNA of the telomerase complex was well known[19]. Since 2010, research on lncRNAs has accelerated, leading to the discovery of multiple lncRNAs with a large variety of functions as well as the discovery of potential disease biomarker lncRNAs.

2.3.2 Long noncoding RNA characteristics

Long noncoding RNAs are arbitrarily defined as RNA molecules with a length of at least 200 nucleotides and without protein coding potential. Several algorithms, for example CPC[64], CPAT[65] and PhyloCSF[66], can confirm this lack of protein coding potential. LncRNAs share several characteristics with mRNAs, as most of them are multi-exonic and transcribed by RNA polymerase II, giving them the 5' Cap-structure and 3' poly-A tail. In comparison to mRNAs however, lncRNAs generally have lower expression levels and show tissue or cell type specific expression patterns. The latter characteristic offers perspectives for lncRNAs as potential targets for therapy, as their specific expression pattern might avoid off-target effects of specific lncRNA inhibitors.

One possible classification of lncRNAs is based on the location of the lncRNA gene in comparison to protein coding genes (Figure 6). Intergenic lncRNAs or 'lincRNAs' do not overlap with protein coding genes. As these lincRNAs are quite easy to define, the first described lncRNAs were intergenic, for example *XIST*[20, 21], *H19*[63], and *MALAT1*[67]. Like miRNAs and snoRNAs, lncRNAs can also be located in introns of protein coding genes. One example of this category is *COLDAIR*, one of the first described lncRNAs in plant biology, which is located in the first intron

of the Flowering Locus C (*FLC)* gene and is necessary for the epigenetic repression of *FLC* through the recruitment of PRC2 (Polycomb Repressive Complex 2)[68]. The third class of lncRNAs includes the lncRNA genes that overlap with a known protein-coding gene in the antisense direction, of which *HOTAIR* (HOX transcript antisense RNA) is a well-known example. *HOTAIR* is located in the *HOXC* locus and regulates the repression of the *HOXD* locus through PRC2 recruitment[69]. Some lncRNAs have also been detected that overlap with a protein-coding gene in the sense direction, with an exon of the protein-coding gene located in an intron of the lncRNA gene (sense-overlapping lncRNAs, for example *SOX2-OT*[70]). Another category includes lncRNAs of which the transcription start site is located in the vicinity of a protein-coding gene, but is transcribed from the opposite strand. These lncRNAs are called bi-directional lncRNAs.

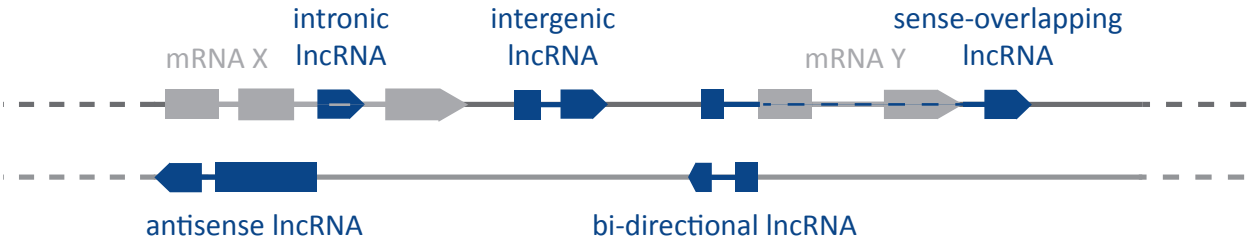


Figure 6. Long noncoding RNA classes. LncRNA genes can be classified based on their location according to protein coding genes: intronic lncRNAs, intergenic lncRNAs, antisense lncRNAs, bi-directional lncRNAs or sense-overlapping lncRNAs. (Figure based on [71])

2.3.3 Long noncoding RNA functions

Long noncoding RNAs have been described in several processes of the cellular machinery, mainly acting as guides, scaffolds or decoys (Figure 7). Guiding lncRNAs bind to transcription factors or epigenetic regulators and "guide" them to a genomic location (ex. *XIST* and the epigenetic regulator *PRC2*), scaffolds bring several proteins together in one complex (ex. *TERC* in the telomerase complex) and decoys bind to proteins by which they inhibit the binding of the protein to the DNA (ex. *GAS5* and the glucocorticoid receptor).

Many lncRNAs identified so far are implicated in transcriptional regulation. Of particular interest, many lncRNAs have been shown to interact with the PRC2-complex, thus possibly guiding PRC2 to a specific locus on the genome, leading to the epigenetic silencing of that locus by PRC2-mediated H3K27 trimethylation[72, 73]. Another class of lncRNAs are the so-called enhancer RNAs or eRNAs. These eRNAs are transcribed from enhancer sites and aid in the attraction of transcription factors to the enhancers and the transcription start site of the nearby gene. Furthermore, eRNAs can also help in the chromatin looping bringing the enhancer site closer to the transcription start site[74, 75]. However, these eRNAs are nowadays more and more considered as a separate entity of RNA molecules as several eRNAs appear to have some characteristics devi-

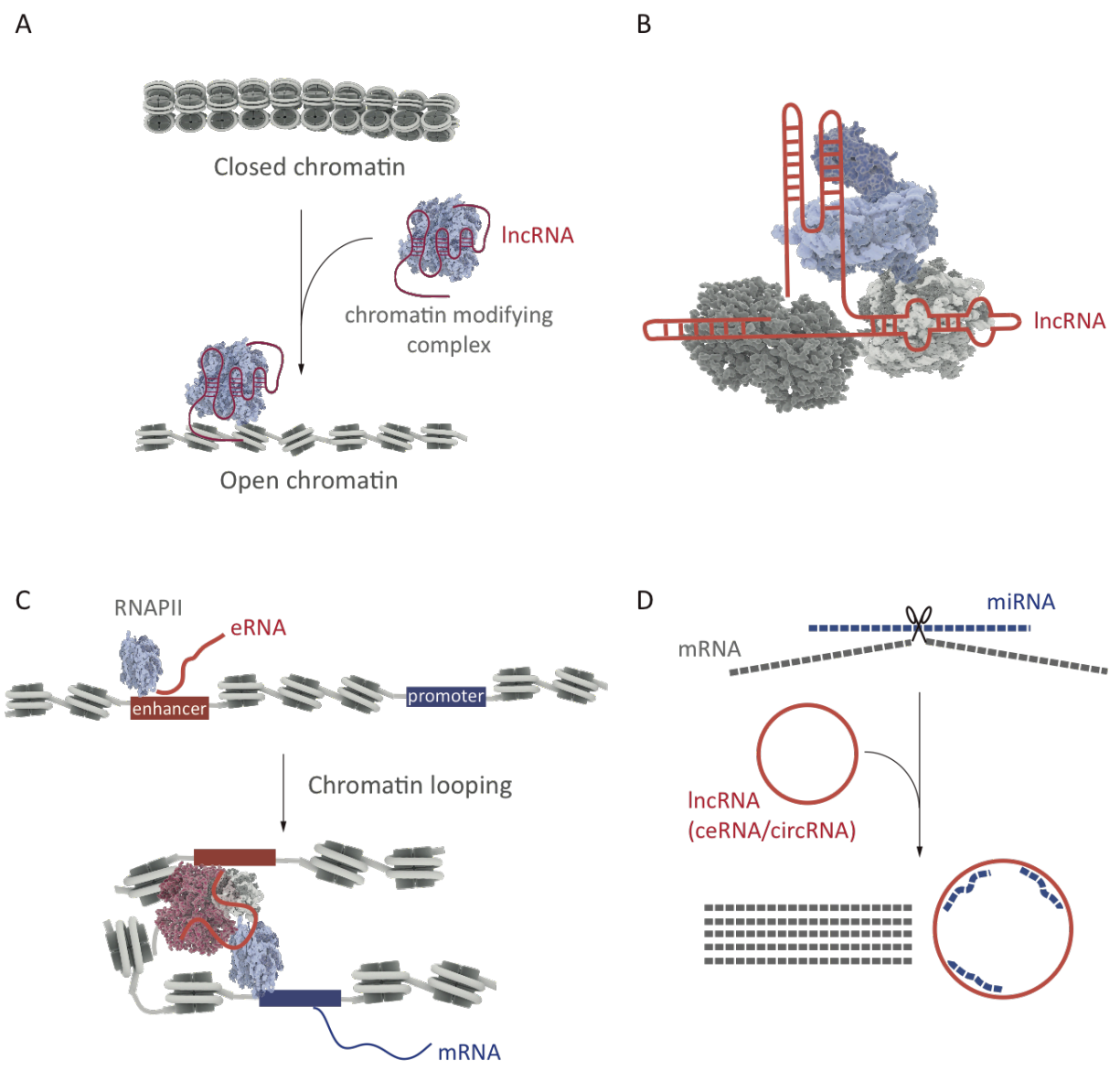


Figure 7. Long noncoding RNA functions. (A) LncRNAs can interact with chromatin modifying complexes and guide them to specific loci on the genome. (B) LncRNAs can be scaffolds, bringing several proteins into one complex. (C) Enhancer RNAs are transcribed from enhancer loci and can aid in chromatin looping, bringing transcription factors in close proximity of the promoter of a neighboring gene. (D) CircRNAs or ceRNAs bind miRNAs, titrating them away from their mRNA target, leading to an increased mRNA expression.

ating from general lncRNAs (ex. some are only 50 nucleotides in length and most of them are single-exonic). Nevertheless, the role of eRNAs is still a matter of debate and further investigations are required to clarify to what extent they are the result of divergent RNA Pol II mediated transcription from the target locus rather than an RNA with a specific functional role in enhancer biology. In the end, both might be true, and some enhancer RNAs might be critically involved in enhancer functions whereas others might be transcriptional by-products. (Figure 7C)

Also in several post-transcriptional steps lncRNAs can be involved, for example in splicing control (ex. *MALAT1*) or RNA-editing. One interesting class of lncRNAs comprises the competitive endogenous RNAs (ceRNAs). These ceRNAs contain several miRNA binding sites that can keep

miRNAs away from their normal mRNA target, leading to the enhanced expression of the mRNA target. The *PTENP1* pseudogene of the tumor suppressor gene *PTEN* is an example of a ceRNA as it acts as a decoy for miRNAs that target the *PTEN* gene, implying also a tumor suppressor role for *PTENP1*[35, 76]. (Figure 7D)

LncRNAs have been studied in different developmental processes, as they are known to have a very specific expression patterns. One landmark study by the group of John Rinn made 18 knockout mouse models of intergenic lncRNAs that were conserved between humans and mice[77]. Three homozygous knockout models were incompatible with life, showing an important role for the lncRNA in viability (*Fendrr*$^{-/-}$, *Peril*$^{-/-}$ and *Mdgt*$^{-/-}$) and two other lncRNA knockouts had a major impact on development (*linc-Pint*$^{-/-}$ and *linc-Brn1b*$^{-/-}$). In Table 3, an overview is presented of some of the lncRNAs with a role in the cell and in organ development.

Table 3. Long noncoding RNAs in normal development. (A list of abbreviations can be found at the beginning of this thesis.)

lncRNA	Function	Mechanism	Mouse models	Ref.
Fendrr	?	Interacts with PRC2	Knockout is perinatally lethal due to respiratory defects (defects in lung maturation and vasculature)	62, 77, 78
H19	Allelic imprinting at the IGN locus (Imprinted Gene Network)	Interacts with MBD1 that brings additional histone methyl-transferases	Knockout results in overweight	63, 79, 80
HOTAIR	Inactivation of *HOXD* locus and some imprinted loci	Recruits PRC2 (H3K27 trimethylation) Interacts with the LSD1 complex (H3K4 demethylation)	Deletion leads to spine and bone malformation	69, 81, 82
HOTTIP	Activation of 5' *HOXA* genes	Recruits the WDR5/MLL complex by chromosomal looping to the *HOXA* locus (H3K4 trimethylation)	Knockdown shortens forelimbs	83
MALAT1	Alternative splicing Transcription control	Interaction with SRSF1 modulating it by phosphorylation Modulates nuclear speckle association of various splicing factors Interaction with chromatin readers at genome loci	Knockdown has no effect on mouse development	84-87
NEAT1	Nuclear paraspeckle formation	Interacts with paraspeckle proteins and transcriptional repressors	Knockout leads to loss of paraspeckles, but has no phenotype	58, 88-90
TERC	Telomere maintenance	Scaffold for the telomerase complex Template for telomere extension	Knockout leads to early aging	19, 91
XIST	Inactivation of one of the X chromosomes	Recruits PRC2 (H3K27 trimethylation)	Partial deletion is embryonic lethal in female mice	21, 92, 93

2.3.4 *Long noncoding RNA deregulation in cancer*

Together with the discovery of lncRNAs playing a role in normal cell homeostasis, several lncRNAs were discovered with a differential expression pattern in one or more cancer types. Some identified cancer-related lncRNAs also had a function in normal development, but several were only expressed in that specific cancer type, making them ideal candidates for targeted therapy. Furthermore, lncRNAs were described as biomarkers for cancer type, stage or prognosis. Still most of the cancer-associated lncRNAs are only identified based on expression profiling and have unknown action mechanisms. However, with the small amount of lncRNAs that do have a known role in cancer development, it became clear that they can have a role in different cancer pathways that contribute to the hallmarks of cancer (see Figure 8, Table 4)[94, 95].

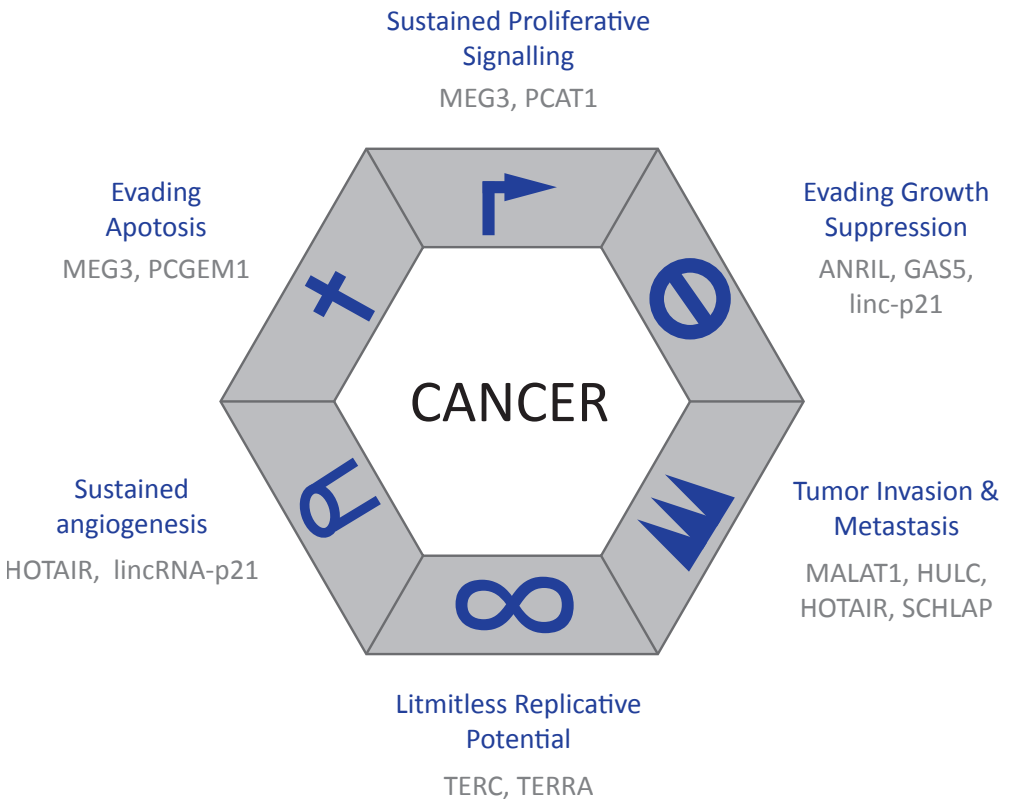


Figure 8. Long noncoding RNAs associated with hallmarks of cancer. LncRNAs have been discovered in the several processes that lead to cancer development. (Figure based on [95, 96])

The deregulated expression of lncRNAs in cancer is mostly due to a change in transcriptional regulation as several lncRNAs are regulated by known oncogenes and tumor suppressor genes. However, there are also lncRNAs that are upregulated by different mechanisms. For example, the *NBAT1* (Neuroblastoma-Associated Transcript 1) gene contains a high-risk neuroblastoma single nucleotide polymorphism (SNP), decreasing its expression and leading to higher proliferation and invasion of neuroblastoma cells[97].

To gain insight into lncRNA function, 'guilt-by-association' analysis can offer a first glimpse by comparing the expression pattern of a specific lncRNA with all protein coding genes and linking the lncRNA to specific oncogenic pathways[57]. However, this only provides a correlation and *in*

vitro and *in vivo* studies should be performed to further assess the role of the lncRNA in the oncogenic process. Challenges for *in vitro* studies include low success rates of shRNA/siRNA-mediated knockdown as the RNAi machinery is present in the cytoplasm and most lncRNAs are localized in the nucleus, but also the uncertainty of the annotation of lncRNA transcript sequence and *cis*-regulatory functions make overexpression challenging. Furthermore, poor species conservation for many lncRNAs often precludes *in vivo* studies. These challenges and solutions are discussed in more detail in the review in the next part of this introduction. Despite these challenges, several lncRNAs could convincingly be linked to several cancer processes, such as the melanoma-specific lncRNA *SAMMSON* (as discussed in Box 2 in more detail)[98]. An overview of a selection of lncRNAs with a role in cancer can be found in Table 4. It should however be stated that some lncRNAs have different functions in different cancer types and not everything could be included. An overview of lncRNAs with a specific role in T-ALL can be found in the review in the last part of this introduction.

Box 2. *SAMMSON*: a melanoma-specific lncRNA with therapeutic potential[98].

The SAMMSON lncRNA is focally amplified in 10% of melanoma cases together with the *MITF* oncogene. Next to that, it appeared to be expressed in over 90% of melanoma cases, as a result of SOX10-mediated transcription, without a correlation to MITF expression. *SAMMSON* was also exclusively expressed in melanoma samples of the 'The Cancer Genome Atlas' (TCGA) cohort.

In vitro knockdown of the lncRNA reduced the clonogenicity of melanoma cultures and forced overexpression of *SAMMSON* could rescue this effect. This indicates that *SAMMSON* exerts a pro-survival function *in trans*. Further *in vitro* experiments showed that *SAMMSON* interacts with p32, a protein involved in mitochondrial metabolism, and enhances its mitochondrial localization and function.

Intravenous treatment of a patient-derived xenograft (PDX) melanoma model with an LNA GapmeR targeting *SAMMSON* could suppress the growth of melanoma tumors, decrease cell proliferation and increase apoptosis. Most interestingly, as *SAMMSON* is melanoma specific expressed, this treatment did not cause any adverse reaction in mice. The therapeutic use of these antisense drugs targeting *SAMMSON* might be additive to the BRAF inhibitors used today as cells that acquired BRAF inhibitor resistance are still in need of *SAMMSON* signaling.

Table 4. Long noncoding RNAs with an oncogenic or tumor suppressor role in cancer. Inspired by recent reviews[96, 99]. (A list of abbreviations can be found at the beginning of this thesis.)

lncRNA	Cancer assocation	Mechanism	Phenotype	Ref.
ANRIL	High expression in prostate cancer, gastric cancer and leukemia (cancer-associated SNP)	Silencing of *p15* by PRC2 and CBX7 recruitment	↑ proliferation	100-106
CCAT1	Oncogenic in gastric, colorectal and gall bladder cancer	MYC regulation by chromatin loops Sponge for *miR-218-5p*, increasing BMI1 expression	↑ proliferation ↑ invasion	107-110
CRNDE	Upregulated in colorectal cancer, glioma, liver cancer, medulloblastoma …	Interacts with PRC2, reduces *miR-384* and *miR-186* expression	↑ cell growth ↑ invasion ↓ cell growth ↓ cell cycle progression	111-115
GAS5	Downregulated in breast, prostate, gastric, cervical and renal cell cancers	Inhibits DNA binding of the glucocorticoid receptor	↑ cell arrest ↑ apoptosis ≠ cell metabolism	116-121
H19	Oncogenic in several cancer types (liver, breast, bladder, prostate...)	Interacts with MBD1 that brings additional histone methyltransferases (imprinting)	↑ proliferation ↓ apoptosis	79, 80, 122
HOTAIR	Overexpressed in liver cancer and metastatic breast, lung and pancreatic cancers	Silencing of *HOXD* and other genes through PRC2 and LSD1 binding	↑ metastasis	82, 123-128
HULC	Biomarker for liver cancer	Activated by CREB and sponge for miRNAs	↑ proliferation ↑ invasion ↓ apoptosis	129-132
Linc-PINT	Downregulated in colorectal cancer	Activated by p53, interacts with PRC2 for gene silencing	↑ apoptosis ↓ proliferation	133
LincRNA-p21	Downregulated in several cancer types (lung, cancer, lymphoma …)	Activated by p53, interacts with hnRNP-K for gene silencing	↑ apoptosis	134
MALAT1	Overexpressed in several cancer types (colorectal, breast…). SNP linked to hepatocellular carcinoma.	Alternative splicing	↑ metastasis ↑ proliferation	67, 84, 135
MEG3	Downregulated in several cancer types (prostate, bladder cancer…)	Downregulates *MDM2* and upregulates *p53*. Recruits PRC2 for gene silencing. Regulates TGF-β pathway genes.	↓ proliferation	136-139
NBAT1	SNP in high-risk neuroblastoma reduces expression	Silencing of REST by PRC2	↓ proliferation ↓ invasion	97

Table 4 (Continued). Long noncoding RNAs with an oncogenic or tumor suppressor role in cancer. Inspired by recent reviews[96, 99]. (A list of abbreviations can be found at the beginning of this thesis.)

lncRNA	Cancer assocation	Mechanism	Phenotype	Ref.
NEAT1	Upregulated in several cancer types (prostate and liver cancer, leukemia …)	Paraspeckle formation and gene regulation	Replication stress response ↓ chemosensitivity ↑ proliferation ↑ invasion	88, 140
PCA3	Biomarker for prostate cancer	Androgen receptor signaling modulation	↑ cell survival	141, 142
PCAT1	Upregulated in prostate cancer with disease-associated SNPs	Recruits PRC2 for gene silencing. Inhibits homologous recombination. Activates *c-MYC* and inhibits *BRCA2*.	↑ proliferation	143-145
PCGEM	Upregulated in prostate cancer	Androgen receptor transcriptional activation and *c-MYC* activation	↑ cell growth	146-149
PTENP1	Deleted in melanoma, colon cancer and prostate cancer	Sponge for miRNAs targeting the tumor suppressor gene *PTEN*	↓ proliferation ↓ migration ↓ invasion ↓ tumor growth	35, 150
SAMMSON	Upregulated/amplified in melanoma	Interaction with p32 to regulate mitochondrial metabolism	↑ cell viability	98
SCHLAP1	Upregulated in prostate cancer	Interacts with SWI/SNF complex and inhibits its binding to the genome	↑ metastasis ↑ invasion	151
XIST	Downregulated in breast, ovarian and cervical cancer, testicular cancer, leukemias and lymphomas	PRC2 recruitment to silence one X-chromosome	X-chromosome aneuploidy	21, 152-155

2.4 LncRNAs and trying to see the forest for the trees

Since the discovery of lncRNAs, the scientific area has been divided between 'believers' and 'non-believers' of their importance. One major concern that was raised was the lack of species conservation of these lncRNAs. However, it was stated that these lncRNAs could fold into specific secondary and tertiary structures and that these structures are more important than the sequence itself[156]. Furthermore, promoters of lncRNAs appeared to show a higher level of conservation than the transcript itself, implying that modulation of transcription of the lncRNA might be more important than the transcript itself[157-159]. The Mattick team on the other hand showed that the number of noncoding RNAs was positively correlated to the complexity of the organism, implying that this noncoding part of the genome was responsible for the evolution of complex organisms[22], however this not fully explains the low lncRNA conservation between e.g. human and mice.

Some teams also looked into the probability that these lncRNAs might actually be coding for small peptides as it was discovered that several lncRNAs interact with ribosomes[160-162]. In addition, putative open reading frames in the lncRNAs bound by ribosomes resemble small peptides that evolved very recently in evolution, which could imply that lncRNAs are only an intermediate step to the development of novel proteins[163]. However, others observed that even though lncRNAs are bound by ribosomes, these ribosomes do not act in the same way as for mRNAs and also true functional lncRNAs, like *TERC*, and small RNAs involved in splicing were bound by ribosomes. Therefore, it was concluded that ribosomes are not able to make a distinction between lncRNAs and mRNAs for binding[164]. It was also suggested that the ribosomal interaction would lead to the degradation of the lncRNA[162]. In recent years however, several micropeptides were detected that are less than 100 amino acids long and some of these were translated from lncRNAs, for example CRNDEP, an 84 amino acid peptide translated from the *CRNDE* lncRNA[165-167]. These micropeptides were not detected in previous studies, given the general assumption that proteins should at least contain 100 amino acids and the protein purification methods that discard these smaller fractions.

Also several other concerns arose. For example, some lncRNAs appeared to be part of the 3'UTR of mRNAs of which a specific transcript was not yet detected, due to its low expression[61, 168]. Some lncRNAs were also discovered to be read-through transcripts from neighboring genes due to spurious Pol II transcription[169]. Furthermore, several groups identified lncRNAs or eRNAs of which only the act of transcription, but not the sequence, is necessary to enhance the transcription of a nearby gene (see Box 3)[157, 170, 171]. One way to test this is to perform two knockout strategies: full deletion of lncRNA and promoter or insertion of a poly-A signal next to the transcription start site. If only the first set-up has a phenotype, the sequence of the lncRNA has no impact on it.

After more than a decade of lncRNA research, it becomes clear that both the 'believers' and 'non-believers' have some interesting viewpoints and both might be right to some extent. Several of the detected noncoding transcripts appear to be by-products of neighboring gene regulation[157, 170]. However, these studies only looked at the potential *cis*-effects of lncRNAs and do

not rule out possible *trans* effects of the transcript. The recent discovery of the *SAMMSON* lncRNA, that might be targeted for melanoma therapy, proves on the other hand that there are lncRNAs with a role in cell biology and shows the importance of further focus on the discovery of cancer-specific lncRNAs for targeted therapy. As lncRNA research is a relatively recent field, several novel insights in lncRNA biology will still be gained in the coming years, with novel technologies as a major driver of new findings.

Box 3. *Cis*-regulation by lncRNAs: when the sequence doesn't matter.

The Lander team and the Olson team studied the mechanism of *cis*-regulation of lncRNAs on their neighboring genes and published their findings in Nature in 2016.

In the first study, neither the sequence nor the transcription of 3 *cis*-acting lncRNAs influenced the expression of the neighboring genes. It appeared that only the DNA regulatory elements in the promoter of the lncRNA locus were responsible for the modulation of the neighboring gene expression. For another lncRNA, the 5' splice site and the process of transcription, but not the lncRNA sequence, were important for its *cis*-regulatory function. From none of the lncRNA loci included in this study, the sequence mediated the effect on expression level of neighbouring genes. However, this doesn't state that the lncRNA transcripts do not have an effect other than gene regulation as this phenomenon was also detected for several protein coding genes.[156]

The Olsen group focused on *Hand2*, a regulator of heart development, and its bidirectional lncRNA *Uph*. *In vivo* experiments indicated that *Uph* transcription is necessary for embryonic survival and *Hand2* regulation, but *in vitro* knockdown of *Uph* did not alter *Hand2* expression. This again implies that the sequence of the transcript is not of importance for *cis*-regulatory effects. Furthermore, it was shown that *Uph* transcription was necessary for the H3K4me1 and H3K27ac marks of the locus by recruitment of GATA4 and for the RNAPII elongation along the *Hand2* locus.[169]

2.5 Translating noncoding RNAs to the patient

In recent years, precision oncology has rapidly emerged and gained much attention. Insights into patient tumor specific genetic lesions of expressed mRNAs opened the way for targeted therapies using immunotherapy and small molecules. Given the high tissue specificity of many lncRNAs and addiction of cancer cells to some oncogenic lncRNAs, novel opportunities for RNA based therapies using antisense technologies are emerging (ex. siRNAs, ASO and ribozymes)[172, 173]. The ASO technology (antisense oligonucleotides) is being tested for several mRNAs and miRNAs by IONIS pharmaceuticals™, with some mRNA-targeting ASOs already FDA approved. However, the delivery of the antisense molecules to the specific tissues remains challenging.

Several lncRNAs are known to be specifically expressed in a certain cancer type, without any expression in healthy tissue. One example, *PCA3* is now already used in the clinic as a non-invasive biomarker for the detection of prostate cancer as it is only present in the urine of affected patients[141]. The tissue specific expression pattern of lncRNAs also makes them interesting candidates for targeted therapy, as side effects should be minimal when the lncRNA is not expressed in healthy tissue. This is also the current mindset for a lot of lncRNA researchers, in which they want to find ectopically expressed lncRNAs for a specific disease (ex. *SAMMSON*[98]).

2.6 References

1. Harris H. Transfer of Radioactivity from Nuclear to Cytoplasmic Ribonucleic Acid. Nature. 1964;202:249-52.
2. Bryson V, Vogel HJ. Evolving Genes and Proteins. Science. 1965;147(3653):68-71.
3. Holley RW. Structure of an alanine transfer ribonucleic acid. JAMA. 1965;194(8):868-71.
4. Holley RW, Everett GA, Madison JT, Zamir A. Nucleotide Sequences in the Yeast Alanine Transfer Ribonucleic Acid. J Biol Chem. 1965;240:2122-8.
5. Nobel Prizes for Medicine, 1968. Nature. 1968;220(5165):324-5.
6. Crick F. Central dogma of molecular biology. Nature. 1970;227(5258):561-3.
7. Cech TR. Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena. Biosci Rep. 2004;24(4-5):362-85.
8. Golden BL, Kim H, Chase E. Crystal structure of a phage Twort group I ribozyme-product complex. Nat Struct Mol Biol. 2005;12(1):82-9.
9. Strobel SA. Ribonucleic general acid. Nat Chem Biol. 2005;1(1):5-6.
10. Christov CP, Gardiner TJ, Szuts D, Krude T. Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. Mol Cell Biol. 2006;26(18):6993-7004.
11. Christov CP, Trivier E, Krude T. Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. Br J Cancer. 2008;98(5):981-8.
12. Karijolich J, Yu YT. Spliceosomal snRNA modifications and their function. RNA Biol. 2010;7(2):192-204.
13. Kiss T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell. 2002;109(2):145-8.
14. Reichow SL, Hamma T, Ferre-D'Amare AR, Varani G. The structure and function of small nucleolar ribonucleoproteins. Nucleic Acids Res. 2007;35(5):1452-64.
15. Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell. 2008;31(6):785-99.
16. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 2011;471(7340):602-7.
17. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339(6121):819-23.
18. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339(6121):823-6.
19. Greider CW, Blackburn EH. A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature. 1989;337(6205):331-7.
20. Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, et al. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell. 1992;71(3):515-26.
21. Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence 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. 1992;71(3):527-42.
22. Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays. 2007;29(3):288-99.
23. Bird AP. Gene number, noise reduction and biological complexity. Trends Genet. 1995;11(3):94-100.
24. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22(9):1775-89.
25. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5):843-54.
26. Lee RC, Ambros V. An extensive class of small RNAs in Caenorhabditis elegans. Science. 2001;294(5543):862-4.
27. Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science. 2001;294(5543):858-62.
28. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of tissue-specific microRNAs from mouse. Curr Biol. 2002;12(9):735-9.
29. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell. 2003;115(7):787-98.
30. Ivanovska I, Cleary MA. Combinatorial microRNAs: working together to make a difference. Cell cycle. 2008;7(20):3137-42.
31. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014;15(8):509-24.

32. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer. 2006;6(4):259-69.
33. Karube Y, Tanaka H, Osada H, Tomida S, Tatematsu Y, Yanagisawa K, et al. Reduced expression of Dicer associated with poor prognosis in lung cancer patients. Cancer Sci. 2005;96(2):111-5.
34. Melo SA, Moutinho C, Ropero S, Calin GA, Rossi S, Spizzo R, et al. A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells. Cancer Cell. 2010;18(4):303-15.
35. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010;465(7301):1033-8.
36. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002;99(24):15524-9.
37. Bonci D, Coppola V, Musumeci M, Addario A, Giuffrida R, Memeo L, et al. The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities. Nat Med. 2008;14(11):1271-7.
38. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A. 2004;101(9):2999-3004.
39. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, et al. RAS is regulated by the let-7 microRNA family. Cell. 2005;120(5):635-47.
40. Park SM, Shell S, Radjabi AR, Schickel R, Feig C, Boyerinas B, et al. Let-7 prevents early cancer progression by suppressing expression of the embryonic gene HMGA2. Cell cycle. 2007;6(21):2585-90.
41. Sampson VB, Rong NH, Han J, Yang Q, Aris V, Soteropoulos P, et al. MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. Cancer research. 2007;67(20):9762-70.
42. Bui TV, Mendell JT. Myc: Maestro of MicroRNAs. Genes Cancer. 2010;1(6):568-75.
43. Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkeland SJ, et al. Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. Cell. 2008;132(5):875-86.
44. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, et al. A microRNA polycistron as a potential human oncogene. Nature. 2005;435(7043):828-33.
45. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. Nature. 2005;435(7043):839-43.
46. Schulte JH, Horn S, Otto T, Samans B, Heukamp LC, Eilers UC, et al. MYCN regulates oncogenic MicroRNAs in neuroblastoma. Int J Cancer. 2008;122(3):699-704.
47. Yan HL, Xue G, Mei Q, Wang YZ, Ding FX, Liu MF, et al. Repression of the miR-17-92 cluster by p53 has an important function in hypoxia-induced apoptosis. EMBO J. 2009;28(18):2719-32.
48. Mu P, Han YC, Betel D, Yao E, Squatrito M, Ogrodowski P, et al. Genetic dissection of the miR-17~92 cluster of microRNAs in Myc-induced B-cell lymphomas. Genes Dev. 2009;23(24):2806-11.
49. Mogilyansky E, Rigoutsos I. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. Cell Death Differ. 2013;20(12):1603-14.
50. Garzon R, Calin GA, Croce CM. MicroRNAs in Cancer. Annu Rev Med. 2009;60:167-79.
51. Hayes J, Peruzzi PP, Lawler S. MicroRNAs in cancer: biomarkers, functions and therapy. Trends Mol Med. 2014;20(8):460-9.
52. Jansson MD, Lund AH. MicroRNA and cancer. Mol Oncol. 2012;6(6):590-610.
53. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291(5507):1304-51.
54. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
55. Rinn JL, Euskirchen G, Bertone P, Martone R, Luscombe NM, Hartman S, et al. The transcriptional activity of human Chromosome 22. Genes Dev. 2003;17(4):529-40.
56. Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, et al. Large-scale transcriptional activity in chromosomes 21 and 22. Science. 2002;296(5569):916-9.
57. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009;458(7235):223-7.
58. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011;477(7364):295-300.
59. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448(7153):553-60.
60. Rando OJ, Chang HY. Genome-wide views of chromatin structure. Annu Rev Biochem. 2009;78:245-71.
61. van Bakel H, Nislow C, Blencowe BJ, Hughes TR. Most "dark matter" transcripts are associated with known genes. PLoS Biol. 2010;8(5):e1000371.

62. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A. 2009;106(28):11667-72.
63. Brannan CI, Dees EC, Ingram RS, Tilghman SM. The product of the H19 gene may function as an RNA. Mol Cell Biol. 1990;10(1):28-36.
64. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007;35(Web Server issue):W345-9.
65. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013;41(6):e74.
66. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics. 2011;27(13):i275-82.
67. Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene. 2003;22(39):8031-41.
68. Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science. 2011;331(6013):76-9.
69. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007;129(7):1311-23.
70. Fantes J, Ragge NK, Lynch SA, McGill NI, Collin JR, Howard-Peebles PN, et al. Mutations in SOX2 cause anophthalmia. Nat Genet. 2003;33(4):461-3.
71. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012;81:145-66.
72. Davidovich C, Cech TR. The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. RNA. 2015;21(12):2007-22.
73. Brockdorff N. Noncoding RNA and Polycomb recruitment. RNA. 2013;19(4):429-42.
74. Li W, Notani D, Rosenfeld MG. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat Rev Genet. 2016;17(4):207-23.
75. Lam MT, Li W, Rosenfeld MG, Glass CK. Enhancer RNAs and regulated transcriptional programs. Trends Biochem Sci. 2014;39(4):170-82.
76. Yu G, Yao W, Gumireddy K, Li A, Wang J, Xiao W, et al. Pseudogene PTENP1 functions as a competing endogenous RNA to suppress clear-cell renal cell carcinoma progression. Mol Cancer Ther. 2014;13(12):3086-97.
77. Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, et al. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife. 2013;2:e01749.
78. Grote P, Herrmann BG. The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. RNA Biol. 2013;10(10):1579-85.
79. Ripoche MA, Kress C, Poirier F, Dandolo L. Deletion of the H19 transcription unit reveals the existence of a putative imprinting control element. Genes Dev. 1997;11(12):1596-604.
80. Monnier P, Martinet C, Pontis J, Stancheva I, Ait-Si-Ali S, Dandolo L. H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1. Proc Natl Acad Sci U S A. 2013;110(51):20693-8.
81. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010;329(5992):689-93.
82. Li L, Liu B, Wapinski OL, Tsai MC, Qu K, Zhang J, et al. Targeted disruption of Hotair leads to homeotic transformation and gene derepression. Cell Rep. 2013;5(1):3-12.
83. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature. 2011;472(7341):120-4.
84. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell. 2010;39(6):925-38.
85. Yang L, Lin C, Liu W, Zhang J, Ohgi KA, Grinstein JD, et al. ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. Cell. 2011;147(4):773-88.
86. West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev RI, Wang PI, et al. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Mol Cell. 2014;55(5):791-802.
87. Zhang B, Arun G, Mao YS, Lazar Z, Hung G, Bhattacharjee G, et al. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. Cell Rep. 2012;2(1):111-23.
88. Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell. 2009;33(6):717-26.
89. Nakagawa S, Naganuma T, Shioi G, Hirose T. Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. J Cell Biol. 2011;193(1):31-9.

90. Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res. 2009;19(3):347-59.
91. Wong KK, Maser RS, Bachoo RM, Menon J, Carrasco DR, Gu Y, et al. Telomere dysfunction and Atm deficiency compromises organ homeostasis and accelerates ageing. Nature. 2003;421(6923):643-8.
92. Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 1997;11(2):156-66.
93. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science. 2008;322(5902):750-6.
94. Gutschner T, Diederichs S. The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol. 2012;9(6):703-19.
95. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
96. Fatima R, Akhade VS, Pal D, Rao SM. Long noncoding RNAs in development and cancer: potential biomarkers and therapeutic targets. Mol Cell Ther. 2015;3:5.
97. Pandey GK, Kanduri C. Fighting Neuroblastomas with NBAT1. Oncoscience. 2015;2(2):79-80.
98. Leucci E, Vendramin R, Spinazzi M, Laurette P, Fiers M, Wouters J, et al. Melanoma addiction to the long non-coding RNA SAMMSON. Nature. 2016;531(7595):518-22.
99. Huarte M. The emerging role of lncRNAs in cancer. Nat Med. 2015;21(11):1253-61.
100. Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, et al. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene. 2011;30(16):1956-62.
101. Iacobucci I, Sazzini M, Garagnani P, Ferrari A, Boattini A, Lonetti A, et al. A polymorphism in the chromosome 9p21 ANRIL locus is associated to Philadelphia positive acute lymphoblastic leukemia. Leuk Res. 2011;35(8):1052-9.
102. Pasmant E, Sabbagh A, Masliah-Planchon J, Ortonne N, Laurendeau I, Melin L, et al. Role of noncoding RNA ANRIL in genesis of plexiform neurofibromas in neurofibromatosis type 1. Journal of the National Cancer Institute. 2011;103(22):1713-22.
103. Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, 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. 2010;38(5):662-74.
104. Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B. Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression. PLoS Genet. 2010;6(4):e1000899.
105. Zhang EB, Kong R, Yin DD, You LH, Sun M, Han L, et al. Long noncoding RNA ANRIL indicates a poor prognosis of gastric cancer and promotes tumor growth by epigenetically silencing of miR-99a/miR-449a. Oncotarget. 2014;5(8):2276-92.
106. Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature. 2008;451(7175):202-6.
107. Xiang JF, Yin QF, Chen T, Zhang Y, Zhang XO, Wu Z, et al. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. Cell Res. 2014;24(5):513-31.
108. He X, Tan X, Wang X, Jin H, Liu L, Ma L, et al. C-Myc-activated long noncoding RNA CCAT1 promotes colon cancer cell proliferation and invasion. Tumour Biol. 2014;35(12):12181-8.
109. Ma MZ, Chu BF, Zhang Y, Weng MZ, Qin YY, Gong W, et al. Long non-coding RNA CCAT1 promotes gallbladder cancer development via negative modulation of miRNA-218-5p. Cell Death Dis. 2015;6:e1583.
110. Nissan A, Stojadinovic A, Mitrani-Rosenbaum S, Halle D, Grinbaum R, Roistacher M, et al. Colon cancer associated transcript-1: a novel RNA expressed in malignant and pre-malignant human tissues. Int J Cancer. 2012;130(7):1598-606.
111. Song H, Han LM, Gao Q, Sun Y. Long non-coding RNA CRNDE promotes tumor growth in medulloblastoma. Eur Rev Med Pharmacol Sci. 2016;20(12):2588-97.
112. Zheng J, Liu X, Wang P, Xue Y, Ma J, Qu C, et al. CRNDE Promotes Malignant Progression of Glioma by Attenuating miR-384/PIWIL4/STAT3 Axis. Mol Ther. 2016;24(7):1199-215.
113. Liu T, Zhang X, Gao S, Jing F, Yang Y, Du L, et al. Exosomal long noncoding RNA CRNDE-h as a novel serum-based biomarker for diagnosis and prognosis of colorectal cancer. Oncotarget. 2016.
114. Zheng J, Li XD, Wang P, Liu XB, Xue YX, Hu Y, et al. CRNDE affects the malignant biological characteristics of human glioma stem cells by negatively regulating miR-186. Oncotarget. 2015;6(28):25339-55.
115. Graham LD, Pedersen SK, Brown GS, Ho T, Kassir Z, Moynihan AT, et al. Colorectal Neoplasia Differentially Expressed (CRNDE), a Novel Gene with Elevated Expression in Colorectal Adenomas and Adenocarcinomas. Genes Cancer. 2011;2(8):829-40.
116. Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal. 2010;3(107):ra8.

117. Mourtada-Maarabouni M, Hedge VL, Kirkham L, Farzaneh F, Williams GT. Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). Journal of cell science. 2008;121(Pt 7):939-46.
118. Mourtada-Maarabouni M, Pickard MR, Hedge VL, Farzaneh F, Williams GT. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. Oncogene. 2009;28(2):195-208.
119. Qiao HP, Gao WS, Huo JX, Yang ZS. Long non-coding RNA GAS5 functions as a tumor suppressor in renal cell carcinoma. Asian Pac J Cancer Prev. 2013;14(2):1077-82.
120. Cao S, Liu W, Li F, Zhao W, Qin C. Decreased expression of lncRNA GAS5 predicts a poor prognosis in cervical cancer. Int J Clin Exp Pathol. 2014;7(10):6776-83.
121. Sun M, Jin FY, Xia R, Kong R, Li JH, Xu TP, et al. Decreased expression of long noncoding RNA GAS5 indicates a poor prognosis and promotes cell proliferation in gastric cancer. BMC Cancer. 2014;14:319.
122. Matouk IJ, Halle D, Gilon M, Hochberg A. The non-coding RNAs of the H19-IGF2 imprinted loci: a focus on biological roles and therapeutic potential in Lung Cancer. J Transl Med. 2015;13:113.
123. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010;464(7291):1071-6.
124. Zhang H, Cai K, Wang J, Wang X, Cheng K, Shi F, et al. MiR-7, inhibited indirectly by lincRNA HOTAIR, directly inhibits SETDB1 and reverses the EMT of breast cancer stem cells by downregulating the STAT3 pathway. Stem Cells. 2014;32(11):2858-68.
125. Liu XH, Liu ZL, Sun M, Liu J, Wang ZX, De W. The long non-coding RNA HOTAIR indicates a poor prognosis and promotes metastasis in non-small cell lung cancer. BMC Cancer. 2013;13:464.
126. Geng YJ, Xie SL, Li Q, Ma J, Wang GY. Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. J Int Med Res. 2011;39(6):2119-28.
127. Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. Cancer research. 2011;71(20):6320-6.
128. Lu L, Zhu G, Zhang C, Deng Q, Katsaros D, Mayne ST, et al. Association of large noncoding RNA HOTAIR expression and its downstream intergenic CpG island methylation with survival in breast cancer. Breast Cancer Res Treat. 2012;136(3):875-83.
129. Panzitt K, Tschernatsch MM, Guelly C, Moustafa T, Stradner M, Strohmaier HM, et al. Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. Gastroenterology. 2007;132(1):330-42.
130. Wang J, Liu X, Wu H, Ni P, Gu Z, Qiao Y, et al. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res. 2010;38(16):5366-83.
131. Xie H, Ma H, Zhou D. Plasma HULC as a promising novel biomarker for the detection of hepatocellular carcinoma. Biomed Res Int. 2013;2013:136106.
132. Zhao Y, Guo Q, Chen J, Hu J, Wang S, Sun Y. Role of long non-coding RNA HULC in cell proliferation, apoptosis and tumor metastasis of gastric cancer: a clinical and in vitro investigation. Oncol Rep. 2014;31(1):358-64.
133. Marin-Bejar O, Marchese FP, Athie A, Sanchez Y, Gonzalez J, Segura V, et al. Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. Genome Biol. 2013;14(9):R104.
134. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010;142(3):409-19.
135. Gutschner T, Hammerle M, Diederichs S. MALAT1 -- a paradigm for long noncoding RNA function in cancer. J Mol Med (Berl). 2013;91(7):791-801.
136. Zhou Y, Zhang X, Klibanski A. MEG3 noncoding RNA: a tumor suppressor. J Mol Endocrinol. 2012;48(3):R45-53.
137. Zhou Y, Zhong Y, Wang Y, Zhang X, Batista DL, Gejman R, et al. Activation of p53 by MEG3 non-coding RNA. J Biol Chem. 2007;282(34):24731-42.
138. Ying L, Huang Y, Chen H, Wang Y, Xia L, Chen Y, et al. Downregulated MEG3 activates autophagy and increases cell proliferation in bladder cancer. Mol Biosyst. 2013;9(3):407-11.
139. Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, et al. MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures. Nat Commun. 2015;6:7743.
140. Adriaens C, Standaert L, Barra J, Latil M, Verfaillie A, Kalev P, et al. p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. Nat Med. 2016;22(8):861-8.
141. Hessels D, Klein Gunnewiek JM, van Oort I, Karthaus HF, van Leenders GJ, van Balken B, et al. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. Eur Urol. 2003;44(1):8-15; discussion -6.

142. Ferreira LB, Palumbo A, de Mello KD, Sternberg C, Caetano MS, de Oliveira FL, et al. PCA3 noncoding RNA is involved in the control of prostate-cancer cell survival and modulates androgen receptor signaling. BMC Cancer. 2012;12:507.
143. Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nature biotechnology. 2011;29(8):742-9.
144. Prensner JR, Chen W, Han S, Iyer MK, Cao Q, Kothari V, et al. The long non-coding RNA PCAT-1 promotes prostate cancer cell proliferation through cMyc. Neoplasia. 2014;16(11):900-8.
145. Prensner JR, Chen W, Iyer MK, Cao Q, Ma T, Han S, et al. PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. Cancer research. 2014;74(6):1651-60.
146. Srikantan V, Zou Z, Petrovics G, Xu L, Augustus M, Davis L, et al. PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. Proc Natl Acad Sci U S A. 2000;97(22):12216-21.
147. Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, et al. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature. 2013;500(7464):598-602.
148. Hung CL, Wang LY, Yu YL, Chen HW, Srivastava S, Petrovics G, et al. A long noncoding RNA connects c-Myc to tumor metabolism. Proc Natl Acad Sci U S A. 2014;111(52):18697-702.
149. Petrovics G, Zhang W, Makarem M, Street JP, Connelly R, Sun L, et al. Elevated expression of PCGEM1, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients. Oncogene. 2004;23(2):605-11.
150. Poliseno L, Haimovic A, Christos PJ, Vega YSdMEC, Shapiro R, Pavlick A, et al. Deletion of PTENP1 pseudogene in human melanoma. J Invest Dermatol. 2011;131(12):2497-500.
151. Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, et al. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat Genet. 2013;45(11):1392-8.
152. Huang KC, Rao PH, Lau CC, Heard E, Ng SK, Brown C, et al. Relationship of XIST expression and responses of ovarian cancer to chemotherapy. Mol Cancer Ther. 2002;1(10):769-76.
153. Kawakami T, Okamoto K, Ogawa O, Okada Y. XIST unmethylated DNA fragments in male-derived plasma as a tumour marker for testicular cancer. Lancet. 2004;363(9402):40-2.
154. McDonald HL, Gascoyne RD, Horsman D, Brown CJ. Involvement of the X chromosome in non-Hodgkin lymphoma. Genes Chromosomes Cancer. 2000;28(3):246-57.
155. Yildirim E, Kirby JE, Brown DE, Mercier FE, Sadreyev RI, Scadden DT, et al. Xist RNA is a potent suppressor of hematologic cancer in mice. Cell. 2013;152(4):727-42.
156. Diederichs S. The four dimensions of noncoding RNA conservation. Trends Genet. 2014;30(4):121-3.
157. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature. 2016;539(7629):452-5.
158. Amaral PP, Leonardi T, Han N, Vire E, Gascoigne DK, Arias-Carrasco R, et al. Genomic positional conservation identifies topological anchor point (tap)RNAs linked to developmental loci. bioRxiv. 2016.
159. Johnsson P, Lipovich L, Grander D, Morris KV. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochim Biophys Acta. 2014;1840(3):1063-71.
160. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell. 2011;147(4):789-802.
161. van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, et al. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. Genome Biol. 2014;15(1):R6.
162. Carlevaro-Fita J, Rahim A, Guigo R, Vardy LA, Johnson R. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. RNA. 2016;22(6):867-82.
163. Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM. Long non-coding RNAs as a source of new peptides. Elife. 2014;3:e03523.
164. Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell. 2013;154(1):240-51.
165. Olexiouk V, Crappe J, Verbruggen S, Verhegen K, Martens L, Menschaert G. sORFs.org: a repository of small ORFs identified by ribosome profiling. Nucleic Acids Res. 2016;44(D1):D324-9.
166. Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell. 2015;160(4):595-606.
167. Szafron LM, Balcerak A, Grzybowska EA, Pienkowska-Grela B, Felisiak-Golabek A, Podgorska A, et al. The Novel Gene CRNDE Encodes a Nuclear Peptide (CRNDEP) Which Is Overexpressed in Highly Proliferating Tissues. PLoS One. 2015;10(5):e0127475.
168. Mercer TR, Clark MB, Crawford J, Brunck ME, Gerhardt DJ, Taft RJ, et al. Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nat Protoc. 2014;9(5):989-1009.
169. Ebisuya M, Yamamoto T, Nakajima M, Nishida E. Ripples from neighbouring transcription. Nat Cell Biol. 2008;10(9):1106-13.

170. Anderson KM, Anderson DM, McAnally JR, Shelton JM, Bassel-Duby R, Olson EN. Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. Nature. 2016;539(7629):433-6.
171. Paralkar VR, Taborda CC, Huang P, Yao Y, Kossenkov AV, Prasad R, et al. Unlinking an lncRNA from Its Associated cis Element. Mol Cell. 2016;62(1):104-10.
172. Parasramka MA, Maji S, Matsuda A, Yan IK, Patel T. Long non-coding RNAs as novel targets for therapy in hepatocellular carcinoma. Pharmacol Ther. 2016;161:67-78.
173. Christopher AF, Kaur RP, Kaur G, Kaur A, Gupta V, Bansal P. MicroRNA therapeutics: Discovering novel targets and developing specific therapy. Perspect Clin Res. 2016;7(2):68-74.

3 T-ALL AND THYMOCYTES: A MESSAGE OF NONCODING RNA

Annelynn Wallaert[1,2], Kaat Durinck[1,2], Tom Taghon[2,3], Pieter Van Vlierberghe[1,2], Frank Speleman[1,2]

[1] *Center for Medical Genetics, Ghent University, Ghent Belgium*
[2] *Cancer Research Institute Ghent, Ghent, Belgium*
[3] *Department of Clinical Chemistry, Microbiology and Immunology, Ghent University, Ghent, Belgium*

Adapted from Journal of Hematology and Oncology 2017, 10:66.

Abstract

In the last decade, the role for noncoding RNAs in disease was clearly established, starting with microRNAs and later expanded towards long noncoding RNAs. This was also the case for T-cell acute lymphoblastic leukemia, which is a malignant blood disorder arising from oncogenic events during normal T-cell development in the thymus. By studying the transcriptomic profile of protein coding genes, several oncogenic events leading to T-ALL could be identified. In recent years, it became apparent that several of these oncogenes function via microRNAs and long noncoding RNAs. In this review, we give a detailed overview of the studies that describe the noncoding RNAome in T-ALL oncogenesis and normal T-cell development.

3.1 Background

3.1.1 The noncoding RNAome

For decades, it was thought that only 2% of the human genome was functional given its coding potential for proteins. The remaining 98% of the genome was considered 'junk DNA'. More recently, many studies have indicated that a large portion (up to 75% or more) of the human genome is actively transcribed while not coding for proteins [1]. These so-called noncoding RNAs consist of several distinct families including microRNAs, small nuclear RNAs, PIWI-interacting RNAs and long noncoding RNAs. Interestingly, the detection of noncoding RNAs led to a solution for the G-value paradox that states that there is no correlation between the amount of coding genes and the complexity of the organism [2], while we do observe a correlation between the complexity of the organism and the ratio of the amount of noncoding genes to the total genomic DNA. This finding, indirectly, suggests that the increasing amount of noncoding RNA genes in the genome of organisms can account for their complexity [3]. In the last decade, the role of several of these noncoding RNA families has been intensively studied and key functions in both normal development and disease were determined.

MicroRNAs

The Ambros team described the identification of the first microRNA (miRNA) *lin-4* in *C. elegans* in 1993 [4]. However, only from the beginning of this century, the process of miRNA biogenesis was studied in depth showing that miRNAs play an important role in development and disease. MicroRNAs are short non-coding RNAs of approximately 22 nucleotides, that function as post-transcriptional repressors of their target genes. MiRNA biogenesis starts with transcription by RNA POLII of a primary miRNA (pri-miRNA), an RNA molecule consisting of one to several hairpin structures [5, 6]. This pri-miRNA is subsequently cleaved to one hairpin by the enzyme Drosha leading to the formation of a precursor miRNA (pre-miRNA), which is then translocated to the cytoplasm by Exportin-5 and processed by the Dicer complex to a mature double stranded miRNA [6[7-12]. In order to execute its gene regulatory function, one strand is incorporated into the 'RNA-induced silencing complex' (RISC) thus guiding RISC to its target messenger RNA (mRNA) through complementary base pairing between the 3'UTR of the target mRNA and the miRNA. In most cases, this interaction will eventually lead to mRNA degradation or inhibition of

protein translation (for further details on miRNA biogenesis we refer to a review by Ha & Kim [13]). Remarkably, it has also been shown that the miRNA interaction with mRNAs during cell cycle arrest can recruit translation activators instead of translation repressors [14].

Typically, 3'UTRs of protein coding genes harbor multiple bona fide seed sequences for different miRNAs. While the overall effect of a given miRNA on mRNA levels or translation may be modest, action of multiple miRNAs on a single 3'UTR may significantly alter the mRNA or protein level of a gene. At the same time, the nature of this regulatory process creates the possibility of time and context specific gene regulation, which, amongst others, is critical in normal development and cellular functions. As such, it is not surprising that microRNAs are implicated in various diseases, including cancer [15-17]. Indeed, miRNAs can act as oncogenes by repressing the expression of tumor suppressors in the cell. The prototypical miRNA oncogene is the *miR-17∼92* cluster (*oncomiR-1*) encompassing 6 different miRNAs that are overexpressed in several cancer entities [18]. This polycistron is directly activated by the MYC transcription factor or repressed by p53 [19-21] and controls a plethora of target genes including *PTEN*, *BIM* and *p21* (*CDKN1A*), thereby broadly impacting on the phenotype of cells [22, 23]. One of the first described tumor suppressor miRNAs is encoded by the *miR-15a/16-1* cluster and this locus is affected by recurrent 13q14 deletions in more than half of chronic lymphocytic leukemia (CLL) cases with the *BCL2* oncogene as a primary target [24, 25]. Following these landmark discoveries, many additional miRNAs have been identified to act as oncomirs or tumor suppressor miRNAs (reviewed in [16, 26, 27]). Recent studies also linked another class of small noncoding RNAs, the PIWI-interacting RNAs (piRNAs) to cancer. PiRNAs were detected to be upregulated in several cancer types, which could be linked to a poor prognosis. The specific mechanism of action of piRNAs in cancer biology should however be further investigated to find out if and how they are driving cancer development [28, 29].

Given the role for miRNAs in several cancer types and promising preclinical studies, further initiatives towards implementing miRNA-based therapies using miRNA mimics or miRNA antisense inhibitors are taken (eg. Mirna Therapeutics - www.mirnatherapeutics.com and miRagen Therapeutics - www.miragentherapeutics.com). Remaining challenges are the risk of a miRNA to act as both an oncogene or tumor suppressor depending on the cancer type, off-target effects and the bioavailability of miRNA mimics/inhibitors (reviewed in [30]). Further, miRNAs or miRNA signatures can be used for prognostic evaluation of cancer entities, as it was for example detected that miRNA expression profiles of acute myeloid leukemia (AML) patients clustered the samples in different groups that could be linked to cytogenetic risk categories [31].

Long noncoding RNAs

While the existence of certain long noncoding RNAs (lncRNAs) such as *XIST* (implicated in X-chromosome inactivation) has been known for some time [32, 33], the full recognition for lncRNAs towards functionally active molecules has only emerged more recently. Generally speaking, lncRNAs, in contrast to miRNAs, represent a functionally very heterogeneous group of RNA molecules that are defined by their length of at least 200 nucleotides and lack of protein-coding potential. While many lncRNA genes share characteristics with protein coding genes in

relation to splicing and poly-adenylation, many lncRNAs are expressed at low levels and show poor species conservation compared to protein coding genes. These characteristics have caused skepticism on the actual functional relevance of lncRNAs. On the other hand, their complex secondary and tertiary structures hint towards functional active molecules [34, 35], a notion that is also further supported by their remarkable tissue or cell type specific expression pattern. Indeed, for an increasing number of lncRNAs, the normal function and its putative implication in certain diseases has been reported, but for the vast majority of lncRNAs such functionalities remain to be discovered.

The total number of annotated lncRNAs is enormous and may exceed 100,000 transcripts [36]. LncRNAs can be classified according to their location and orientation relative to protein coding genes. LncRNAs overlapping a protein-coding gene are categorized as 'sense' or 'antisense' depending on their transcriptional orientation compared to the protein-coding gene. 'Intronic' lncRNAs are transcribed from an intron of another transcript, whereas 'intergenic' lncRNAs are located between two coding genes without any overlap. A fifth category is represented by those lncRNAs that are transcribed on the opposite strand of a protein coding gene, with the transcription start sites located less than 1 kb from each other. These lncRNAs are categorized as 'bidirectional'. (Reviewed in [37])

At present, in-depth insights into the function of specific lncRNAs are rather limited. These studies however illustrate the broad possible cellular functions of lncRNAs, with putative functions in transcription, shaping genome architecture or epigenetic regulation. Modulation (activation or repression) of transcription by lncRNAs can be either through binding and regulation of chromatin-modifying complexes (ex. PRC2 recruitment [38, 39]) or transcription factors [40] or by inhibition of the general transcription machinery [41, 42]. Also, a specific class of lncRNAs transcribed at enhancers, the so-called eRNAs, has been described [43, 44]. Post-transcriptionally, lncRNAs have been detected to aid mRNA processing and direct splicing [45, 46] and also effects on translation or mRNA degradation have been encountered 21307942. Some lncRNAs also have several binding sites for a miRNA. These lncRNAs are called competitive endogenous RNAs (ceRNAs) as they titrate miRNAs away from their conventional target mRNA [47-49]. (Review on lncRNA functions in [50])

As indicated above, long non-coding RNAs play a role in normal cell development, but also in several types of heritable diseases and cancer [51]. One of the most well-characterized and described lncRNAs in cancer is the 'metastasis-associated lung adenocarcinoma transcript 1' (*MALAT1*), a rather atypical lncRNA with a high expression and species conservation. *MALAT1* is expressed in nuclear speckles and plays a role in nuclear organization, transcription and alternative splicing. It has been shown that *MALAT1* is upregulated in several cancer types, enhancing cancer metastasis and high *MALAT1* expression is correlated with poor prognosis [52]. Another example of an oncogenic lncRNA involved in several cancer types is 'homeobox transcript antisense RNA' (*HOTAIR*). It has been shown that *HOTAIR* functions in the recruitment of the 'polycomb repressive complex 2' (PRC2) to specific loci in the genome, which leads to H3K27 trimethylation and transcriptional silencing of these loci [53].

In the last years, it was shown that a key subset of lncRNAs is expressed from enhancer sites [54]. Enhancers are loci on the genome that are bound by specific factors that modulate the transcriptional activity of a nearby gene. These loci are demarcated in the genome by specific histone modifications (e.g. H3K27ac and H3K4me1) and binding of key transcription factors (e.g. p300 and the Mediator complex). The RNA molecules expressed from these enhancers have been coined enhancer RNAs (eRNAs) and are between 50 and 2000 nucleotides in length. It is hypothesized that several eRNAs are necessary for the activity of enhancers, by recruiting transcription factors to the enhancers and aiding in the chromosomal looping to bring the enhancer bound transcription factors to the gene promoter [44, 55]. However, some concerns about the functionality of these eRNA transcripts arose, as it appears that only the act of transcription, but not the sequence, has an influence on the function of the enhancer [56, 57].

3.1.2 T-cell acute lymphoblastic leukemia

T-cell acute lymphoblastic leukemia (T-ALL) is a hematological malignancy caused by oncogenic transformation of developing thymocytes. Normal T-cell development is a strictly regulated process that occurs in the thymus. Immature thymocytes enter the thymus from the bone marrow and migrate through several thymic niches that drive specific stages of T-cell development [58-68]. During these stages, specific markers are present at the membrane of these immature thymocytes and genomic rearrangements attribute to the formation of a functional T-cell receptor, leading to a broad range of different mature T-cell types characterized by a specific T-cell receptor. During these stages of T-cell development, abnormal activation of oncogenes or inactivation of tumor suppressor genes, can lead to a differentiation arrest and an uncontrolled expansion of immature thymocytes evolving to fully transformed T-ALL lymphoblasts [69].

Different genetic lesions have been identified as driving events, marking specific subgroups of T-ALL with distinct gene expression patterns [70-74]: the *TAL*-rearranged subgroup, the *TLX1* subgroup, the *TLX3* subgroup and the *HOXA*-overexpressing subgroup. Recently, a fifth subgroup with a poor prognosis, the immature T-ALL subgroup, has been added with an early T-cell progenitor phenotype, but no single specific oncogenic driver event. This subgroup is marked by the overexpression of multiple oncogenic factors as *MEF2C*, *LMO2*, *LYL1* and/or *HHEX* in several patients. To establish a full-blown leukemia, several other oncogenic effects cooperate with these subtype specific driver events. For example, constitutive activation of the NOTCH1-signaling pathway is present in over half of all T-ALL patients, regardless of the subtype, indicating that hyperactive NOTCH1 signaling plays a central role in T-ALL biology [75]. The NOTCH signaling cascade is necessary in the early stages of T-cell development [58], but sustained NOTCH activation leads to the malignant transformation of thymocytes. For more in depth information on T-ALL, we refer to several good reviews [72, 76-78].

Noncoding RNAs have been extensively studied in leukemia and normal hematopoiesis [79-85], here we will focus on the role of miRNAs and lncRNAs in T-ALL and T-cell development.

3.2 Methodological approaches in miRNA and long noncoding RNA research

3.2.1 MicroRNAs

Analytical platforms

MiRNA expression studies have initially used RT-qPCR or microarray platforms, which enable simultaneous detection of several hundreds of miRNAs. More recently, advances in next-generation sequencing technology made it also possible to determine the expression profiles of miRNAs by means of small RNA-sequencing. A major advantage of small RNA-sequencing is that also novel miRNAs and isomiRs (miRNAs with small variations compared to a reference miRNA sequence) get detected [86, 87]. The recently published miRQC study gives a detailed overview of the strengths and weaknesses of the different miRNA detection methods and platforms [88].

In silico target gene prediction

After the identification of miRNAs of interest, their potential target mRNAs are usually identified based upon the miRNA seed sequence, a seven-nucleotide sequence mostly situated at positions 2-7 from the 5'-end that can interact through complementary basepairing with the 3'UTR of the miRNA. For this, several online tools can be used, including miRDB (mirDB.org) [89], miRanda (microRNA.org) [90], TargetScan (targetscan.org) [91] and the recently developed miSTAR (mi-star.org) [92], which also enable the identification of all miRNAs that potentially target an mRNA of interest. The disadvantage of these *in silico* prediction algorithms is that they focus on the interaction between the 5' miRNA seed sequence and the 3'UTR of the miRNA, but it has been shown that these interaction can also take place in the 5'UTR or coding sequence of the mRNA, that in only 60% of the cases the seed interactions are perfectly complementary (others contain bulged or mismatched nucleotides) and that sometimes the 3' end and not the 5' end of the miRNA is used for base pairing [93]. Furthermore, these methods do not take into account the site accessibility as other RNA binding proteins might block the miRNA binding site [94]**.**

Wet lab validation of miRNA target genes

Target prediction can also be achieved through several *in vitro* methods. High-troughput sequencing methods used for miRNA-mRNA interaction detection are for example HITS-CLIP (High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation) [95] or PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) [96], which are methods developed to identify the specific binding sites of RNA binding proteins. In miRNA research, these are specifically used to pull down the RNA that interacts with proteins from the RISC-complex, mostly AGO2. By comparing the pulled down RNA after overexpression or knockdown of a miRNA with a mock control, the exact interaction partners of a miRNA can be identified [96-98]. Next to these methods, all miRNA-mRNA interactions can be directly mapped using the CLASH (crosslinking, ligation and sequencing of hybrids) technology [93]. In this method, AGO-associated miRNA-target duplexes are ligated, resulting in a chimeric RNA

molecule that is subsequently sequenced immediately revealing the exact miRNA-mRNA interaction site [93]. While these methods obtain valuable novel information on miRNA targets, they are however labor intensive and technically challenging. An overview of other *in vitro* methods can be found in the review by Thomson et al. [94].

To validate potential miRNA-mRNA interactions, a luciferase reporter assay is the method of choice. In this assay, the 3'UTR of the target mRNA is cloned next to a reporter gene (e.g. luciferase) and a functional miRNA-mRNA interaction should result in a decrease of the reporter gene signal after overexpression of the miRNA of interest. A direct interaction between the miRNA and the 3'UTR of the target gene could then be confirmed if the decrease in signal is rescued by mutations of the miRNA binding sites.

This reporter assay has also been applied in 3'UTR library screens in order to detect possible interactions of known miRNAs with a certain gene of interest. In such 3'UTR library screens, a plasmid containing the luciferase report gene with the 3'UTR of the gene of interest and a miRNA library are transfected together in HEK293T cells [99]. Subsequent screening of the luciferase signal intensity allows for the identification of potential functional miRNA-mRNA interactions, which also need to be validated through subsequent mutagenesis assays.

In vivo studies of miRNA function in T-ALL development

NOTCH1 activating mutations are frequently detected in human T-ALL and it has been shown that *NOTCH1* serves as a potent oncogene that can drive T-ALL development in mice. To study the role of miRNAs in T-ALL *in vivo*, a NOTCH1-sensitized mouse model was used [100]. To establish this model, fetal liver cells with hematopoietic progenitor cells (HPCs) are isolated from pregnant mice. These HPCs are then transduced with *ICN1* (active NOTCH1) and a vector containing an antagomiR or premiR. After irradiation of the recipient mice, these transduced HPCs are injected by tail vein injection. The leukemia onset of these mice, compared with control mice (*ICN1* + a negative control miRNA), gives an indication of the oncogenic or tumor suppressive potential of the tested miRNA.

Junker et al. showed that the leukemia onset of this mouse model is dependent on *Dicer1*-mediated biogenesis of miRNAs [101]. When HPCs with ICN1 overexpression were injected in conditional *Dicer1* knock out mice, where *Dicer1* is inactivated when thymocytes or leukemic cells transit from the double negative to double positive ($CD4^+CD8^+$) stage, there was no leukemic onset compared with control mice that all developed leukemia in less than 100 days.

3.2.2 LncRNAs

Analytical platforms

In some early studies, lncRNAs have been investigated using dedicated microarrays [102, 103]. More recently, RNA sequencing has become the method of choice, particularly given the significant tissue and spatial specific expression of lncRNAs. RNA-sequencing detection of lncRNAs requires a higher read depth in comparison to protein coding mRNAs, given the low expression levels of most lncRNAs [104]. In addition to standard poly(A) RNA-sequencing, total RNA se-

quencing (with ribosomal RNA depletion) is the preferred sequencing technique for more exploratory studies, as it appears that several lncRNAs don't have a poly(A) tail [105]. The major advantage of RNA-sequencing is the ability to detect novel lncRNAs or different splicing variants of a known lncRNA [104, 106].

Guilt-by-association analysis

One of the major challenges in lncRNA research is the selection of candidate lncRNAs for further functional studies. Guilt-by-association analysis has been applied to detect potential pathways in which a certain lncRNA of interest is involved [107, 108]. This analysis is based on the correlation of the candidate lncRNA expression pattern in a sufficient large number of (patient) samples to the expression of all protein coding genes. Strong positive and/or negative correlations between the lncRNA and several protein coding genes could hint towards the involvement of the lncRNA in the same pathways as these protein coding genes.

In vitro studies of lncRNAs

The cellular localization of the lncRNA can be determined by means of RNA-FISH (fluorescence *in situ* hybridization) or cell fractionation. Nuclear lncRNAs are probably involved in gene regulation or splicing control, whereas cytoplasmic lncRNAs might have a plethora of other functions such as miRNA sequestration, regulation of translation or protein complex formation. To detect the interaction of the lncRNA with DNA, other RNAs or proteins, several techniques have been published. These are based on the use of biotinylated oligonucleotides complementary to the RNA of interest to pull down its associated DNA, RNA or proteins (ChIRP: chromatin isolation by RNA purification [109]; CHART: capture hybridization analysis of RNA targets [110]; RAP: RNA antisense purification [111]). On the other hand, lncRNAs that are interacting with a protein of interest can be detected by means of RIP (RNA immunoprecipitation)[112]. These technologies and many more are nicely reviewed by Chu et al. [113]. Furthermore, the change in transcriptional profiles after lncRNA knockdown could already hint towards potential roles for the lncRNAs. However, it should be noted that knockdown of lncRNAs is not always as straightforward as for protein coding genes. One major disadvantage is the nuclear location of several lncRNAs, which makes knockdown by siRNAs less efficient. The use of antisense oligonucleotides (ASOs) could be a solution for this problem as ASOs activate the RNaseH mechanism in the nucleus to cut the RNA target. The use of the CRISPR/Cas9 technology to knock out lncRNAs also imposes some obstacles, as lncRNAs might overlap with protein coding genes (sense or antisense) or with regulatory elements (ex. enhancers). CRISPRi [91, 114], using an inactivated Cas9 protein linked to a transcription repressor, could be a possible solution to inhibit the expression of the lncRNA. Here, the guide RNA is targeted to the transcription start site of the lncRNA, inhibiting its expression.

In vivo studies of lncRNAs

The lack of sequence conservation of lncRNAs between human and mice makes it very difficult to find orthologous lncRNAs for *in vivo* studies. However, for several lncRNAs the preservation of secondary structures, sequence domains or interacting proteins could be detected, as re-

viewed by Johnsson et al. [35]. One remarkable feature detected by several groups is that the promoter of lncRNAs showed a higher degree of sequence conservation than the exons and that this promoter conservation was similar to the promoter conservation of protein coding genes [104, 107, 115, 116]. This topic is also reviewed by I. Ulitsky [117]. One way to study the oncogenic potential of lncRNAs *in vivo* without the knowledge of the mouse orthologous lncRNA, is the use of xenografts by implanting human cell lines in mice. These cell lines could be modulated by means of knockdown or overexpression of the lncRNA and cancer progression could be monitored. In T-ALL, a competition assay could be used where wild type and modulated cell lines with specific fluorescent markers are mixed and consequently injected in mice. After a few weeks, the fluorescent signal ratios can then be measured by flow cytometric analysis of the blast cells [118].

3.3 Oncomirs and tumor suppressor miRNAs in T-ALL

3.3.1 T-ALL miRNA oncogenes

MiR-19b was one of the first oncogenic miRNAs described in T-ALL by the Wendel team [119]. This miRNA is part of the above-mentioned *mir-17~92* cluster. The oncogenic role of the cluster in T-ALL was strongly suggested through the finding of a new translocation t(13;14)(q32;q11) that juxtaposed the *miR-17~92* cluster to the *TCRA/D* locus thereby placing it under the immediate control of the strong TCRA/D enhancer. This translocation occurred together with a t(9;14)(q34;q11) translocation that contributes to the aberrant activation of the *NOTCH1* gene. The coexistence of these two translocations hinted towards the collaboration of *NOTCH1* and the *miR-17~92* cluster in T-ALL development. In order to define which members of the cluster effectively contributed to T-ALL formation, cytokine dependent FL5-12 lymphocytes were transduced with individual miRNAs of the cluster followed by IL-3 withdrawal. In these assays, *miR-19b* showed the strongest oncogenic capacity, which is in line with the fact that *miR-19b* shows the highest expression of all members of the *miR-17~92* cluster in human T-ALL. A *NOTCH1*-sensitized mouse model was subsequently used to confirm the oncogenic role of *miR-19b in vivo*. Finally, target prediction algorithms in combination with functional validation experiments identified different components of the PI(3)K signaling pathway as direct *miR-19b* targets, including *PP2A*, *PRKAA1*, *BIM* and *PTEN*.

A few years later, Ye et al. performed a large bio-informatics screening to point out central hubs in the T-ALL network [120]. To this end, combinations of genes and miRNAs known to be involved in T-ALL were tested with target prediction algorithms. Furthermore, possible transcription factor regulatory relationships (feed forward and feedback loops) were determined based on predicted transcription factor binding sites near T-ALL genes and miRNAs. This led to a complex network that contained 21 T-ALL genes, 21 T-ALL miRNAs and 28 transcription factors. The main hubs in this *in silico* established network contained 4 miRNAs of the *miR-17~92* cluster, again revealing an important role for this cluster in T-ALL. In addition, these authors revealed that *miR-19* could regulate NF-κB signaling through direct targeting of *CYLD*.

In 2011, a more in depth study was performed towards identifying oncogenic miRNAs targeting known tumor suppressor genes in T-ALL [121]. In this study, miRNA expression data was compared with an unbiased miRNA library screen, computational target prediction analyses and *in vivo* modeling, to identify the most promising candidates. Eventually, this resulted in the identification of a network of 5 oncogenic miRNAs (*miR-19b*, *miR-20a*, *miR-26a*, *miR-92* and *miR-223*), which shared a panel of direct tumor suppressor target genes previously implicated in T-ALL biology (*IKZF1*, *PTEN*, *BIM*, *PHF6*, *NF1* and *FBXW7*). MiRNAs with the same target genes also showed a cooperative effect on cell viability. Three of these miRNAs (*miR-19b*, *miR-20a* and *miR-92*) belong to the oncogenic *miR-17~92* cluster, whereas *miR-223* was subsequently shown to be activated by TAL1 [122, 123] and NOTCH1 [124], two important T-ALL oncogenes (discussed below), further supporting the original observations of this study.

In subsequent studies, additional miRNAs with an oncogenic role in the development of T-ALL have been reported. *MiR-128-3p* is highly expressed in T-ALL patients and has increased expression in T-ALL samples compared to healthy donor thymocytes. *MiR-128-3p* directly inhibits the expression of the tumor suppressor *PHF6* and overexpression caused accelerated leukemia onset in the *NOTCH1*-sensitized mouse model [99]. *MiR-21* is highly expressed in both murine and human T-ALL and is involved in the inhibition of apoptosis, probably by regulating *Pdcd4,* known to play a role in the apoptosis pathway by inhibition of *BCL-xL* translation [101]. Another example is *miR-142-3p*, which is upregulated in T-ALL patient samples compared to thymocytes of healthy donors and is one of the top expressed miRNAs in T-ALL [121, 125]. *MiR-142-3p* plays a role in cell proliferation through an indirect inhibition of *cAMP* (cyclic AMP) and *PKA* (protein kinase A), an inhibitor of T-cell leukemia proliferation. Furthermore, *miR-142-3p* directly targets glucocorticoid receptor alpha (*GRα*), with high *miR-142-3p* levels being involved in glucocorticoid resistance and is linked to poor prognosis [125]. *MiR-149** was detected as being upregulated in T-ALL cell lines and bone marrow of T-ALL patients in comparison to peripheral blood. This miRNA promotes cell proliferation and reduces cell apoptosis and might perform this oncogenic function by its direct targeting of *JunB* [126].

Finally, other studies made use of correlation analyses between miRNA and mRNA expression in T-ALL patient samples to detect potential novel oncomiRs. For example, the expression of *miR-590* is negatively correlated with *RB1* expression and it was found that *miR-590* plays an oncogenic role in cell proliferation and migration and invasion, by directly targeting *RB1* [127]. A second example is the negative correlation between *miR-181a* and *EGR1*, a tumor suppressor in several other cancer entities. The *miR-181a*/*EGR1* pair probably has a role in cell cycle regulation [128]. *miR-181a* is also linked to the NOTCH1 signaling pathway which is discussed further in more detail [129].

3.3.2 T-ALL tumor suppressor miRNA

The Wendel team also conducted a screening for miRNAs with a tumor suppressor function [130]. They selected abundantly expressed miRNAs in thymocytes from healthy donors that had at least a 10-fold lower expression in primary T-ALL samples. Further selection was performed by *in vitro* proliferation assays after overexpression of the miRNAs. This approach eventually led

to the identification of 5 miRNAs (*miR-29*, *miR-31*, *miR-150*, *miR-155* and *miR-200*) with tumor suppressive effects *in vitro* and *in vivo*. To identify the potential mRNA targets by which these miRNAs performed their tumor suppressive effect, predicted targets with higher expression in T-ALL patients as compared to healthy donors, were selected, in keeping with a potential oncogenic function of the targets. The known T-ALL oncogene *MYB* (for *miR-150*, *miR-155* and *miR-200*) and also, a potential new oncogene in T-ALL, *HBP1* (for *miR-29*, *miR-31, miR-155* and *miR-200*) appeared to be key targets of this tumor suppressive miRNA network. Remarkably, it was also shown that the oncogenic NOTCH1/c-MYC pathway inhibited the expression of *miR-31, miR-150* and *miR-155*.

To further evaluate the post-transcriptional regulation of the T-ALL oncogene *MYB* by miRNAs, the Speleman team performed a miRNA library screen testing the putative interaction of 470 miRNAs with the 3'UTR of *MYB* by a luciferase reporter assay. Combined with mRNA and miRNA expression profiling data from 64 T-ALL patient samples, *miR-193b-3p* was detected as a direct negative regulator of *MYB*. *MiR-193b-3p* was also lower expressed in TAL-rearranged T-ALL patients, in keeping with *MYB* upregulation in this T-ALL genetic subtype. Importantly, inhibition of *miR-193b-3p* in the *NOTCH1*-sensitized mouse model significantly increased leukemia onset [131].

In another study, *miR-204* was detected as a potential tumor suppressive miRNA as it was lower expressed in T-ALL patient samples compared to normal T-cells from peripheral blood. This was further supported by the observation that *miR-204* could inhibit proliferation, migration and invasion of T-ALL cell lines and directly targets SOX4, a protein involved in tumorigenesis of AML [132].

3.3.3 T-ALL subtype specific miRNAs

As indicated above, T-ALL samples can be classified in different genetic subtypes, which display unique gene expression signatures [70-74]. Although different studies have linked miRNAs to specific genetic subtypes of human T-ALL, a comprehensive study on the expression of subtype specific miRNAs in human T-ALL remains to be accomplished.

One of the first papers that described miRNAs in T-ALL linked high expression of the *miR-17~92* cluster to *TLX1*, *TLX3* and *NKX2-5* overexpressing T-ALL primary samples and cell lines. This miRNA cluster seems to be activated by these transcription factors and imposes increased cell survival through the inhibition of *E2F1* [133].

Schotte et al. linked *miR-196b* to the HOXA-overexpressing subtype with MLL-rearrangements, *CALM-AF10* or *SET-NUP214* fusions or inversion on chromosome 7 [134]. Since *miR-196b* is located in the *HOXA*-locus, this link might be due to co-activation. High expression of *miR-196a* and *miR-196b* was subsequently also linked to T-ALL samples with an early immunophenotype and concomitant expression of CD34 and CD33 [135].

Furthermore, *miR-223* has been linked to a myeloid-like T-ALL phenotype[136], but has also been identified as a target of the TAL1 transcription factor oncogene [122, 123]. Moreover, high expression of *miR-221* and *miR-222* has been linked to the poor prognostic subtype of hu-

man ETP-ALL, and it was discovered that miR-222 directly inhibits the expression of the proto-oncogene *ETS1* [137]. In the same study, *miR-19a* and *miR-363* were detected as specifically downregulated in ETP-ALL.

3.3.4 miRNAs in the NOTCH1 regulatory network

As mentioned in the introduction, *NOTCH1*-activating mutations are present in over half of all T-ALL patients. A plethora of canonical NOTCH1 downstream protein coding targets have been described over the last decade. More recently, it became apparent that also miRNAs play a role in the NOTCH1 regulatory network in the context of T-ALL development.

Li et al. described *miR-451* and *miR-709* as possible tumor suppressor miRNAs in murine T-ALL. These miRNAs are downregulated in T-ALL and show a dynamic expression pattern during normal T-cell development. The tumor suppressor role of these miRNAs was further established by a delayed leukemia onset after overexpression of *miR-451* or *miR-709* in the *NOTCH1*-sensitized mouse model. *MiR-451* and *miR-709* directly target *c-Myc*, a known oncogene activated by NOTCH1 in T-ALL. Next to *c-MYC*, *miR-709* also directly targets *Ras-GRF1* and *Akt*. Motif analysis followed by ChIP-sequencing revealed the positive regulation of these miRNAs by *E2A*, which itself is inhibited by NOTCH1 signaling. The NOTCH1/miR-451/c-MYC axis also plays a role in human T-ALL (*miR-709* has no human homologue) [138].

Later, this network was further expanded by adding a feed forward loop between *NOTCH1* and *c-MYC* that was regulated by the tumor suppressive miRNA, *miR-30a*[139]. The expression of *miR-30a* is lower in T-ALL patient samples with hyperactive NOTCH1 compared to *NOTCH1* wild type cases. NOTCH1 signaling activates the expression of *c-MYC* and *c-MYC* inhibits *miR-30a* expression [140]. Target prediction analysis and reporter assays then demonstrated that *miR-30a* targets *NOTCH1*. This implies that oncogenic activation of *NOTCH1* leads to an overexpression of *c-MYC*, followed by a *miR-30a* downregulation. This then releases the inhibition of *NOTCH1* expression by *miR-30a* [139].

It has also been shown that NOTCH1-induced murine T-ALL development was hampered by the deletion of the *miR-181a-1/b-1* gene. Remarkably, the effects of *miR-181a-1/b-1* change depending on the expression level of *Notch1*. If the expression of *Notch1* is high, the deletion of the *miR-181a-1/b-1* gene strongly delays T-ALL development, whereas the deletion leads to a full inhibition of T-ALL if *Notch1* expression is lower. *miR-181a* regulates Notch signaling by inhibition of *Nrarp*, which is a negative regulator of the NOTCH1 downstream signaling. Furthermore, *miR-181a* was also necessary in early T-cell development, where it inhibits negative regulators of pre-T-cell receptor signaling (ex. *Dusp5* and *Dusp6*) [129].

MiR-223 was detected as differentially expressed in murine Notch-modulated T-ALL models. Motif analysis and ChIP-sequencing showed the binding of the ICN1-complex and NF-κB to the promoter of *miR-223*, which leads to the activation of transcription of this miRNA. *MiR-223* itself, further negatively regulates *FBXW7*, a known tumor suppressor gene in T-ALL. In contrast to this finding, γ-secretase inhibitor (GSI) treatment (which inhibits downstream NOTCH1-signaling) showed upregulation of miR-223 in GSI-resistant T-ALL cell lines [124]. These contra-

dictory results could later be explained by the activation of C/EBPα after GSI treatment, which can activate *miR-223* as well [141]. *MiR-223* is also important in the TAL1 downstream pathway, which will be discussed in the next paragraph.

3.3.5 miRNAs up- and downstream of the TAL1 oncogene

TAL1/SCL overexpression is one of the major oncogenic events in T-ALL, which could delineate a specific T-ALL subtype. Mansour et al. studied the downstream miRNAs of TAL1 [122]. In this study *miR-223* was the most promising candidate as it was most strongly differentially expressed upon *TAL1*-knockdown and direct binding of *TAL1* to the *miR-223* promoter was shown. Next to that, *TAL1*-positive T-ALL cells needed *miR-223* for their sustained cell survival. They also showed that the expression of *TAL1* and *miR-223* is strongly correlated during normal T-cell development, implicating that the expression of *miR-223* is high in early T-cell progenitors and low from the DN3a-stage on to more mature T-cell stages. Furthermore, it was proven that *miR-223* directly inhibits the expression of *FBXW7* and in this way supports the oncogenic function of *TAL1*[121, 122]. By means of TAL1-overexpression, Correia et al. also showed the direct activation of *miR-223*, but also direct repression of *miR-146b-5p* by TAL1. Direct or indirect TAL1-regulated miRNAs were predicted (by *in silico* analysis) to target several genes in the *TAL1* downstream pathways [123].

Because several T-ALL patients show *TAL1*-overexpression without a known cause, Correia and colleagues hypothesized that the downregulation of miRNAs that target *TAL1* might be a novel oncogenic event in T-ALL. Target prediction algorithms revealed several miRNAs with potential binding sites in the 3'UTR of *TAL1*. Five of these miRNAs (*miR-101*, *miR-520d-5p*, *miR-140-5p*, *miR-448* and *miR-485-5p*) could be validated as direct inhibitors of *TAL1*, of which 4 miRNAs (not miR-520d-5p) where lower expressed in T-ALL patient samples compared to normal bone marrow cells [142] .

An overview of the described miRNA-mRNA interaction can be found in Figure 9 and in Table 3.

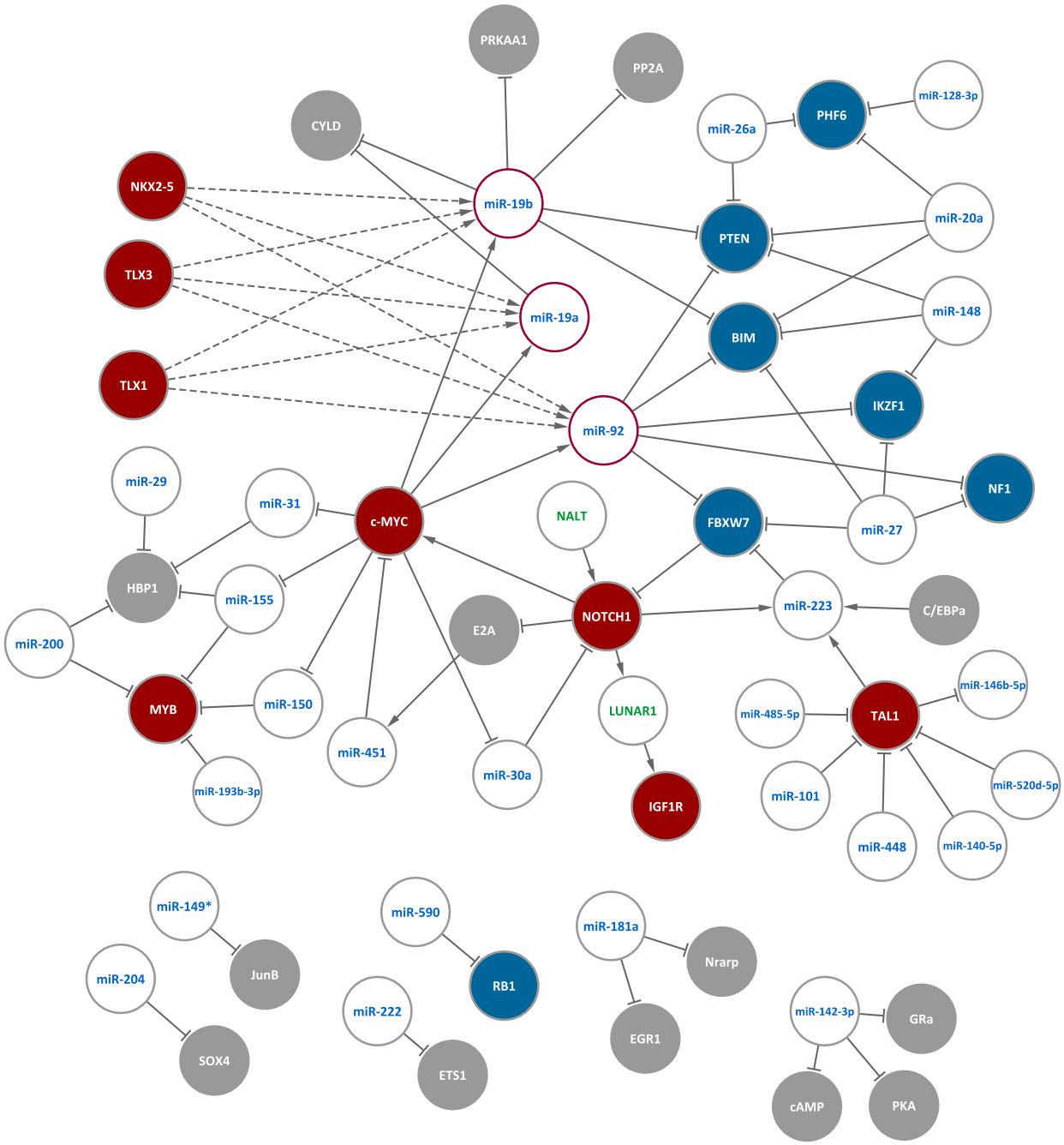


Figure 9. Overview of the noncoding RNAs implicated in T-ALL. miRNAs (blue text) and lncRNAs (green text) studied in T-ALL oncogenesis. The mRNAs linked to the ncRNAs are annotated in the filled circles. Bonafide oncogenes and tumorsuppressor genes in T-ALL are annotated in respectively red and blue background. Dashed lines represent indirect interactions. miRNAs of the miR-17~92 cluster are highlighted with a red circle.

Table 3. miRNAs implicated in T-ALL biology. ONC: oncogenic miRNA, TSG: tumor suppressor miRNA, ETP: ETP-ALL, TAL-R: TAL-rearranged T-ALL, IMM: Immature T-ALL, HOXA: HOXA-overexpressing T-ALL. The most recent miRBase annotation was retrieved using the miRBase Tracker, www.mirbasetracker.org [143].

	miRBase release 21	Function	Direct targets	Refs
miR-19a	hsa-miR-19a-3p	ONC ETP low	CYLD	[120, 137]
miR-19b	hsa-miR-19b-3p	ONC	BIM, CYLD, PP2A, PRKAA1, PTEN	[119-121]
miR-20a	hsa-miR-20a-5p	ONC	BIM, PHF6, PTEN	[121]
miR-21	hsa-miR-21-5p	ONC	PDCD4	[101]
miR-26a	hsa-miR-26a-5p	ONC	BIM, PHF6, PTEN	[121]
miR-29	hsa-miR-29a-3p	TSG	HBP1	[129]
miR-30a	hsa-miR-30a-5p hsa-miR-30a-3p	MYC repressed Targets NOTCH1	NOTCH1	[139, 140]
miR-31	hsa-miR-31-5p	TSG	HBP1	[130]
miR-92	hsa-miR-92a-3p	ONC	BIM, FBXW7, IKZF1, NF1, PTEN	[121]
miR-101	hsa-miR-101-3p	Targets TAL1	TAL1	[142]
miR-128-3p	hsa-miR-128-3p	ONC	PHF6	[99]
miR-140-5p	hsa-miR-140-5p	Targets TAL1	TAL1	[142]
miR-142-3p	hsa-miR-142-3p	ONC	cAMP, GRα, PKA	[125]
miR-146b-5p	hsa-miR-146b-5p	TAL1 repressed		[123]
miR-149*	hsa-miR-149-3p	ONC	JunB	[126]
miR-150	hsa-miR-150-5p	TSG	MYB	[130]
miR-155	hsa-miR-155-5p	TSG	HBP1, MYB	[130]
miR-181a	hsa-miR-181a-5p	ONC	EGR1, NRARP	[128]
miR-193b-3p	hsa-miR-193b-3p	TSG TAL-R low	MYB	[131]
miR-196a	hsa-miR-196a-5p	IMM high	ERG	[135]
miR-196b	hsa-miR-196b-5p	HOXA high IMM high	ERG	[134, 135]
miR-200c	hsa-miR-200c-3p	TSG	HBP1, MYB	[130]
miR-204	hsa-miR-204-5p	TSG	SOX4	[132]
miR-221	hsa-miR-221-3p	ETP high		[137]
miR-222	hsa-miR-222-3p	ETP high	ETS1	[137]
miR-223	hsa-miR-223-3p	ONC Myeloid high TAL-R high NOTCH1 activated TAL1 activated	FBXW7	[121-124, 136]
miR-363	hsa-miR-363-3p	ETP low		[137]
miR-448	hsa-miR-448	Targets TAL1	TAL1	[142]
miR-451	hsa-miR-451a	NOTCH1 repressed	c-MYC	[138]
miR-485-5p	hsa-miR-485-5p	Targets TAL1	TAL1	[142]
miR-520d-5p	hsa-miR-520d-5p	Targets TAL1	TAL1	[142]
miR-590	hsa-miR-590-5p	ONC	RB1	[127]
miR-92	hsa-miR-92a-3p	ONC	BIM, FBXW7, IKZF1, NF1, PTEN	[121]

3.4 Long non-coding RNAs implicated in T-ALL

In contrast to miRNAs, lncRNAs emerged more recently on the cancer scene and fewer studies have been published so far. Moreover, the possible functions of lncRNAs are most probably very diverse as exemplified by those described so far in the lncRNA field. Moreover, modulating lncRNAs and identifying their function can be notoriously difficult and requires extensive investigations.

3.4.1 NOTCH1 driven lncRNAs

The first comprehensive study of lncRNAs in T-ALL comprised mRNA and lncRNA expression profiles of T-ALL cell lines and primary T-ALL patient samples by means of deep total RNA-sequencing. Direct NOTCH1-regulated lncRNAs were determined by pharmacological inhibition of the NOTCH1 pathway by means of GSIs in two T-ALL cell lines and by NOTCH/RBPJκ ChIP-sequencing. Trimarchi et al. prioritized *LUNAR1* (*leukemia induced noncoding activator RNA 1*) as a NOTCH1 induced candidate oncogenic lncRNA for further functional analysis. This was based on its strong correlated expression with *IGF1R*, as *IGF1R* was already previously linked to T-ALL development. In addition, the *LUNAR1* locus is characterized by an active promoter based on the chromatin structure as determined in cell lines with hyperactive NOTCH1-signalling and the transcript structure of *LUNAR1*, as determined by 'Rapid Amplification of cDNA Ends' (RACE), has no protein coding potential. Hi-C and 3C (chromosome conformation capture) proved a physical interaction between the *LUNAR1* promoter and an active enhancer in the last intron of its neighboring gene *IGF1R*. Also, knockdown of *LUNAR1* led to a decrease in expression of *IGF1R*, whereas overexpression of *LUNAR1* did not have any effect on *IGF1R*, in keeping with a *cis*-acting role of *LUNAR1*. Next, in depth *in vitro* and *in vivo* experiments could unravel the mechanism by which *LUNAR1* has an oncogenic role in T-ALL. Xenograft assays with a mix of human T-ALL cells with or without knockdown of *LUNAR1* revealed tumors with a significant loss of representation of cells where *LUNAR1* was depleted, again proving an oncogenic role of *LUNAR1* in T-ALL development. On a molecular level, Trimarchi and colleagues could show that *LUNAR1* is involved in the recruitment of the Mediator complex and RNA Pol II to the enhancer located in the last intron of *IGF1R*, leading to full transcriptional activation of the *IGF1R* gene [118].

In a parallel study, the repertoire of NOTCH1-driven lncRNAs in T-ALL was further unraveled by Durinck et al., through characterization of lncRNAs of which the expression was affected by GSI treatment of T-ALL cell lines and under control of NOTCH-signaling in $CD34^+$ thymocytes [102]. By means of RNA-sequencing a set of known and novel lncRNAs that are directly regulated by NOTCH1 in both normal and malignant T-cell development were identified, with one of the most prominent NOTCH1 candidate lncRNAs apparent from both *in vitro* model systems being the previously described *LUNAR1*. Integration of the obtained RNA-seq profiles of GSI-treated cell lines and NOTCH1 stimulated CD34+ T-cell progenitors with NOTCH1 ChIP-sequencing profiles showed that the majority of the identified NOTCH1-regulated lncRNAs showed ICN1 binding in the vicinity of their promoter. In addition, a subset of those was also bound by MED1 and

BRD4, hinting towards a potential role of enhancer RNAs for a subset of the identified NOTCH1 regulated lncRNAs [102].

In addition to the above studies focusing on NOTCH1 controlled lncRNAs, yet another investigation identified *NALT* (*Notch1 associated lncRNA in T-ALL*) as a lncRNA involved in the regulation of *NOTCH1* expression. It is located 400 bp upstream of the *NOTCH1* locus in the antisense direction and is higher expressed in T-ALL patient bone marrow compared to healthy control samples. *In vitro* and *in vivo* knockdown experiments could further show a potential role for *NALT* as a transcriptional activator involved in cell proliferation [144].

3.4.2 T-ALL subtype specific lncRNAs

As indicated above, gene expression studies have been shown to allow genetic subgroup classification. To explore this for lncRNA expression profiles, the Speleman team screened a cohort of 64 primary T-ALL patient samples for expression of all protein coding genes and 13,000 lncRNAs [145]. This cohort consisted of 15 immature, 17 *TLX1/3*, 25 *TAL*-rearranged and 7 *HOXA* overexpressing T-ALL cases. This study allowed defining subsets of lncRNAs specific for each of the T-ALL genetic subtypes. Furthermore, the authors linked the lncRNA expression pattern in these T-ALL subtypes to the different stages of healthy T-cell development in the thymus. As the immature T-ALL subtype lymphoblasts occur from a differentiation arrest early during T-cell development ($CD34^+$ thymocytes), it appeared that several lncRNAs that are upregulated in the immature T-ALL subtype are also higher expressed in the $CD34^+$ thymocytes compared to later stages during T-cell development. These lncRNAs might be involved in normal T-cell development. On the other hand, lncRNAs were identified in the immature T-ALL subtype group with significantly higher expression in immature T-ALL as compared to $CD34^+$ thymocytes, revealing a potential oncogenic role during T-ALL development. The same comparisons could be made for the *TAL*-rearranged patients that resemble a later differentiation arrest during T-cell development, the double positive $CD4^+CD8^+$ stage [103].

3.5 Noncoding RNAs in T-cell development

Normal thymopoiesis is a tightly regulated developmental process that is initiated with $CD34^+$ early T-cell progenitors that migrate from the bone marrow towards the thymus. Within this thymic microenvironment, discrete developmental stages of T cell development can be identified through a combination of cell surface markers (CD34, CD4, CD8, CD3 etc.) and each of these stages contains a distinct transcriptional profile [58-68] (Figure 10). As noncoding RNAs show a very tissue and cell type specific expression pattern, the possible involvement of miRNAs and lncRNAs in the clearly distinct steps of T-cell development is quite obvious.

3.5.1 MiRNAs in T-cell development

Several miRNAs have been discovered over the last years that are involved in normal T-cell development in the thymus. Their overall relevance was nicely illustrated by a study by Cobb et al. that showed that deletion of *Dicer* in early T-cell progenitors in mice led to a decrease in thymic cellularity as a consequence of reduced survival of the αβ T-cell lineage, thus revealing a role

for miRNAs in the double negative to double positive stage transition [146, 147]. In addition, miRNA processing by Dicer is also necessary for the positive selection of thymocytes and the transition from the double positive to the $CD8^+$ single positive stage, as shown in conditional *Dicer* knock-out mice with a CD4-Cre transgene [148]. It has also been shown that miRNAs and even isomiRs change in expression during T-cell development, indicating that not only the expression, but also the processing of the miRNAs is altered during thymopoiesis [149].

Despite their clear importance, data on the role of individual miRNAs is rare. *MiR-181a*, however, not only plays a role in *NOTCH1*-driven T-ALL, but also appears to be involved during normal T-cell development. The expression of *miR-181a* is high at the double positive T-cell stage and decreases during development, with almost no expression in differentiated T-cells [150, 151]. Furthermore, it has been shown that *miR-181a* increases thymocyte sensitivity by directly inhibiting the expression of *DUSP5*, *DUSP6*, *SHP2* and *PTPN22* which are negative regulators of TCR-signaling [152]. *miR-181a* also appears to be involved in the regulation of positive and negative selection of thymocytes [151, 152]. (Figure 10)

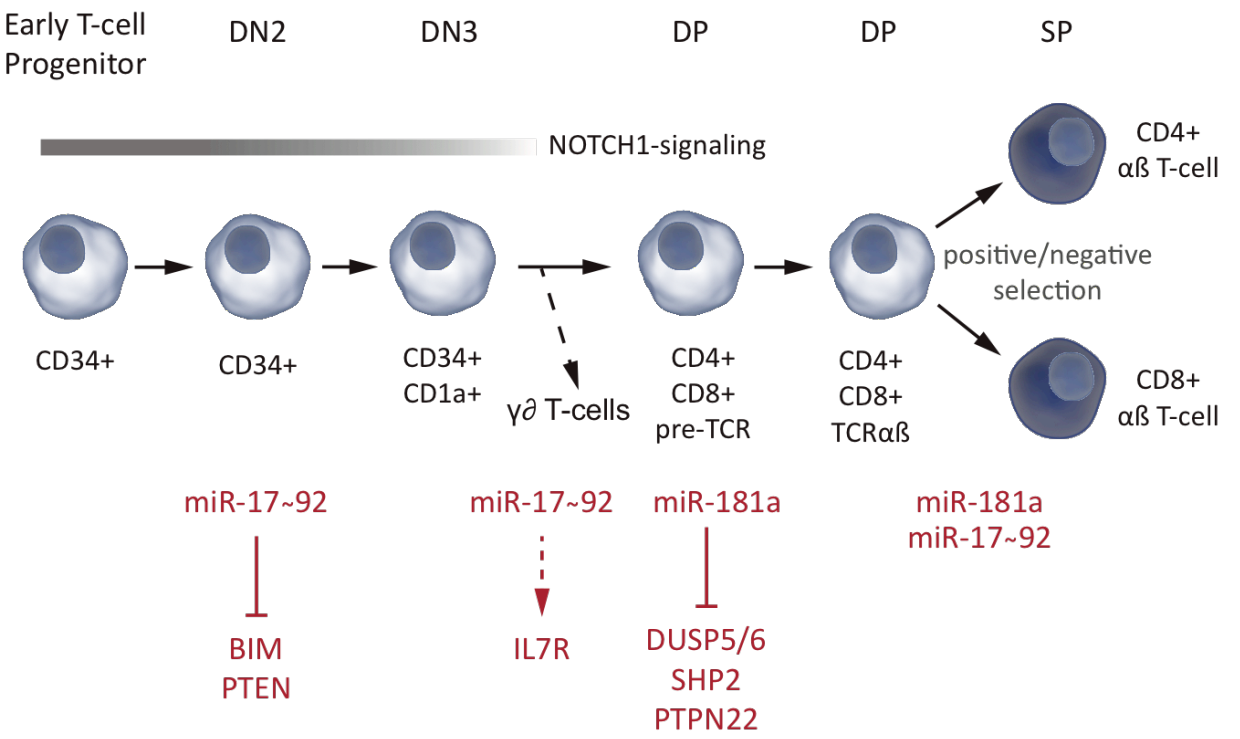


Figure 10. MicroRNAs involved in T-cell development in the thymus. Immature T-cells migrate from the bone marrow to the thymus where they go through several stages of differentiation (early T-cell prognitors (ETP), double negative T-cells (DN), double positive T-cells (DP) and single positive T-cells (SP)), which are marked by different membrane receptors (CD34, CD1a, CD4, CD8, TCR etc). Mature T-cells leave the thymus as either CD4+ or CD8+ αβ T-cells or γδ T-cells and perform several function in the immune defense of the body. MicroRNAs play a role during this process, with proven function for the miR-17~92 cluster and miR-181a.

The *miR-17~92* cluster inhibits the expression of *PTEN* and of the pro-apoptotic protein BIM, which leads to T-cell survival at the DN2 stage of T-cell development [153]. It has also been shown that the *miR-17~92* cluster is necessary for cell survival at the double negative to double positive transition of T cell development by regulating the IL7R receptor surface expression and the response to IL-7 [154]. Furthermore, this cluster is also involved in positive and negative selection of thymocytes [153]. Recently it has been shown that the expression of the *miR-17~92* cluster is regulated by TCR-signaling and, in this way, indirectly by *miR-181a*. The expression of *miR-17~92* can inhibit CD69 expression, which is also activated by TCR-signaling. With this feed forward loop, cell-to-cell variation in the thymocytes is regulated. This further marks the importance of miRNAs during normal T-cell development [155]. (Figure 10)

3.5.2 LncRNAs in T-cell development

As lncRNAs are known to be expressed tissue specifically, it should be no surprise that also lncRNAs are involved in this specific developmental process. However, not much is known about lncRNAs involved in human thymopoiesis. Several studies profiled either sorted thymocytes from mice or differentiation stages of mature T-cells [156, 157], thereby already revealing fluctuations in lncRNA expression during T-cell development.

The T-ALL oncogene *NOTCH1* is also necessary for T-cell lineage commitment in the first stages of T-cell development. NOTCH1 signaling is high in the $CD34^+$ thymocytes, but drops significantly during the β-selection process when the cells differentiate towards $CD4^+CD8^+$ double positive thymocytes. In addition to the role of NOTCH1 in lncRNA expression in T-ALL (see above), Durinck et al. also examined lncRNAs in T-cell development[102]. Human thymic $CD34^+$ progenitor T-cells were plated on an OP9 stromal cell layer that expresses the NOTCH-ligand DLL-1 (delta-like ligand 1)[61], leading to activation of NOTCH1 signaling. RNA-sequencing was performed after 48 hours of co-culture and showed a clear shift in lncRNA expression. Furthermore, *ex vivo* purified human thymocyte subsets ($CD34^+CD1^-CD4^+$, $CD34^+CD1^+CD4^+$, $CD4^+CD8^+CD3^-$ and $CD4^+CD8^+CD3^+$) were profiled on an expression array. The NOTCH-regulated lncRNAs selected from the co-culture experiment clearly followed the expression pattern of *DTX1*, a protein coding NOTCH1 target gene that is expressed in $CD34^+$ thymocytes but not in $CD4^+CD8^+$ double positive T-cells. With this study, the importance of NOTCH1 in the regulation of lncRNA expression during T-cell development was clearly shown. Expression profiling of these T-cell subsets also revealed several other lncRNAs with dynamic expression patterns during human T cell development, suggesting that these also have specific roles during the T-cell maturation process [103].

To elucidate the molecular mechanisms that control early hematopoietic lineage choices in human, Casero et al. performed RNA-sequencing on several stages of B- and T-cell development [158]. Also here, stage-specific patterns of lncRNA expression were identified during the different stages of T-cell development. Remarkably, cell-type specific lncRNAs, and not highly expressed lncRNAs, were characterized by high densities of H3K4me1 and H3K4me3 histone modifications (marks for respectively active enhancers and promoters). Another interesting dif-

ference between lncRNAs and protein coding genes was detected if the samples were clustered based on differentially expressed genes. For protein coding genes, the $CD34^+$ thymic progenitor cells segregated with the $CD34^+$ populations in the bone marrow and not with more mature ($CD34^-$) thymic progenitor cells. However, lncRNAs clearly made the distinction between thymic cells and bone marrow derived cells. With these data, the authors could show that the cell type-specific nature of lncRNA expression could be used to define developmental relationships.

Despite the low amount of studies describing the role of miRNAs and lncRNAs in early T-cell development, they already suggest that noncoding RNAs complement protein coding genes in their ability to guide early T-cell progenitors through the different maturation stages.

3.6 Conclusions and future perspectives

The role for miRNAs and long noncoding RNAs has been described in several cancer entities and in developmental processes. However, it remains a challenge to define the functional activities of these noncoding RNAs, especially for long noncoding RNAs since their potential mechanism of action can be very broad. Nevertheless, the oncogenic roles for several miRNAs (ex. the *miR-17~92* cluster [23]) and lncRNAs (ex. *MALAT1*, *HOTAIR* … [52, 53]) have been described in detail for several cancer entities.

In T-ALL, the role for miRNAs is already explored in depth. One landmark publication by the Wendel team could link several miRNAs to protein coding genes with a known tumor suppressive role in T-ALL, also showing the cooperative effect of several miRNAs on the same mRNA [121]. This paved the way for several other studies that could expand this miRNA-mRNA network. Also in T-cell development, there seems to be a role for miRNAs, however, more in depth studies should be performed to profile the miRNAs that have key roles during these developmental steps.

The role for lncRNAs in T-ALL and T-cell development is less established in comparison for miRNAs. Nevertheless, the discovery of *LUNAR1*, a NOTCH1-activated lncRNA that regulates the expression of IGF1R in T-ALL [118], proves that there are lncRNAs involved in the oncogenic development of T-ALL. Furthermore, several studies identified lncRNAs with a specific expression pattern in T-ALL and T-cell development, but the functional mechanisms of these lncRNAs have not been discovered. This is partly due to the lack of species conservation of lncRNAs, which makes it difficult to study them in mouse models, but also because a detailed study describing an in-depth and full transcriptome of all discrete stages of human T cell development is still missing. Another obstacle is the broad range of possible functional mechanisms that lncRNAs could have, which is not the case for miRNAs, and the lack of robust genetic tools in human primary hematopoietic precursors cells to functionally study the role of individual lncRNAs.

As more and more functions for miRNAs and lncRNAs are discovered, several possible pharmacological inhibitory mechanisms, for example the usage of anti-sense oligonucleotides, are being tested to target these noncoding RNAs [159, 160]. The interesting feature of noncoding RNAs is that their expression is more tissue specific than most protein coding oncogenes. Drugs

targeting these tissue specific RNAs could then result in less off-target effects of the therapy. Because of this, noncoding RNA research with a focus on these ectopic expressed noncoding RNAs should be further established, taken into account that there should also be a possibility to identify patients in clinic that could benefit from these specific treatments.

3.7 References

1. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al: Landscape of transcription in human cells. *Nature* 2012, 489:101-108.
2. Hahn MW, Wray GA: The g-value paradox. *Evol Dev* 2002, 4:73-75.
3. Taft RJ, Pheasant M, Mattick JS: The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 2007, 29:288-299.
4. Lee RC, Feinbaum RL, Ambros V: The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 1993, 75:843-854.
5. Cai X, Hagedorn CH, Cullen BR: Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* 2004, 10:1957-1966.
6. Lee Y, Jeon K, Lee JT, Kim S, Kim VN: MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* 2002, 21:4663-4670.
7. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN: The nuclear RNase III Drosha initiates microRNA processing. *Nature* 2003, 425:415-419.
8. Yi R, Qin Y, Macara IG, Cullen BR: Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* 2003, 17:3011-3016.
9. Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U: Nuclear export of microRNA precursors. *Science* 2004, 303:95-98.
10. Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ: Processing of primary microRNAs by the Microprocessor complex. *Nature* 2004, 432:231-235.
11. Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R: The Microprocessor complex mediates the genesis of microRNAs. *Nature* 2004, 432:235-240.
12. Landthaler M, Yalcin A, Tuschl T: The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. *Curr Biol* 2004, 14:2162-2167.
13. Ha M, Kim VN: Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* 2014, 15:509-524.
14. Vasudevan S, Tong Y, Steitz JA: Switching from repression to activation: microRNAs can up-regulate translation. *Science* 2007, 318:1931-1934.
15. Jansson MD, Lund AH: MicroRNA and cancer. *Mol Oncol* 2012, 6:590-610.
16. Hayes J, Peruzzi PP, Lawler S: MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol Med* 2014, 20:460-469.
17. Lin S, Gregory RI: MicroRNA biogenesis pathways in cancer. *Nat Rev Cancer* 2015, 15:321-333.
18. Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkeland SJ, Newman J, Bronson RT, Crowley D, Stone JR, et al: Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell* 2008, 132:875-886.
19. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT: c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 2005, 435:839-843.
20. Schulte JH, Horn S, Otto T, Samans B, Heukamp LC, Eilers UC, Krause M, Astrahantseff K, Klein-Hitpass L, Buettner R, et al: MYCN regulates oncogenic MicroRNAs in neuroblastoma. *Int J Cancer* 2008, 122:699-704.
21. Yan HL, Xue G, Mei Q, Wang YZ, Ding FX, Liu MF, Lu MH, Tang Y, Yu HY, Sun SH: Repression of the miR-17-92 cluster by p53 has an important function in hypoxia-induced apoptosis. *EMBO J* 2009, 28:2719-2732.
22. Mu P, Han YC, Betel D, Yao E, Squatrito M, Ogrodowski P, de Stanchina E, D'Andrea A, Sander C, Ventura A: Genetic dissection of the miR-17~92 cluster of microRNAs in Myc-induced B-cell lymphomas. *Genes Dev* 2009, 23:2806-2811.
23. Mogilyansky E, Rigoutsos I: The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ* 2013, 20:1603-1614.
24. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, et al: Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 2002, 99:15524-15529.
25. Bonci D, Coppola V, Musumeci M, Addario A, Giuffrida R, Memeo L, D'Urso L, Pagliuca A, Biffoni M, Labbaye C, et al: The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities. *Nat Med* 2008, 14:1271-1277.
26. Garzon R, Calin GA, Croce CM: MicroRNAs in Cancer. *Annu Rev Med* 2009, 60:167-179.
27. Esquela-Kerscher A, Slack FJ: Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* 2006, 6:259-269.

28. Mei Y, Clark D, Mao L: Novel dimensions of piRNAs in cancer. *Cancer Lett* 2013, 336:46-52.
29. Moyano M, Stefani G: piRNA involvement in genome stability and human cancer. *J Hematol Oncol* 2015, 8:38.
30. Naidu S, Magee P, Garofalo M: MiRNA-based therapeutic intervention of cancer. *J Hematol Oncol* 2015, 8:68.
31. Cancer Genome Atlas Research N: Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013, 368:2059-2074.
32. Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF: 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* 1992, 71:527-542.
33. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N: Requirement for Xist in X chromosome inactivation. *Nature* 1996, 379:131-137.
34. Diederichs S: The four dimensions of noncoding RNA conservation. *Trends Genet* 2014, 30:121-123.
35. Johnsson P, Lipovich L, Grander D, Morris KV: Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta* 2014, 1840:1063-1071.
36. Volders PJ, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, Mestdagh P: An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res* 2015, 43:4363-4364.
37. Rinn JL, Chang HY: Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 2012, 81:145-166.
38. Davidovich C, Cech TR: The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. *RNA* 2015, 21:2007-2022.
39. Brockdorff N: Noncoding RNA and Polycomb recruitment. *RNA* 2013, 19:429-442.
40. Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP: Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* 2010, 3:ra8.
41. Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A: Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 2007, 445:666-670.
42. Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA: Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* 2008, 29:499-509.
43. Li W, Notani D, Rosenfeld MG: Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* 2016, 17:207-223.
44. Lam MT, Li W, Rosenfeld MG, Glass CK: Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci* 2014, 39:170-182.
45. Beltran M, Puig I, Pena C, Garcia JM, Alvarez AB, Pena R, Bonilla F, de Herreros AG: A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev* 2008, 22:756-769.
46. Munroe SH, Lazar MA: Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA. *J Biol Chem* 1991, 266:22083-22086.
47. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP: A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 2010, 465:1033-1038.
48. Yu G, Yao W, Gumireddy K, Li A, Wang J, Xiao W, Chen K, Xiao H, Li H, Tang K, et al: Pseudogene PTENP1 functions as a competing endogenous RNA to suppress clear-cell renal cell carcinoma progression. *Mol Cancer Ther* 2014, 13:3086-3097.
49. Sanchez-Mejias A, Tay Y: Competing endogenous RNA networks: tying the essential knots for cancer biology and therapeutics. *J Hematol Oncol* 2015, 8:30.
50. Geisler S, Coller J: RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* 2013, 14:699-712.
51. Maass PG, Luft FC, Bahring S: Long non-coding RNA in health and disease. *J Mol Med (Berl)* 2014, 92:337-346.
52. Gutschner T, Hammerle M, Diederichs S: MALAT1 -- a paradigm for long noncoding RNA function in cancer. *J Mol Med (Berl)* 2013, 91:791-801.
53. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, et al: Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010, 464:1071-1076.
54. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al: Widespread transcription at neuronal activity-regulated enhancers. *Nature* 2010, 465:182-187.
55. Lai F, Shiekhattar R: Enhancer RNAs: the new molecules of transcription. *Curr Opin Genet Dev* 2014, 25:38-42.

56. Paralkar VR, Taborda CC, Huang P, Yao Y, Kossenkov AV, Prasad R, Luan J, Davies JO, Hughes JR, Hardison RC, et al: Unlinking an lncRNA from Its Associated cis Element. *Mol Cell* 2016, 62:104-110.
57. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES: Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 2016, 539:452-455.
58. Taghon T, Waegemans E, Van de Walle I: Notch signaling during human T cell development. *Curr Top Microbiol Immunol* 2012, 360:75-97.
59. Dik WA, Pike-Overzet K, Weerkamp F, de Ridder D, de Haas EF, Baert MR, van der Spek P, Koster EE, Reinders MJ, van Dongen JJ, et al: New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J Exp Med* 2005, 201:1715-1723.
60. Germain RN: T-cell development and the CD4-CD8 lineage decision. *Nat Rev Immunol* 2002, 2:309-322.
61. Van de Walle I, De Smet G, De Smedt M, Vandekerckhove B, Leclercq G, Plum J, Taghon T: An early decrease in Notch activation is required for human TCR-alphabeta lineage differentiation at the expense of TCR-gammadelta T cells. *Blood* 2009, 113:2988-2998.
62. Van de Walle I, Dolens AC, Durinck K, De Mulder K, Van Loocke W, Damle S, Waegemans E, De Medts J, Velghe I, De Smedt M, et al: GATA3 induces human T-cell commitment by restraining Notch activity and repressing NK-cell fate. *Nat Commun* 2016, 7:11171.
63. Blom B, Spits H: Development of human lymphoid cells. *Annu Rev Immunol* 2006, 24:287-320.
64. Garcia-Peydro M, de Yebenes VG, Toribio ML: Sustained Notch1 signaling instructs the earliest human intrathymic precursors to adopt a gammadelta T-cell fate in fetal thymus organ culture. *Blood* 2003, 102:2444-2451.
65. Jones ME, Zhuang Y: Acquisition of a functional T cell receptor during T lymphocyte development is enforced by HEB and E2A transcription factors. *Immunity* 2007, 27:860-870.
66. Yui MA, Rothenberg EV: Developmental gene networks: a triathlon on the course to T cell identity. *Nat Rev Immunol* 2014, 14:529-545.
67. Rothenberg EV, Kueh HY, Yui MA, Zhang JA: Hematopoiesis and T-cell specification as a model developmental system. *Immunol Rev* 2016, 271:72-97.
68. Rothenberg EV, Ungerback J, Champhekar A: Forging T-Lymphocyte Identity: Intersecting Networks of Transcriptional Control. *Adv Immunol* 2016, 129:109-174.
69. Koch U, Radtke F: Mechanisms of T cell development and transformation. *Annu Rev Cell Dev Biol* 2011, 27:539-562.
70. Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC, Behm FG, Pui CH, Downing JR, Gilliland DG, et al: Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* 2002, 1:75-87.
71. Meijerink JP: Genetic rearrangements in relation to immunophenotype and outcome in T-cell acute lymphoblastic leukaemia. *Best Pract Res Clin Haematol* 2010, 23:307-318.
72. Van Vlierberghe P, Pieters R, Beverloo HB, Meijerink JP: Molecular-genetic insights in paediatric T-cell acute lymphoblastic leukaemia. *Br J Haematol* 2008, 143:153-168.
73. Soulier J, Clappier E, Cayuela JM, Regnault A, Garcia-Peydro M, Dombret H, Baruchel A, Toribio ML, Sigaux F: HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). *Blood* 2005, 106:274-286.
74. Homminga I, Pieters R, Langerak AW, de Rooi JJ, Stubbs A, Verstegen M, Vuerhard M, Buijs-Gladdines J, Kooi C, Klous P, et al: Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. *Cancer Cell* 2011, 19:484-497.
75. Weng AP, Ferrando AA, Lee W, Morris JPt, Silverman LB, Sanchez-Irizarry C, Blacklow SC, Look AT, Aster JC: Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science* 2004, 306:269-271.
76. Tzoneva G, Ferrando AA: Recent advances on NOTCH signaling in T-ALL. *Curr Top Microbiol Immunol* 2012, 360:163-182.
77. Karrman K, Johansson B: Pediatric T-Cell Acute Lymphoblastic Leukemia. *Genes Chromosomes Cancer* 2016.
78. Van Vlierberghe P, Ferrando A: The molecular basis of T cell acute lymphoblastic leukemia. *J Clin Invest* 2012, 122:3398-3406.
79. Wei S, Wang K: Long noncoding RNAs: pivotal regulators in acute myeloid leukemia. *Exp Hematol Oncol* 2015, 5:30.
80. Morlando M, Ballarino M, Fatica A: Long Non-Coding RNAs: New Players in Hematopoiesis and Leukemia. *Front Med (Lausanne)* 2015, 2:23.
81. Zhang X, Hu W: Long noncoding RNAs in hematopoiesis. *F1000Res* 2016, 5.
82. Satpathy AT, Chang HY: Long noncoding RNA in hematopoiesis and immunity. *Immunity* 2015, 42:792-804.

83. Schotte D, Pieters R, Den Boer ML: MicroRNAs in acute leukemia: from biological players to clinical contributors. *Leukemia* 2012, 26:1-12.
84. Zhao H, Wang D, Du W, Gu D, Yang R: MicroRNA and leukemia: tiny molecule, great function. *Crit Rev Oncol Hematol* 2010, 74:149-155.
85. Montagner S, Deho L, Monticelli S: MicroRNAs in hematopoietic development. *BMC Immunol* 2014, 15:14.
86. Zhang H, Yang JH, Zheng YS, Zhang P, Chen X, Wu J, Xu L, Luo XQ, Ke ZY, Zhou H, et al: Genome-wide analysis of small RNA and novel MicroRNA discovery in human acute lymphoblastic leukemia based on extensive sequencing approach. *PLoS One* 2009, 4:e6849.
87. Schotte D, Akbari Moqadam F, Lange-Turenhout EA, Chen C, van Ijcken WF, Pieters R, den Boer ML: Discovery of new microRNAs by small RNAome deep sequencing in childhood acute lymphoblastic leukemia. *Leukemia* 2011, 25:1389-1399.
88. Mestdagh P, Hartmann N, Baeriswyl L, Andreasen D, Bernard N, Chen C, Cheo D, D'Andrade P, DeMayo M, Dennis L, et al: Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat Methods* 2014, 11:809-815.
89. Wong N, Wang X: miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res* 2015, 43:D146-152.
90. Betel D, Wilson M, Gabow A, Marks DS, Sander C: The microRNA.org resource: targets and expression. *Nucleic Acids Res* 2008, 36:D149-153.
91. Agarwal V, Bell GW, Nam JW, Bartel DP: Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 2015, 4.
92. Van Peer G, De Paepe A, Stock M, Anckaert J, Volders PJ, Vandesompele J, De Baets B, Waegeman W: miSTAR: miRNA target prediction through modeling quantitative and qualitative miRNA binding site information in a stacked model structure. *Nucleic Acids Res* 2016.
93. Helwak A, Kudla G, Dudnakova T, Tollervey D: Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 2013, 153:654-665.
94. Thomson DW, Bracken CP, Goodall GJ: Experimental strategies for microRNA target identification. *Nucleic Acids Res* 2011, 39:6845-6853.
95. Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, et al: HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 2008, 456:464-469.
96. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M, et al: Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010, 141:129-141.
97. Chi SW, Zang JB, Mele A, Darnell RB: Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 2009, 460:479-486.
98. Tan LP, Seinen E, Duns G, de Jong D, Sibon OC, Poppema S, Kroesen BJ, Kok K, van den Berg A: A high throughput experimental approach to identify miRNA targets in human cells. *Nucleic Acids Res* 2009, 37:e137.
99. Mets E, Van Peer G, Van der Meulen J, Boice M, Taghon T, Goossens S, Mestdagh P, Benoit Y, De Moerloose B, Van Roy N, et al: MicroRNA-128-3p is a novel oncomiR targeting PHF6 in T-cell acute lymphoblastic leukemia. *Haematologica* 2014, 99:1326-1333.
100. Pear WS, Aster JC, Scott ML, Hasserjian RP, Soffer B, Sklar J, Baltimore D: Exclusive development of T cell neoplasms in mice transplanted with bone marrow expressing activated Notch alleles. *J Exp Med* 1996, 183:2283-2291.
101. Junker F, Chabloz A, Koch U, Radtke F: Dicer1 imparts essential survival cues in Notch-driven T-ALL via miR-21-mediated tumor suppressor Pdcd4 repression. *Blood* 2015, 126:993-1004.
102. Durinck K, Wallaert A, Van de Walle I, Van Loocke W, Volders PJ, Vanhauwaert S, Geerdens E, Benoit Y, Van Roy N, Poppe B, et al: The Notch driven long non-coding RNA repertoire in T-cell acute lymphoblastic leukemia. *Haematologica* 2014, 99:1808-1816.
103. Wallaert A, Durinck K, Van Loocke W, Van de Walle I, Matthijssens F, Volders PJ, Avila Cobos F, Rombaut D, Rondou P, Mestdagh P, et al: Long noncoding RNA signatures define oncogenic subtypes in T-cell acute lymphoblastic leukemia. *Leukemia* 2016.
104. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al: The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012, 22:1775-1789.
105. Cui P, Lin Q, Ding F, Xin C, Gong W, Zhang L, Geng J, Zhang B, Yu X, Yang J, et al: A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* 2010, 96:259-265.
106. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008, 5:621-628.

107. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, et al: Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 2009, 458:223-227.
108. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, et al: Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet* 2011, 43:621-629.
109. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY: Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell* 2011, 44:667-678.
110. Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, Borowsky ML, Kuroda MI, Kingston RE: The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci U S A* 2011, 108:20497-20502.
111. Engreitz J, Lander ES, Guttman M: RNA antisense purification (RAP) for mapping RNA interactions with chromatin. *Methods Mol Biol* 2015, 1262:183-197.
112. Gilbert C, Svejstrup JQ: RNA immunoprecipitation for determining RNA-protein associations in vivo. *Curr Protoc Mol Biol* 2006, Chapter 27:Unit 27 24.
113. Chu C, Spitale RC, Chang HY: Technologies to probe functions and mechanisms of long noncoding RNAs. *Nat Struct Mol Biol* 2015, 22:29-35.
114. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA: Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 2013, 152:1173-1183.
115. Pang KC, Frith MC, Mattick JS: Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet* 2006, 22:1-5.
116. Amaral PP, Leonardi T, Han N, Vire E, Gascoigne DK, Arias-Carrasco R, Buscher M, Zhang A, Pluchino S, Maracaja-Coutinho V, et al: Genomic positional conservation identifies topological anchor point (tap)RNAs linked to developmental loci. *bioRxiv* 2016.
117. Ulitsky I: Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet* 2016, 17:601-614.
118. Trimarchi T, Bilal E, Ntziachristos P, Fabbri G, Dalla-Favera R, Tsirigos A, Aifantis I: Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. *Cell* 2014, 158:593-606.
119. Mavrakis KJ, Wolfe AL, Oricchio E, Palomero T, de Keersmaecker K, McJunkin K, Zuber J, James T, Khan AA, Leslie CS, et al: Genome-wide RNA-mediated interference screen identifies miR-19 targets in Notch-induced T-cell acute lymphoblastic leukaemia. *Nat Cell Biol* 2010, 12:372-379.
120. Ye H, Liu X, Lv M, Wu Y, Kuang S, Gong J, Yuan P, Zhong Z, Li Q, Jia H, et al: MicroRNA and transcription factor co-regulatory network analysis reveals miR-19 inhibits CYLD in T-cell acute lymphoblastic leukemia. *Nucleic Acids Res* 2012, 40:5201-5214.
121. Mavrakis KJ, Van Der Meulen J, Wolfe AL, Liu X, Mets E, Taghon T, Khan AA, Setty M, Rondou P, Vandenberghe P, et al: A cooperative microRNA-tumor suppressor gene network in acute T-cell lymphoblastic leukemia (T-ALL). *Nat Genet* 2011, 43:673-678.
122. Mansour MR, Sanda T, Lawton LN, Li X, Kreslavsky T, Novina CD, Brand M, Gutierrez A, Kelliher MA, Jamieson CH, et al: The TAL1 complex targets the FBXW7 tumor suppressor by activating miR-223 in human T cell acute lymphoblastic leukemia. *J Exp Med* 2013, 210:1545-1557.
123. Correia NC, Durinck K, Leite AP, Ongenaert M, Rondou P, Speleman F, Enguita FJ, Barata JT: Novel TAL1 targets beyond protein-coding genes: identification of TAL1-regulated microRNAs in T-cell acute lymphoblastic leukemia. *Leukemia* 2013, 27:1603-1606.
124. Gusscott S, Kuchenbauer F, Humphries RK, Weng AP: Notch-mediated repression of miR-223 contributes to IGF1R regulation in T-ALL. *Leuk Res* 2012, 36:905-911.
125. Lv M, Zhang X, Jia H, Li D, Zhang B, Zhang H, Hong M, Jiang T, Jiang Q, Lu J, et al: An oncogenic role of miR-142-3p in human T-cell acute lymphoblastic leukemia (T-ALL) by targeting glucocorticoid receptor-alpha and cAMP/PKA pathways. *Leukemia* 2012, 26:769-777.
126. Fan SJ, Li HB, Cui G, Kong XL, Sun LL, Zhao YQ, Li YH, Zhou J: miRNA-149* promotes cell proliferation and suppresses apoptosis by mediating JunB in T-cell acute lymphoblastic leukemia. *Leuk Res* 2016, 41:62-70.
127. Miao MH, Ji XQ, Zhang H, Xu J, Zhu H, Shao XJ: miR-590 promotes cell proliferation and invasion in T-cell acute lymphoblastic leukaemia by inhibiting RB1. *Oncotarget* 2016.
128. Verduci L, Azzalin G, Gioiosa S, Carissimi C, Laudadio I, Fulci V, Macino G: microRNA-181a enhances cell proliferation in acute lymphoblastic leukemia by targeting EGR1. *Leuk Res* 2015, 39:479-485.
129. Fragoso R, Mao T, Wang S, Schaffert S, Gong X, Yue S, Luong R, Min H, Yashiro-Ohtani Y, Davis M, et al: Modulating the strength and threshold of NOTCH oncogenic signals by mir-181a-1/b-1. *PLoS Genet* 2012, 8:e1002855.

130. Sanghvi VR, Mavrakis KJ, Van der Meulen J, Boice M, Wolfe AL, Carty M, Mohan P, Rondou P, Socci ND, Benoit Y, et al: Characterization of a set of tumor suppressor microRNAs in T cell acute lymphoblastic leukemia. *Sci Signal* 2014, 7:ra111.
131. Mets E, Van der Meulen J, Van Peer G, Boice M, Mestdagh P, Van de Walle I, Lammens T, Goossens S, De Moerloose B, Benoit Y, et al: MicroRNA-193b-3p acts as a tumor suppressor by targeting the MYB oncogene in T-cell acute lymphoblastic leukemia. *Leukemia* 2015, 29:798-806.
132. Yin JJ, Liang B, Zhan XR: MicroRNA-204 inhibits cell proliferation in T-cell acute lymphoblastic leukemia by down-regulating SOX4. *Int J Clin Exp Pathol* 2015, 8:9189-9195.
133. Nagel S, Venturini L, Przybylski GK, Grabarczyk P, Schmidt CA, Meyer C, Drexler HG, Macleod RA, Scherr M: Activation of miR-17-92 by NK-like homeodomain proteins suppresses apoptosis via reduction of E2F1 in T-cell acute lymphoblastic leukemia. *Leuk Lymphoma* 2009, 50:101-108.
134. Schotte D, Lange-Turenhout EA, Stumpel DJ, Stam RW, Buijs-Gladdines JG, Meijerink JP, Pieters R, Den Boer ML: Expression of miR-196b is not exclusively MLL-driven but is especially linked to activation of HOXA genes in pediatric acute lymphoblastic leukemia. *Haematologica* 2010, 95:1675-1682.
135. Coskun E, von der Heide EK, Schlee C, Kuhnl A, Gokbuget N, Hoelzer D, Hofmann WK, Thiel E, Baldus CD: The role of microRNA-196a and microRNA-196b as ERG regulators in acute myeloid leukemia and acute T-lymphoblastic leukemia. *Leuk Res* 2011, 35:208-213.
136. Chiaretti S, Messina M, Tavolaro S, Zardo G, Elia L, Vitale A, Fatica A, Gorello P, Piciocchi A, Scappucci G, et al: Gene expression profiling identifies a subset of adult T-cell acute lymphoblastic leukemia with myeloid-like gene features and over-expression of miR-223. *Haematologica* 2010, 95:1114-1121.
137. Coskun E, Neumann M, Schlee C, Liebertz F, Heesch S, Goekbuget N, Hoelzer D, Baldus CD: MicroRNA profiling reveals aberrant microRNA expression in adult ETP-ALL and functional studies implicate a role for miR-222 in acute leukemia. *Leuk Res* 2013, 37:647-656.
138. Li X, Sanda T, Look AT, Novina CD, von Boehmer H: Repression of tumor suppressor miR-451 is essential for NOTCH1-induced oncogenesis in T-ALL. *J Exp Med* 2011, 208:663-675.
139. Ortega M, Bhatnagar H, Lin AP, Wang L, Aster JC, Sill H, Aguiar RC: A microRNA-mediated regulatory loop modulates NOTCH and MYC oncogenic signals in B- and T-cell malignancies. *Leukemia* 2015, 29:968-976.
140. Chang TC, Yu D, Lee YS, Wentzel EA, Arking DE, West KM, Dang CV, Thomas-Tikhonenko A, Mendell JT: Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat Genet* 2008, 40:43-50.
141. Kumar V, Palermo R, Talora C, Campese AF, Checquolo S, Bellavia D, Tottone L, Testa G, Miele E, Indraccolo S, et al: Notch and NF-kB signaling pathways regulate miR-223/FBXW7 axis in T-cell acute lymphoblastic leukemia. *Leukemia* 2014, 28:2324-2335.
142. Correia NC, Melao A, Povoa V, Sarmento L, Gomez de Cedron M, Malumbres M, Enguita FJ, Barata JT: microRNAs regulate TAL1 expression in T-cell acute lymphoblastic leukemia. *Oncotarget* 2016, 7:8268-8281.
143. Van Peer G, Lefever S, Anckaert J, Beckers A, Rihani A, Van Goethem A, Volders PJ, Zeka F, Ongenaert M, Mestdagh P, Vandesompele J: miRBase Tracker: keeping track of microRNA annotation changes. *Database (Oxford)* 2014, 2014.
144. Wang Y, Wu P, Lin R, Rong L, Xue Y, Fang Y: LncRNA NALT interaction with NOTCH1 promoted cell proliferation in pediatric T cell acute lymphoblastic leukemia. *Sci Rep* 2015, 5:13749.
145. Volders PJ, Helsens K, Wang X, Menten B, Martens L, Gevaert K, Vandesompele J, Mestdagh P: LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res* 2013, 41:D246-251.
146. Cobb BS, Nesterova TB, Thompson E, Hertweck A, O'Connor E, Godwin J, Wilson CB, Brockdorff N, Fisher AG, Smale ST, Merkenschlager M: T cell lineage choice and differentiation in the absence of the RNase III enzyme Dicer. *J Exp Med* 2005, 201:1367-1373.
147. Chong MM, Zhang G, Cheloufi S, Neubert TA, Hannon GJ, Littman DR: Canonical and alternate functions of the microRNA biogenesis machinery. *Genes Dev* 2010, 24:1951-1960.
148. Muljo SA, Ansel KM, Kanellopoulou C, Livingston DM, Rao A, Rajewsky K: Aberrant T cell differentiation in the absence of Dicer. *J Exp Med* 2005, 202:261-269.
149. Kirigin FF, Lindstedt K, Sellars M, Ciofani M, Low SL, Jones L, Bell F, Pauli F, Bonneau R, Myers RM, et al: Dynamic microRNA gene transcription and processing during T cell development. *J Immunol* 2012, 188:3257-3267.
150. Neilson JR, Zheng GX, Burge CB, Sharp PA: Dynamic regulation of miRNA expression in ordered stages of cellular development. *Genes Dev* 2007, 21:578-589.
151. Ebert PJ, Jiang S, Xie J, Li QJ, Davis MM: An endogenous positively selecting peptide enhances mature T cell responses and becomes an autoantigen in the absence of microRNA miR-181a. *Nat Immunol* 2009, 10:1162-1169.

152. Li QJ, Chau J, Ebert PJ, Sylvester G, Min H, Liu G, Braich R, Manoharan M, Soutschek J, Skare P, et al: miR-181a is an intrinsic modulator of T cell sensitivity and selection. *Cell* 2007, 129:147-161.
153. Xiao C, Srinivasan L, Calado DP, Patterson HC, Zhang B, Wang J, Henderson JM, Kutok JL, Rajewsky K: Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. *Nat Immunol* 2008, 9:405-414.
154. Regelin M, Blume J, Pommerencke J, Vakilzadeh R, Witzlau K, Lyszkiewicz M, Zietara N, Saran N, Schambach A, Krueger A: Responsiveness of Developing T Cells to IL-7 Signals Is Sustained by miR-17 approximately 92. *J Immunol* 2015, 195:4832-4840.
155. Blevins R, Bruno L, Carroll T, Elliott J, Marcais A, Loh C, Hertweck A, Krek A, Rajewsky N, Chen CZ, et al: microRNAs regulate cell-to-cell variability of endogenous target gene expression in developing mouse thymocytes. *PLoS Genet* 2015, 11:e1005020.
156. Hu G, Tang Q, Sharma S, Yu F, Escobar TM, Muljo SA, Zhu J, Zhao K: Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat Immunol* 2013, 14:1190-1198.
157. Ranzani V, Rossetti G, Panzeri I, Arrigoni A, Bonnal RJ, Curti S, Gruarin P, Provasi E, Sugliano E, Marconi M, et al: The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat Immunol* 2015, 16:318-325.
158. Casero D, Sandoval S, Seet CS, Scholes J, Zhu Y, Ha VL, Luong A, Parekh C, Crooks GM: Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. *Nat Immunol* 2015, 16:1282-1291.
159. Parasramka MA, Maji S, Matsuda A, Yan IK, Patel T: Long non-coding RNAs as novel targets for therapy in hepatocellular carcinoma. *Pharmacol Ther* 2016, 161:67-78.
160. Christopher AF, Kaur RP, Kaur G, Kaur A, Gupta V, Bansal P: MicroRNA therapeutics: Discovering novel targets and developing specific therapy. *Perspect Clin Res* 2016, 7:68-74.

CHAPTER 3
RESEARCH OBJECTIVES

Research Objectives

T-cell acute lymphoblastic leukemia (T-ALL) patients currently present with an overall favorable prognosis achieved through intense chemotherapy regimens. Additional challenges that are still posed today concern those patients that present with therapy resistance or relapse. In this perspective it will be crucial to further unravel the molecular basis of T-ALL biology and identify novel targets for development of innovative therapy protocols. Technological advances in the field have opened new possibilities to dissect the T-ALL transcriptome and recent findings underscore the importance of noncoding RNA molecules, such as miRNAs and lncRNAs, next to protein coding genes in various cancer entities and also T-ALL.

In this thesis, my aim was to landscape the expression of these noncoding RNAs in T-ALL to complement the previously published protein coding gene expression profiles. In this way, novel oncogenic aspects in T-ALL could be unraveled, for example when an lncRNA or miRNA is detected in a known T-ALL oncogenic pathway or when it could point at complete novel oncogenic mechanisms. This would lead to novel insight in T-ALL biology and potential novel therapeutic targets.

AIM 1: Defining the NOTCH1-regulated lncRNAome in T-ALL

In more than half of T-ALL cases activating mutations in the *NOTCH1*-oncogene have previously been identified as well as oncogenic aberrations in genes downstream of the NOTCH1-signaling cascade. This makes this pathway an interesting candidate for T-ALL therapy. The first clinical trials with γ-secretase inhibitors (GSI) that strongly reduce NOTCH1-signaling, had to be terminated due to side effects and therapy resistance. Since then, there is a quest to better define the NOTCH1-signaling pathway to find novel, T-ALL specific, targets to complement a low dose GSI treatment.

The role of protein coding genes in the NOTCH1-pathway, both in normal T-cell development as in T-ALL, has already been thoroughly described. Therefore, we now aimed to perform a detailed landscaping of the lncRNAs that are transcriptionally regulated by this pathway **(paper 1, Haematologica, 2014)**. We generated lncRNA expression profiles after GSI treatment of T-ALL cell lines, but also of $CD34^+$ T-cell subsets that encountered a NOTCH1-activating ligand. This was further complemented by lncRNA profiling of healthy donor T-cell subsets and primary T-ALL patient samples.

AIM 2: Unraveling lncRNA expression patterns in the different T-ALL subtypes

Previously published micro-array gene expression profiling data of protein coding genes revealed a specific clustering pattern of T-ALL patient samples that could be linked to specific oncogenic driver events and differentiation arrest during T-cell development at a specific developmental stage with subsequent expansion of the lymphoblast population. To also unravel T-ALL subtype specific lncRNA expression patterns, RNA of 64 T-ALL patient samples was profiled on a custom-designed micro-array platform, detecting the expression of lncRNAs next to protein coding genes **(paper 2, Leukemia, 2016)**. In this way, I was able to identify a set of lncRNAs

that can discriminate the various known T-ALL subgroups and the determination of lncRNA profiles from sorted healthy donor T-cell subsets could further support the elucidation of potential oncogenic lncRNAs amongst the T-ALL subgroups in this study. Previous studies elucidating the mRNA expression profiles of T-ALL subgroups could identify oncogenic drivers for this profiles, we thus hypothesize that some of the subtype specific lncRNAs might also have an oncogenic function in T-ALL development.

Technological advances in next-generation sequencing have pushed forward RNA-sequencing as the state-of-the-art technology to generate gene expression profiles rather than micro-array technology. Therefore we generated in this study novel expression profiles of our T-ALL patient cohort by means of poly(A) RNA-sequencing **(paper 3, in preparation)**. This led to an update of the lncRNA profiles that can differentiate the distinct T-ALL subtypes and I could show that RNA-sequencing is indeed a better option for lncRNA profiling as compared to micro-arrays, as it detects differentially expressed lncRNAs more robustly and it also allows the detection of novel lncRNAs. Furthermore, I included total RNA-sequencing of a small group of T-ALL patient samples that allowed me to also compare both RNA-sequencing technologies.

AIM 3: Completing the T-ALL subtype specific transcriptome by microRNA profiling

In the last decade, several miRNAs have been linked to T-ALL development. These are either oncogenic miRNAs or tumor suppressive miRNAs that target respectively known tumorsuppressor genes or oncogenes in T-ALL. These studies made use of an RT-qPCR platform to detect miRNA expression levels in primary T-ALL cases, but could never define a subtype specific miRNA signature. We therefore measured miRNA expression levels in 48 T-ALL patient samples of different subtypes by small RNA-sequencing **(paper 4, submitted)**. First of all, this detection method allowed to pick up many more miRNAs as the previously used platforms and secondly, this also identified, for the first time, isomiR expression in T-ALL samples. Next to that, in this dataset there was a clear subtype specific miRNA expression signature, especially for the immature T-ALL and the *TAL-R* subtype.

CHAPTER 4
RESULTS

1 DEFINING THE NOTCH1-REGULATED LNCRNAOME IN T-ALL

PAPER 1

The Notch driven long non-coding RNA repertoire in T-cell acute lymphoblastic leukemia

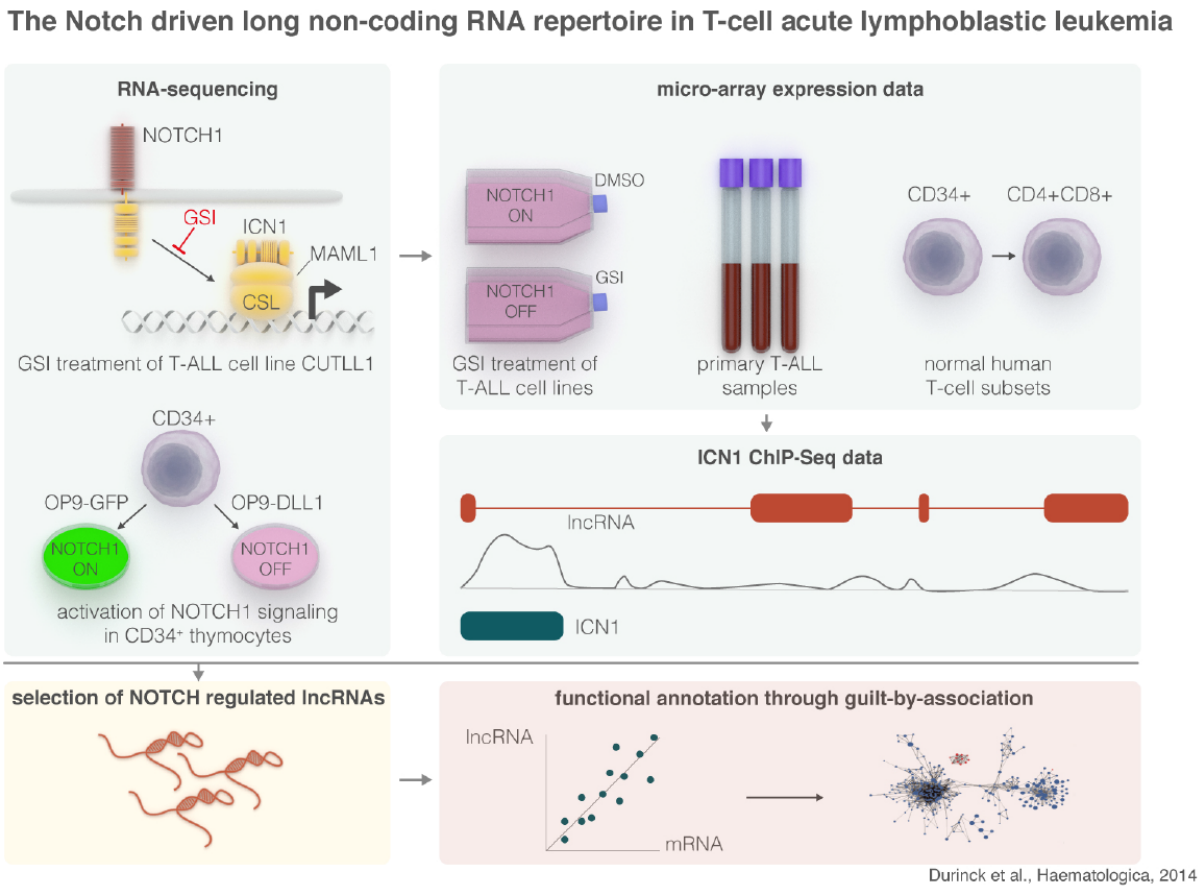
The Notch driven long non-coding RNA repertoire in T-cell acute lymphoblastic leukemia

Kaat Durinck[1*], Annelynn Wallaert[1*], Inge Van de Walle[2], Wouter Van Loocke[1], Pieter-Jan Volders[1], Suzanne Vanhauwaert[1], Ellen Geerdens[3], Yves Benoit[1], Nadine Van Roy[1], Bruce Poppe[1], Jean Soulier[4], Jan Cools[3], Pieter Mestdagh[1], Jo Vandesompele[1], Pieter Rondou[1], Pieter Van Vlierberghe[1], Tom Taghon[2**] and Frank Speleman[1**]

[*]*These authors contributed equally to this work as first authors.*
[*]*These authors contributed equally to this work as senior authors.*

[1]*Center for Medical Genetics, Ghent University, Ghent, Belgium*
[2]*Department of Clinical Chemistry, Microbiology and Immunology, Ghent University, Ghent, Belgium*
[3]*Laboratory for the Molecular Biology of Leukemia, Center for Human Genetics, KU Leuven and Center for the Biology of Disease, VIB, Leuven, Belgium*
[4]*Genome Rearrangements and Cancer Laboratory, U944 INSERM, University Paris Diderot and Hematology Laboratory, Saint-Louis Hospital, Paris, France*

Adapted from Haematologica. 2014 Dec;99(12):1808-16



ABSTRACT

Genetic studies in T-cell acute lymphoblastic leukemia have uncovered a remarkable complexity of oncogenic and loss-of-function mutations. Amongst this plethora of genetic changes, *NOTCH1* activating mutations stand out as the most frequently occurring genetic defect, identified in more than 50% of T-cell acute lymphoblastic leukemias, supporting a role as an essential driver for this gene in T-cell acute lymphoblastic leukemia oncogenesis. In this study, we aimed to establish a comprehensive compendium of the long non-coding RNA transcriptome under control of Notch signaling. For this purpose, we measured the transcriptional response of all protein coding genes and long non-coding RNAs upon pharmacological Notch inhibition in the human T-cell acute lymphoblastic leukemia cell line CUTLL1 using RNA-sequencing. Similar Notch dependent profiles were established for normal human $CD34^+$ thymic T-cell progenitors exposed to Notch signaling activity *in vivo.* In addition, we generated long non-coding RNA expression profiles (array data) from *ex vivo* isolated Notch active $CD34^+$ and Notch inactive $CD4^+CD8^+$ thymocytes and from a primary cohort of 15 T-cell acute lymphoblastic leukemia patients with known *NOTCH1* mutation status. Integration of these expression datasets with publicly available Notch1 ChIP-sequencing data resulted in the identification of long non-coding RNAs directly regulated by Notch activity in normal and malignant T cells. Given the central role of Notch in T-cell acute lymphoblastic leukemia oncogenesis, these data pave the way for the development of novel therapeutic strategies that target hyperactive Notch signaling in human T-cell acute lymphoblastic leukemia.

INTRODUCTION

The Notch pathway comprises a highly conserved signaling pathway that regulates various cellular processes in all metazoans, including stem cell maintenance, regulation of cell fate decisions, cellular proliferation, differentiation, cell death and adult tissue homeostasis.[1] As such, Notch signaling is critically involved in many different tissues including epithelial, neuronal, blood, bone, muscle and endothelial cells.[2] Precise regulation and duration of Notch signaling activity is of critical importance to ensure appropriate execution of the various developmental cues and cellular processes. Consequently, constitutive or acquired perturbation of Notch signaling frequently leads to human disease and cancer.[1-4]

Notch signaling plays multiple roles in hematopoiesis and is essential for the establishment of definitive hematopoiesis through the generation of hematopoietic stem cells,[5] as well as for their subsequent differentiation in an expanding number of blood cell types.[6-9] The role of Notch signaling has been particularly well documented in T-cell development where Notch1/Dll4 interactions are crucial to induce T-lineage differentiation at the expense of other hematopoietic lineages.[10-14]

Subsequently, Notch signaling is implemented in TCR-rearrangements,[15,16] modulation of TCR-αb versus –γδ development,[17-21] and in the support of proliferation during β-selection.[22-24] Sustained activation of Notch1 signaling beyond this developmental checkpoint has been shown to cause T-cell acute lymphoblastic leukemia (T-ALL) and *NOTCH1*

activating mutations are amongst the most frequently observed genetic alterations in T-ALL.[25,26] Importantly, γ-secretase inhibitors (GSIs) that block S3 cleavage of the Notch1 receptor and subsequent release of the intracellular signaling domain (ICN) are the subject of intensive investigation as novel drugs to combat T-ALL. However, single compound therapies almost invariably lead to resistance. Therefore, a deeper understanding of Notch signaling in normal thymocyte maturation[27] and in Notch1 activated T-ALLs could yield novel insights that could make treatment more effective.

Activation of Notch1 converts the intracellular domain (ICN1) of the Notch1 receptor into a transcriptional activator and ICN1 subsequently acts as a direct regulator of multiple target genes.[28] However, despite intensive investigation, the nature of these genes, as well as their context-dependent activation, remains largely elusive. In general, oncogenic Notch signaling promotes leukemic T-cell growth through direct transcriptional upregulation of multiple anabolic genes involved in ribosome biosynthesis, protein translation, and nucleotide and amino acid metabolism. Furthermore, Notch1 positively regulates G1/S cell cycle progression in T-ALL[29-31] and up-regulates several cyclins and CDKs,[30] in addition to the recurrent oncogene MYC.

Furthermore, Notch signaling regulates cell size, glucose uptake and PI3K-AKT activated glycolysis through HES1-mediated *PTEN* repression. Besides direct regulation of *HES1*, Notch1 is also implicated in the control of essential early T-cell genes such as *pre-TCRα* (*PTCRA*) and *IL7R*.[32-34] Taken together, these genes and pathways, as well as a further expanding list controlled by Notch1 in T-ALL and normal T-cell development, illustrate the complexity and vastness of the Notch1 controlled regulatory program.

Recent transcriptome-wide profiling efforts have uncovered an unanticipated pervasiveness of transcription of the human genome, most of which is not translated into protein.[35-38] Evidence is now emerging that more than 60% of the entire genome is transcribed.[39] In addition to previously well-characterized untranslated RNA molecules such as tRNAs, snoRNAs and microRNAs (miRNAs), thousands of so-called long non-coding RNAs (lncRNAs) have been annotated to the human genome.[40, 41] Although functional studies still need to be carried out on the vast majority of these lncRNA sequences, important cellular functions are rapidly being attributed to some of them, including roles in disease processes such as cancer.[42] In contrast to microRNAs, a picture is emerging in which lncRNAs can exhibit a myriad of different functions. These include various regulatory mechanisms of gene transcription, splicing, post-transcriptional control, protein activity and nuclear architecture.[43-45] Despite this initial progress, mechanisms of upstream regulation of lncRNAs have so far remained largely unexplored.

In this study, we investigate the role of Notch in the control of lncRNA transcription in the context of normal T-cell development and T-ALL. To this end, lncRNA expression was measured following modulation of Notch signaling in the T-ALL cell line CUTLL1 as well as in normal human thymocytes, and the recently published data on genome-wide Notch1 binding sites was used to reveal the potential for direct regulation.[34] Using this approach, we identified a total of 40 Notch-driven lncRNAs, thereby revealing a novel layer in the molecular machinery that mediates Notch signaling.

METHODS

GSI treatment of T-ALL cell lines

HPB-ALL, TALL-1, ALL-SIL and CUTLL1 cells (see also Online Supplementary Methods) were seeded at a density of $1x10^6$ cells/mL and treated with either DMSO or 1 µM of Compound E (Enzo Life Sciences). Cells were harvested 12 and 48 h after treatment.

Human thymocytes and OP9-DLL1 co-cultures

Pediatric thymus samples were obtained and used according to the guidelines of the Medical Ethical Commission of the Ghent University Hospital, Belgium. $CD34^+$ thymocytes were purified using magnetic activated cell sorting (MACS, Miltenyi Biotec) to a purity of more than 98% and seeded onto confluent OP9-GFP or OP9-DLL1 plates for 48 h in α-MEM media supplemented with 20% heat-inactivated FCS plus 100 U/mL penicillin, 100 µg/mL streptomycin, 2 mM L-glutamine and the T-lineage supporting cytokines SCF, Flt3-L and IL-7 at 5 ng/mL each.[20] Following 48 h of OP9 co-culture, cells were harvested by forceful pipetting and stained with CD45-PE (Miltenyi) to purify $CD45^+$ human thymocytes through sorting to remove contaminating OP9 stromal cells. For validation of selected lncRNAs, CD34 MACS purified thymocytes were labeled with CD34, CD1 and CD4 to sort $CD34^+CD1^-CD4^-$ uncommitted and $CD34^+CD1^+$ $CD4^-$ committed early thymocytes, while $CD4^+$ $CD8^+CD3^-$ and $CD4^+CD8^+CD3^+$ double positive thymocytes were sorted following CD4, CD8 and CD3 labeling of a total thymus suspension.[20] Sorted cells were lysed in 700 µl QIAzol (Qiagen) and stored at -70°C prior to RNA isolation.

Clinical samples

Diagnostic blood samples of 15 individuals with T-ALL were acquired after informed consent from the Department of Pediatric Hemato-Oncology at Ghent University Hospital, Belgium. This cohort includes 8 wild-type *NOTCH1* cases and 7 mutant *NOTCH1* cases (all *FBXW7* wild type). Sequencing was performed as described by Mavrakis et al.[46] Correlation analysis was performed on bone marrow lymphoblast samples from 64 T-ALL patients (unknown *NOTCH1* mutation status), which were collected after informed consent according to the Declaration of Helsinki from Saint-Louis Hospital, Paris, France. The study was approved by the Institut Universitaire d'Hématologie Institutional Review Board. This primary T-ALL cohort had been previously investigated[47] and the high-quality RNA samples from this cohort were used for lncRNA micro-array based expression profiling.

RNA sequencing

RNA samples from the CUTLL1 cells treated with GSI and thymocytes cultured on OP9-GFP/DLL1 were prepared (see also *Online Supplementary Methods*). RNA-seq was performed after unstranded poly-A library prep with an average coverage of $130x10^6$ paired-end reads. Reads were mapped to the hg19 reference genome using Tophat and transcript assembly was performed with Cufflinks. Normalizaton and differential expression analysis was carried out with DESeq2 in R. The design formula was adjusted to take into account the paired nature of the data.

Micro-array based gene expression profiling

RNA samples (see also *Online Supplementary Methods*) were profiled on a custom desig-

ned Agilent micro-array covering all protein coding genes and 12,000 lncRNAs (23,042 unique lncRNA probes) as described by Volders et al.[48] The data-analysis workflow can be found in the Online Supplementary Methods. The data discussed in this publication have been deposited in the NCBI Gene Expression Omnibus[49] and are accessible through GEO Series accession number GSE62006. Complete details of study methods can be found in the *Online Supplementary Appendix*.

RESULTS

Pharmacological Notch inhibition followed by RNA-sequencing reveals a set of Notch regulated lncRNAs in T-ALL.

To identify lncRNAs that are regulated through Notch signaling activity in the context of T-ALL, we used the γ-secretase (GSI)

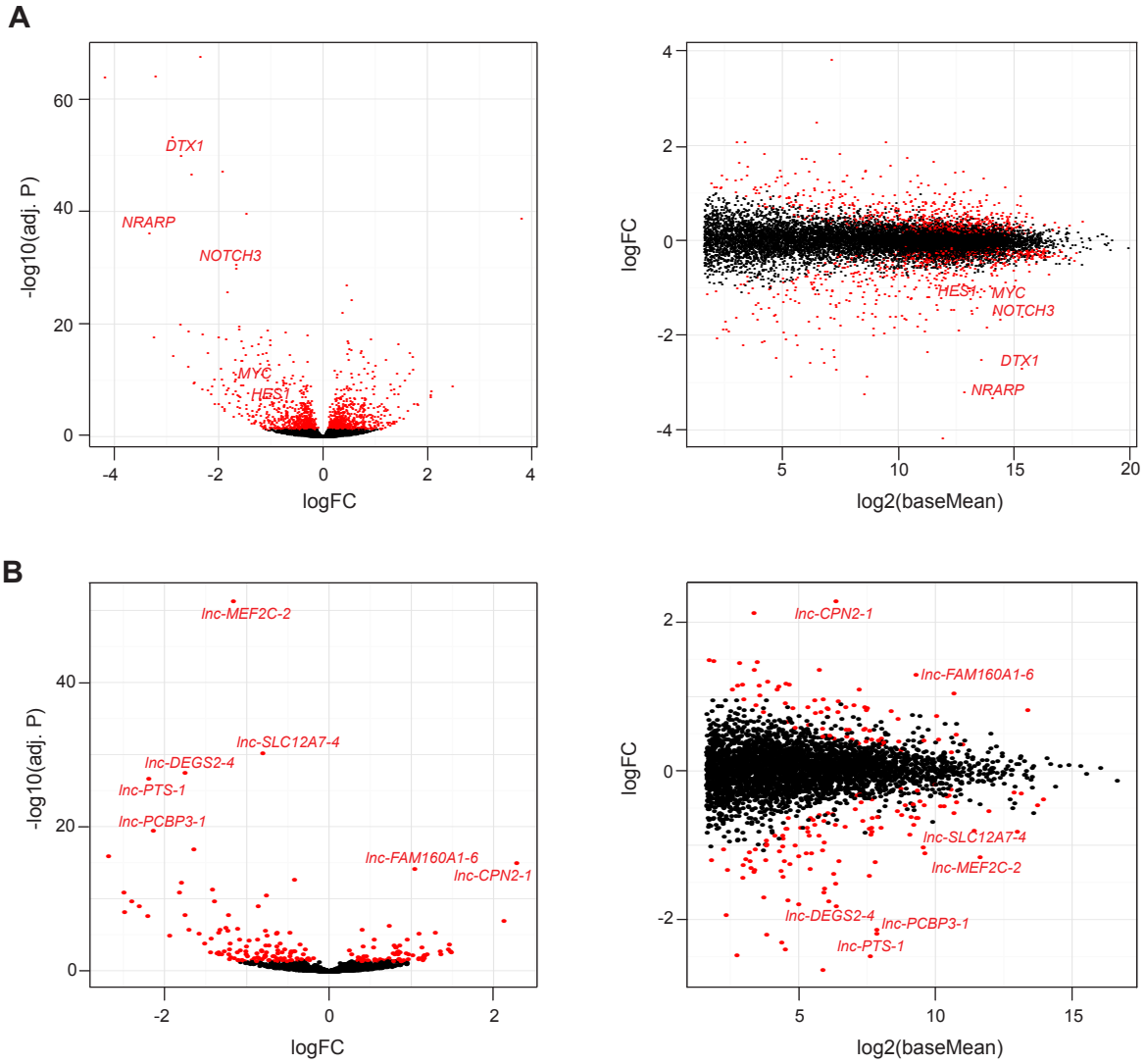


Figure 1. Pharmacological Notch inhibition followed by RNA-sequencing identifies a set of Notch regulated lncRNAs in T-ALL. (**A**) Volcano (left) and MA plot (right) representation of the differential expression of protein coding genes in CUTLL1 cells upon GSI treatment. Red dots represent the significant differentially expressed genes (adjusted P-value <0.05). (**B**) Volcano (left) and MA plot (right) representation of the differential expression of previously annotated lncRNAs in CUTLL1 cells upon GSI treatment. Red dots represent the significant differentially expressed genes (adjusted P-value<0.05; n=2). LncRNAs names depicted in the plots are the top differentially regulated lncRNAs.

inhibitor responsive T-ALL cell line CUTLL1 as a model system, since genome-wide information on this cell line is available with respect to Notch1 binding[34] and the Notch dependent expression of coding genes.[50] CUTLL1 T-ALL cells were treated with GSI for 12h and 48h in triplicate. Genome-wide transcriptional changes determined by performing differential expression analysis on the RNA-seq data (see alignment summary in Online Supplementary Table S1) with DESeq2 using Ensembl (release 75) as a reference, showed robust downregulation of several of the canonical Notch1 protein coding target genes (e.g. DTX1, NRARP, NOTCH3) upon GSI treatment (Figure 1A). A decrease in ICN1 protein levels was shown by Western blot analysis and downregulation of the canonical Notch1 target gene DTX1 upon GSI treatment was further validated by RT-qPCR (Online Supplementary Figure S1). Amongst previously annotated lncRNAs[48] we could detect significant differential expression (adjusted P-value <0.05) for 83 lncRNAs, using a basemean cut off of 100 (Figure 1B). In total, 50 out of the 83 differentially expressed lncRNAs were downregulated after GSI treatment. Besides previously annotated lncRNAs, we also detected differential expression of non-coding transcripts that had not been previously annotated in other databases (Gencode, lncRNAdb, Broad Institute and Ensembl release 64).[48] Differentially expressed lncRNA loci with a basemean higher than 100 and identified as "unknown, intergenic transcript" or "transfrag falling entirely within a reference intron" by Cuffcompare were retained for further analysis. This led to a selection of 134 lncRNA loci of which 74 were downregulated upon GSI treatment.

Transcriptional regulation of Notch regulated lncRNAs in immature normal human thymocytes

Physiological levels of Notch signaling are essential during the earliest stages of T-cell development, but no information is available on the Notch dependent expression of lncRNAs in these cells.[27] Therefore, and in order to have an independent screening method in addition to the CUTLL1 cell line to identify Notch dependent lncRNAs, we used the in *vitro* OP9-DLL1 co-culture system (Figure 2A). Here, *ex vivo* purified $CD34^+$ thymocytes from healthy human donors (n=2) were cultured on a feeder layer of stromal OP9 cells either expressing GFP (as a negative control) or the Notch1 ligand DLL1 to trigger Notch signaling. $CD34^+$ progenitor cells were collected after 48h of co-culture and deep RNA-sequencing was performed (see alignment summary in Online Supplementary Table S2). The set of differentially expressed protein coding genes and lncRNAs was defined as above for the CUTLL1 cells. Detection of differentially expressed protein coding genes known to be regulated by Notch signaling in early human thymocytes,[20,27] also validated our approach in this model system (Figure 2B). Differential expression analysis for previously annotated lncRNAs revealed 131 significantly upregulated lncRNAs as a consequence of Notch activation (Figure 2C). From these 131 lncRNAs, 27 overlapped with the set of downregulated lncRNAs upon GSI treatment of the CUTLL1 cell line (Figure 2D and Online Supplementary Table S3). Furthermore, we identified 156 unannotated lncRNA loci (base mean >100; adjusted P-value <0.05) in $CD34^+$ thymocytes that were upregulated by the Notch ligand DLL1. In total, 13 unique unannotated lncRNA loci were identified to be positively regulated by Notch in both normal and malignant T-cell

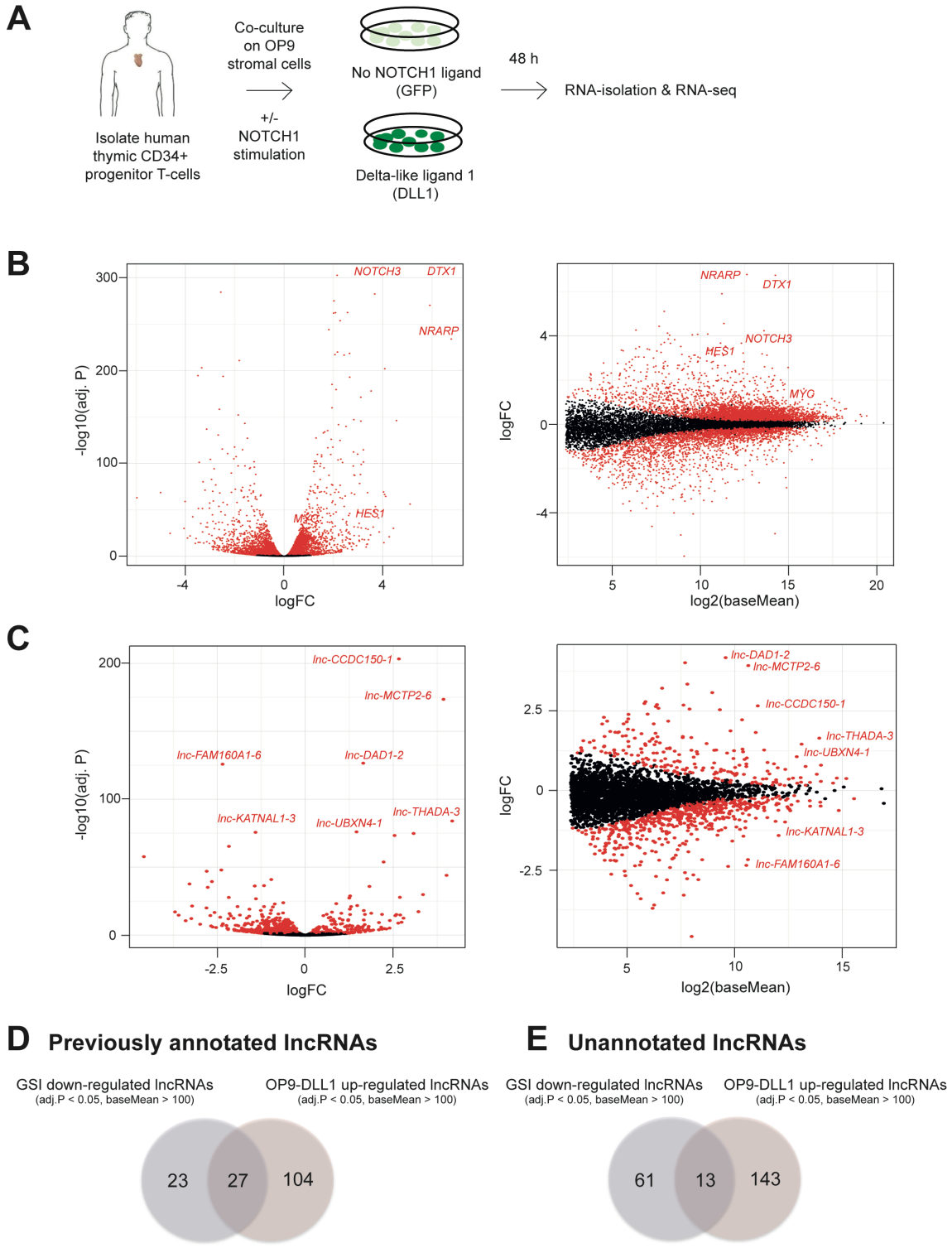


Figure 2. Transcriptional regulation of Notch regulated lncRNAs in immature normal human thymocytes. (**A**) Schematic overview of the OP9-control and -DLL1 co-culture system used to manipulate Notch signaling in healthy human immature CD34^{+} thymocytes. (**B**) Volcano (left) and MAplot (right) representation of the differential expression of protein coding genes in CD34^{+} cells upon Notch signaling induction by an OP9-DLL1 feeder layer. Red dots represent the significant differentially expressed genes (adjusted P-value <0.05; n=2). (**C**) Volcano (left) and MA plot (right) representation of the differential expression of previously annotated lncRNAs in CD34^{+} cells upon Notch activation by an OP9-DLL1 feeder layer. LncRNA names depicted in the figure are the top differentially regulated lncRNAs. (**D**) Venn diagram depicting the overlap between previously annotated lncRNAs that are downregulated upon GSI treatment of the CUTLL1 cell line and up-regulated upon co-culturing of CD34^{+} thymocytes on the OP9-DLL1 feeder layer. (**E**) Venn diagram depicting the overlap between previously unannotated lncRNAs that are downregulated upon GSI treatment of the CUTLL1 cell line and upregulated upon co-culturing of CD34^{+} thymocytes on OP9-DLL1 stromal cells.

development (Figure 2E and Online Supplementary Table S4). Amongst the set of 13 overlapping Notch lncRNA loci, the recently described *LUNAR1*[51] was present, thus supporting the validity of our approach. In addition, we also identified 33 annotated lncRNAs to be up-regulated upon GSI treatment of CUTLL1 cells by RNA-seq, 18 of them overlapping with the set of lncRNAs downregulated in $CD34^+$ T-cell progenitors upon DLL1 exposure in the OP9 *in vitro* culture system (366 in total) (Online Supplementary Table S5 and Figure S2A). In a similar manner, 7 of 57 previously unannotated lncRNAs upregulated upon GSI treatment of CUTLL1 cells overlapped with the set of 320 unannotated lncRNAs downregulated in $CD34^+$ thymocytes with DLL1 exposure (Online Supplementary Table S6 and Figure S2B). Furthermore, we hypothesize that the Notch dependent lncRNAs (both annotated and unannotated) that are not shared between CUTLL1 T-ALL cells and normal human thymocytes can be assumed to have very context-specific functions and should be regarded as potentially interesting for further exploration in future studies. For example, lncRNAs expressed exclusively in T-ALL cells could be restrictively connected to a malignant context. To evaluate the putative protein coding potential of all unannotated lncRNA loci identified by RNA-seq in CUTLL1 T-ALL cells and $CD34^+$ T-cell progenitors cultured on OP9 stromal cells, Phylogenetic Codon Substitution Frequency (PhyloCSF) scores for all loci were calculated and we could confirm that more than 90% of all unannotated lncRNA loci determined are truly 'non-coding' (Online Supplementary Figure S3). Putative unannotated lncRNA loci with a PhyloCSF score higher than the determined threshold score are listed and thus predicted to be 'coding' (Online Supplementary Tables S7 and S8) (see also Supplementary Methods).

Validation of Notch regulated lncRNAs in an extended panel of T-ALL cell lines, normal T-cell subsets and primary T-ALLs

To further validate our data, we used a custom designed Agilent micro-array[48] developed in house that contains probes for 15 of the 27 previously annotated lncRNAs and the recently identified *LUNAR1* lncRNA that were shown to be regulated by Notch in the above described RNA-seq data from the T-ALL and normal thymocyte models. First, we treated the T-ALL cell lines ALL-SIL, TALL-1, HPB-ALL and DND-41 with GSI (Supplementary Figure S4A and B) and carried out gene expression profiling after 12 h and 48 h. Inclusion of the GSI-treated CUTLL1 cell line samples and the samples of 4 donors of $CD34^+$ thymocytes cultured on OP9 stromal cells, revealed that there was a significant overlap between the RNA-sequencing data and the micro-array data as validated by overlapping the protein-coding signatures derived from both datasets by Gene Set Enrichment Analysis (GSEA) (Supplementary Figure S4C and D). Nevertheless, few lncRNAs were significantly Notch dependent over all samples of the extended panel of T-ALL cell lines (ALL-SIL, HPB-ALL, DND-41 and TALL-1), probably related to the difference in the T-ALL genetic subgroup and concomitant differences in maturation arrest of the different cell lines evaluated (Figure 3A). From our selection, only *lnc-PLEKHB2-1* and *lnc-UBXN4-1* were differentially expressed at a significant level, while *lnc-GSDMC-2* and *lnc-CA7-2* narrowly failed to reach significance. As we were able to detect the previously unannotated and recently described lncRNA *LUNAR1*[51] on our custom designed micro-array platform, this lncRNA was one of

the strongest overlapping and significantly differentially expressed lncRNAs amongst the four GSI-treated T-ALL cell lines screened. Secondly, we validated the Notch dependency of selected lncRNAs in normal thymocytes by analyzing their expression in the most immature Notch dependent $CD34^+$ stages in comparison to the Notch independent $CD4^+CD8^+$ double positive stages of human T-cell development. As is evident from the profiles of the Notch target gene *DTX1*, we could show that in these two T-cell subpopulations *LUNAR1* follows the expression pattern of this canonical Notch target (Figure 3B). Remarkably, 9 out of 15 lncRNAs (and *LUNAR1*) from this selection significantly correlated with *DTX1* expression (Spearman rho correlation), supporting their regulation by Notch during early stages of normal T-cell development (Figure 3C and Online Supplementary Table S9).

Moreover, we also had access to 15 primary T-ALL samples of which 7 harbored activating *NOTCH1* mutations while 8 were wild type (all cases are *FBXW7* wild type). There was a significant difference in expression of *LUNAR1* and *lnc-FAM120AOS-1* between *NOTCH1* wildtype and mutant cases (Figure 3D and Online Supplementary Tables S10 and S11). By implying an additional dataset of 64 primary T-ALL patient samples, we could correlate the expression of lncRNAs *lnc-PGBD5-2*, *lnc-FAM120AOS-1*, *lnc-c2orf55-1* and *LUNAR1* with the Notch1 positively regulated gene set Vilimas_NOTCH1 _targets_up[52] by GSEA (Figure 3E). Overall, these independent experiments confirm the Notch dependent regulation of the selected lncRNAs, thereby validating the RNA-seq data from the GSI treated CUTLL1 T-ALL cell line and the Notch perturbed normal human thymocytes.

Genome-wide analysis reveals direct Notch1 binding to selected lncRNAs

To further validate the direct regulation of selected lncRNAs by Notch, publicly available ChIP sequencing (ChIP-seq) data from the CUTLL1 cell line were analyzed for Notch1 binding at specific loci.[34] From the Notch-driven annotated lncRNAs that overlapped between normal and malignant thymocytes (Figure 2D), 13 out of the 27 lncRNAs were bound by ICN1 (Online Supplementary Table S12) as illustrated for *lnc-UBXN4-1* and *lnc-PLEKHB2-1* (Figure 4). Remarkably, 12 out of the 13 lncRNAs with a Notch1 binding peak also show Brd4 and Med1 binding. Notably, from the putative Notch regulated lncRNAs that showed correlated expression with *DTX1* in $CD34^+$ and $CD4^+CD8^+$ normal thymocytes, 6 out of 9 (Figure 4 and Supplementary Figure S5) showed binding of ICN1, suggesting that the majority is a direct Notch target.

In addition, we evaluated the presence of H3K27 acetylation (H3K27ac) ChIP-seq signal at these lncRNA loci, a histone mark indicative for putative enhancer regions. For 18 out of the 27 selected lncRNAs, H3K27ac ChIP-seq signal was present in close proximity of the promoter region (Online Supplementary Table S12), suggesting the presence of enhancer sequences.

Moreover, this public ChIP-seq data also showed LUNAR1 to be directly bound by ICN1, Brd4, Med1 and H3K27ac (Figure 5). We also evaluated ICN1 binding at annotated lncRNA loci up-regulated upon GSI treatment of CUTLL1 T-ALL cells and downregulated in $CD34^+$ T-cell progenitors upon DLL1 exposure in the OP9 *in vitro* co-culture system. Only 4 out of the 18 overlapping annotated lncRNAs

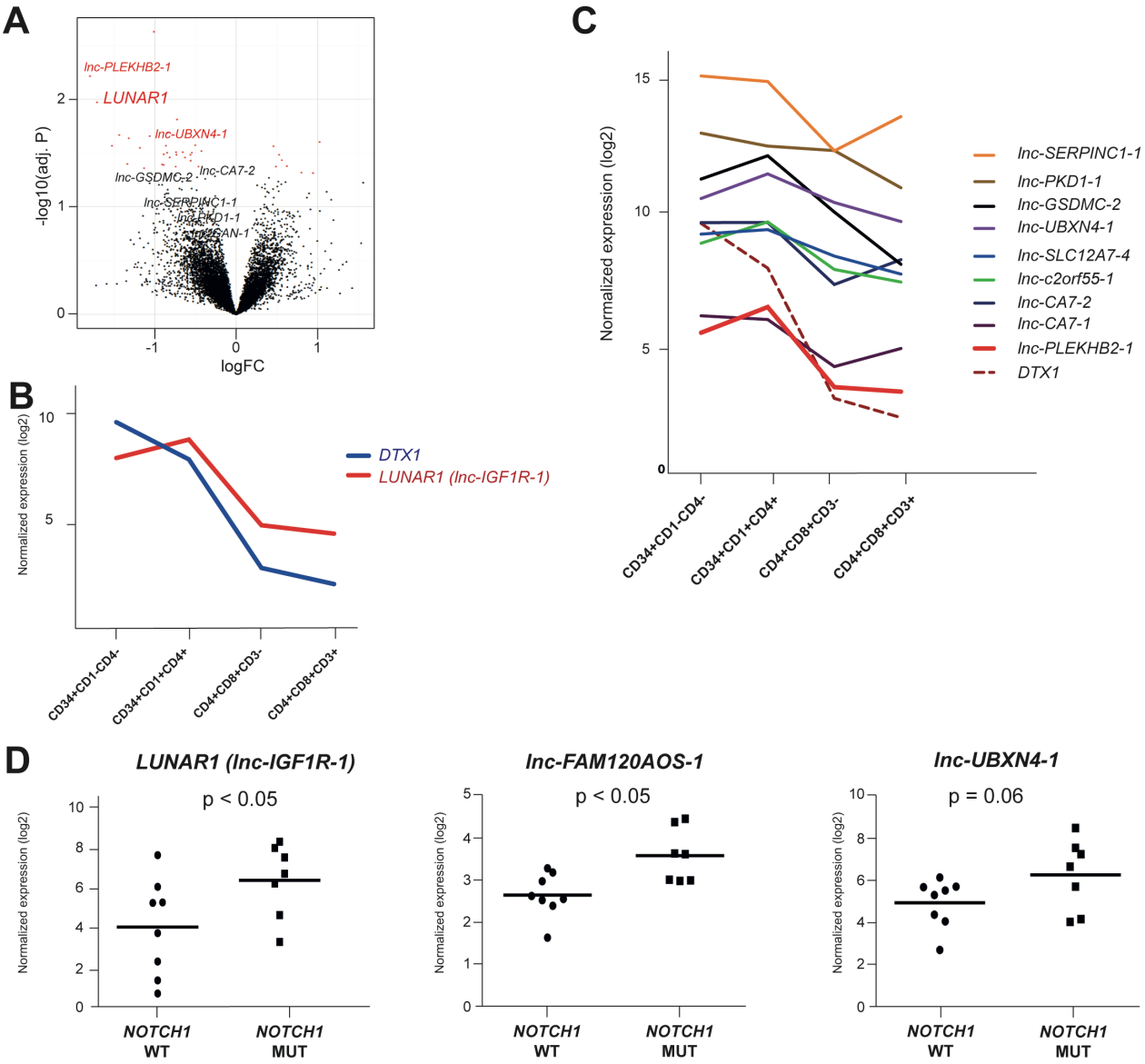


Figure 3. Screening expression of Notch regulated lncRNAs in an extended panel of T-ALL cell lines, normal T-cells subsets and primary T-ALLs. (**A**) Volcano plot representation of the differential expression of lncRNAs upon GSI treatment of ALL-SIL, TALL-1, HPB-ALL and DND-41 cells. Red dots represent the significant differentially expressed genes (adjusted P-value <0.05). *LUNAR1* was amongst the top-differentially expressed lncRNAs across the panel of GSI-treated T-ALL cell lines. The other lncRNA names depicted in the figure are some of the selected lncRNAs from the CUTLL1 GSI treatment and the OP9-DLL1 co-culture system. (**B**) Plot representing the expression of lncRNAs in selected Notch-dependent and -independent stages of normal T-cell development for one healthy donor. LUNAR1 expression is significantly correlated with the expression of DTX1 (see also Online Supplementary Table S6) and the data are representative for 4 independent donors. (**C**) Similar analysis as in (B) for the other lncRNAs that are significantly correlated with the expression of DTX1 (see also Online Supplementary Table S6); data are representative for 4 independent donors. (**D**) Expression of *LUNAR1*, *lnc-FAM120AOS-1* and *lnc-UBXN4-1* in *NOTCH1* wildtype (WT) versus *NOTCH1* mutant (MUT) primary T-ALL samples. (Continued on the next page.)

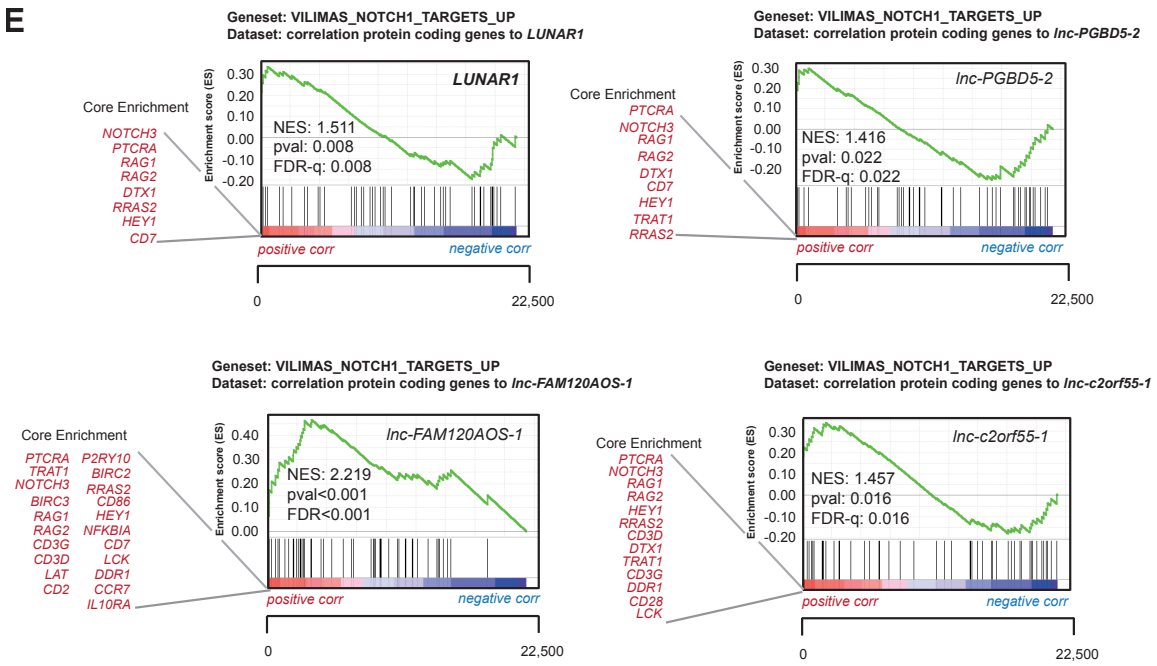


Figure 3 (Continued). Screening expression of Notch regulated lncRNAs in an extended panel of T-ALL cell lines, normal T-cells subsets and primary T-ALLs. (**E**) Gene set enrichment analysis (GSEA) using the public gene set 'VILIMAS_NOTCH1_TARGETS_UP'52 and the Spearman correlations between all protein coding genes and the set of 15 selected annotated candidate Notch lncRNAs was performed. This NOTCH1 signature was significantly enriched within the set of protein coding genes positively correlated to the expression of *lnc-PGBD5-2*, *lnc-FAM120AOS-1* and *lnc-c2orf55-1*. This enrichment was also found for lncRNA *LUNAR1*.

that are negatively regulated by Notch (Supplementary Figure S2A) showed direct binding by ICN1. The same analysis was performed on the set of 7 unannotated lncRNAs repressed by Notch1 signaling (Supplementary Figure S2B). Only 3 out of these 7 lncRNA loci showed ICN1 binding in the proximity of its promoter region. Given the established predominant role of Notch1 as a transcriptional activator, lncRNAs that are negatively affected by Notch1 signaling may actually be indirect targets.

Attributing functional annotation to Notch regulated lncRNAs through guilt-by-association analysis

As described above, we defined a core set of 27 Notch driven and previously annotated lncRNAs by considering only those differentially expressed and positively regulated by Notch signaling in the GSI perturbation model in CUTLL1 cells and the *in vitro* OP9-DLL1 co-culture system. As a next step, we aimed to assign potential functionalities to each of these candidates. To this end, we used the so called 'guilt-by-association' approach (see also Supplementary Methods). As previously mentioned, 15 out of these 27 lncRNAs (together with *LUNAR1*) were detectable by a probe on our custom designed micro-array platform. In a first step, we calculated the Spearman correlation coefficients between the lncRNAs-of-interest and all protein coding genes using the expression data of a primary T-ALL cohort of 64 patients from which we profiled all samples on the custom designed Agilent array.[48] These correlations were subsequently used as an input for a GSEA preranked analysis. Next, the output of this GSEA analysis was further refined into functional clusters of enriched gene sets using the Cytoscape plug-in enrichment mapping. This analysis yielded markedly different functional clustering patterns for each of the 16 lncRNAs

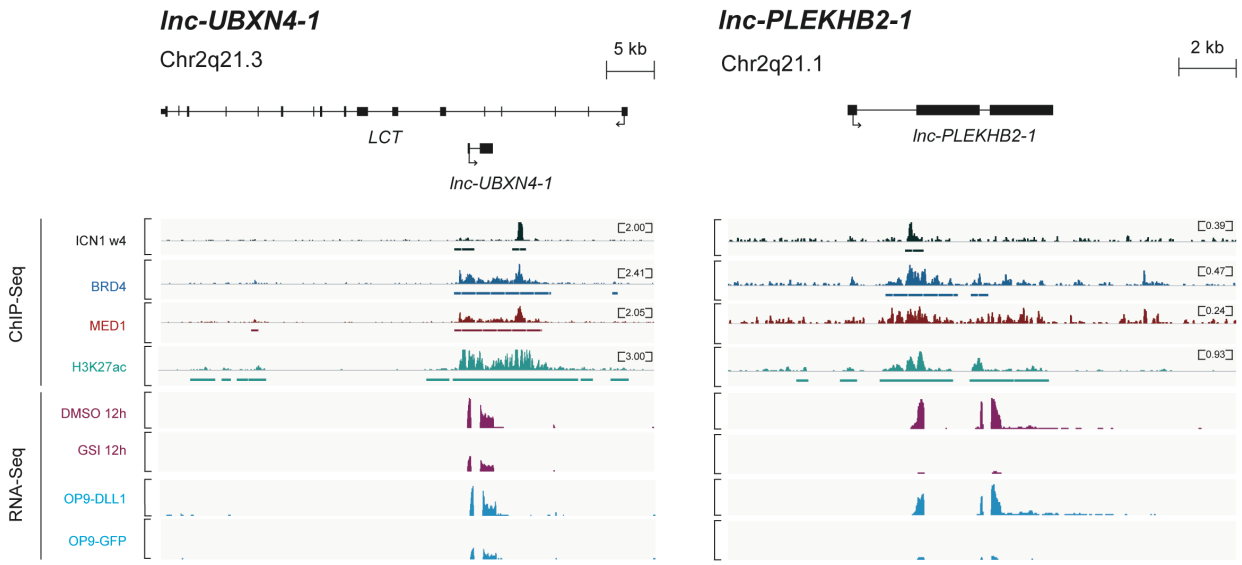


Figure 4. Notch1 ChIP-seq reveals direct binding of Notch1 to a subset of regulated lncRNAs. Representation of ChIP-sequencing tracks34 for Notch1, Brd4, Med1 and H3K27ac and representative RNA-sequencing tracks for CUTLL1 DMSO/GSI treatment and OP9-GFP/DLL1 of *lnc-UBXN4-1* and *lnc-PLEKHB2-1*.

analyzed (including *LUNAR1*). Important putative functionalities were represented in each of the networks as exemplified by TCR-signaling and phospholipid metabolism for *lnc-PLEKHB2-1*, DNA replication and DNA repair for *lnc-UBXN4-1* and splicing and cell cycle regulation for *LUNAR1* (Figure 6A-C and Supplementary Figure S6A-M*).*

DISCUSSION

Non-coding RNAs are emerging as important players in normal development and disease, including cancer. In previous studies, we investigated the role of miRNAs in T-cell acute lymphoblastic leukemia (T-ALL), thereby identifying a small set of miRNAs that is responsible for the cooperative suppression of several tumor suppressor genes.[46]

These miRNAs produced overlapping and cooperative effects with several bona fide T-ALL tumor suppressor genes including *IKZF1*, *PTEN*, *BIM*, *PHF6*, *NF1* and *FBXW7*, and more recently this network was expanded further with *PHF6*.[53] In order to provide some insight into the genetic components driving long non-coding RNAs in T-ALL formation, we performed an integrated analysis of lncRNA profiling data sets from GSI inhibited Notch-driven T-ALL cell lines and Notch-stimulated immature normal human thymocytes using the OP9 co-culture system, together with publicly available genome-wide data on Notch1 binding and specific chromatin marks. In addition, we correlated the expression of Notch-dependent lncRNAs with the Notch-dependent stages of normal thymocytes during T-cell differentiation. Overall, our work establishes a novel lncRNA network that acts downstream of Notch during normal and malignant thymocyte development.

Our study provides a number of fundamental new insights into Notch-dependent regulation of lncRNAs in TALL and normal developing thymocytes. First, we unambiguously demonstrate that a significant number of lncRNAs are directly regulated by Notch signaling activity. Through RNA-sequencing, we identified 40 lncRNAs that are positively regu-

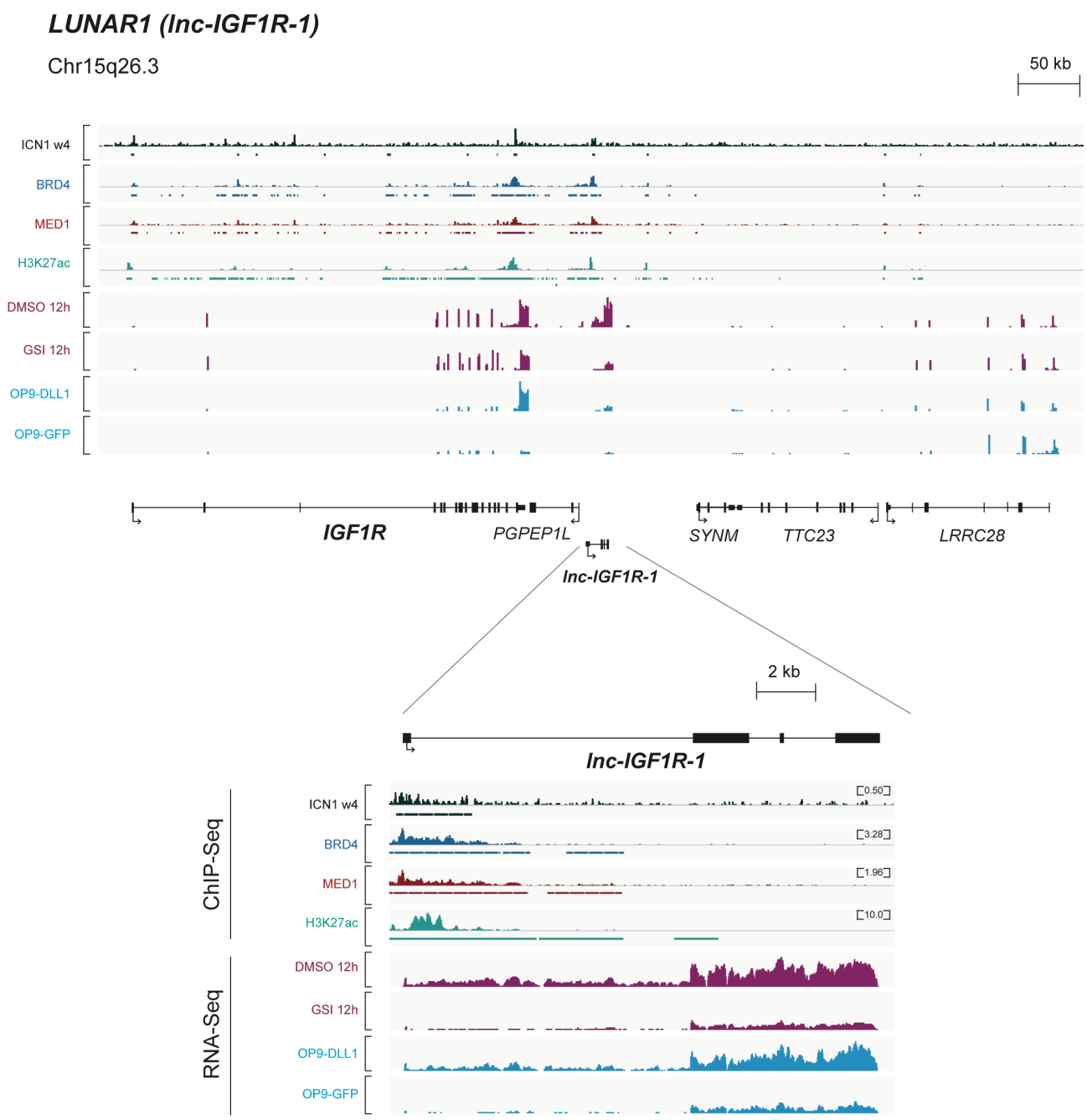


Figure 5. Notch1 ChIP-seq reveals direct binding of Notch1 to LUNAR1. *LUNAR1*[51] (*lnc-IGF1R-1*) was identified amongst the top differentially expressed novel, unannotated lncRNAs in both GSI-treated CUTLL1 cells and CD34+ thymic progenitor cells exposed to DLL1-triggered Notch signaling. Publically available ChIP-seq tracks[34] for ICN1 BRD4, MED1 and H3K27ac as well as representative in house generated RNA-seq data tracks for CUTLL1 DMSO/GSI treated cells and OP9-GFP/DLL1 are shown at the *LUNAR1* locus.

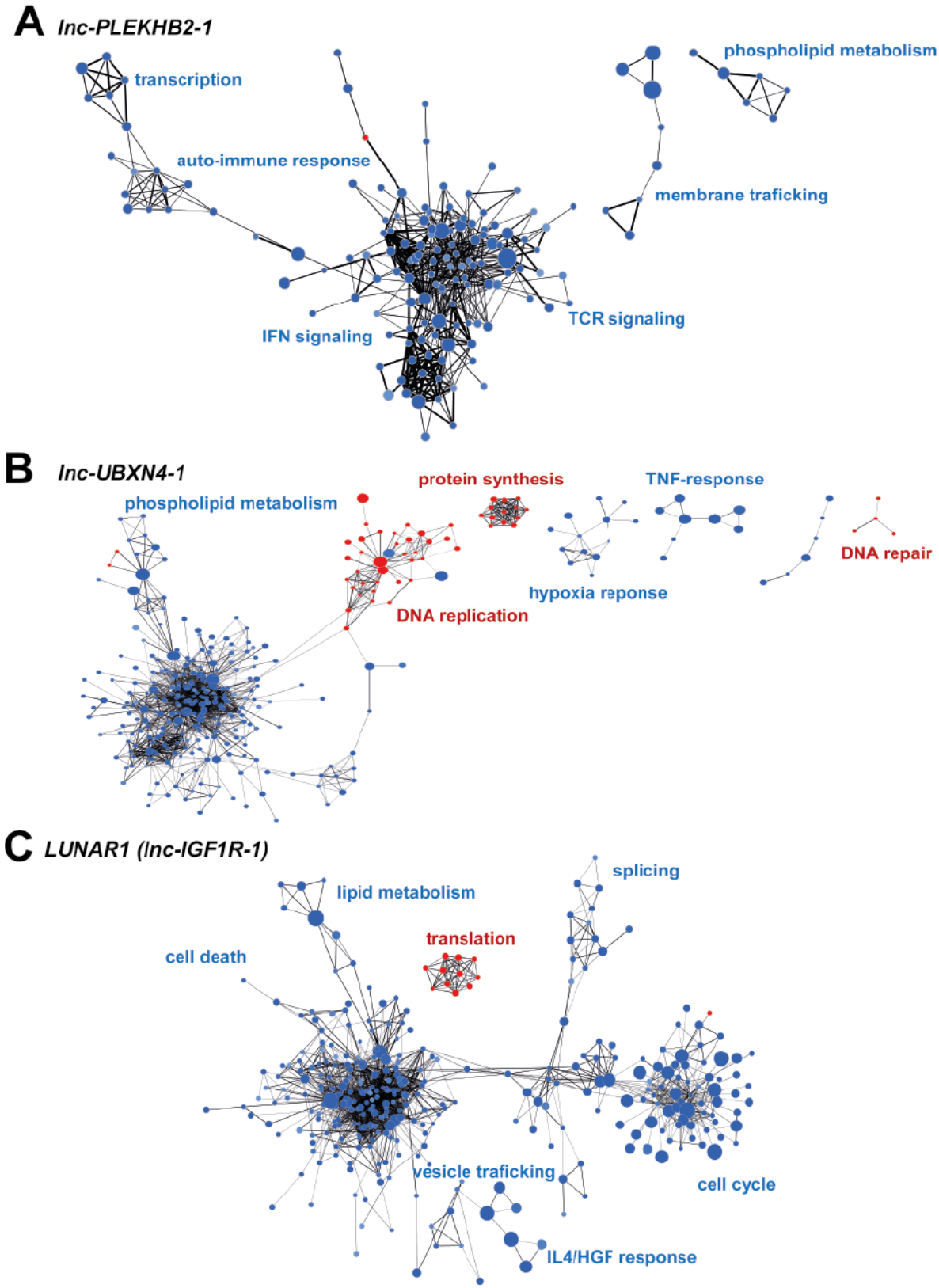


Figure 6. Attributing functional annotation to selected, annotated lncRNAs through guilt-by-association analysis. Enrichment maps of gene sets correlated with the expression of (**A**) *lnc-PLEKHB2-1*, (**B**) *lnc-UBXN4-1* and (**C**) *LUNAR1*. Red nodes represent the positively correlated gene sets to the lncRNA of interest, blue nodes the negatively correlated gene sets. The size of the nodes depicts the size of the gene sets. Nodes that are clustered represent gene sets with the same or similar functional indication.

lated by Notch in both normal and malignant T lymphocytes (annotated as well as previously unannotated lncRNAs), supporting an important role for these lncRNAs in Notch-regulated T-cell biology. This could be related to various functions of Notch signaling, including T-cell lineage specification and commitment, proliferation and differentiation. Importantly, the recently identified lncRNA *LUNAR1*[51] was present amongst the most robustly Notch regulated long non-coding RNAs in our data sets. *LUNAR1* was shown to be required for efficient T-ALL growth as a consequence of its role in enhancing *IGF1R* mRNA expression to sustain IGF1 signaling.[51] As a prelude to assigning functional annotation to the newly assigned Notch-regulated lncRNAs in this study, we applied the so-called 'guilt-by-association' approach in which functions are predicted based upon correlation with known protein coding genes and subsequent gene set enrichment analysis. For the selected lncRNAs, various functions were predicted, several of which are linked to T-cell biology or processes that are perturbed in cancer. This marks these lncRNAs as prime targets for further functional studies in order to unravel their mode of action and assess to what extent they might serve as future therapeutic targets for treatment of T-ALL.

Not all of the 40 overlapping lncRNAs displayed ICN1 binding, as is evident from the publicly available ChIP-seq data.[34] This may relate to the complexity of the chromosomal 3D-structures that are generated when lncRNAs act as *cis*-regulatory elements, as well as to the sensitivity of the Notch1 ChIP procedure. However, for the previously annotated lncRNAs directly bound by ICN1, all but one displayed Brd4 and Med1 binding. Those lncRNAs that are characterized by Brd4 and Med1 ChIP-seq signal are also characterized by the presence of H3K27ac ChIP-seq signal, which could be indicative of an enhancer activity of these loci.

A second aspect of our study involved the identification of novel, previously unannotated lncRNAs. Indeed, previous studies have shown that lncRNAs are often shown to have a very restricted expression pattern, but with biologically high relevant expression.[40-45] This includes expression during very specific time points during development and/or differentiation, as well as restriction to very specific cell subsets. Typically, these lncRNAs are expressed at significant levels in these cells whereas in other cell types their expression is very low or absent. Here, we identified novel lncRNAs in the CUTLL1 cell line and in the OP9-DLL1 co-culture system. Interestingly, 61 of these lncRNAs were present in T-ALL cells only, suggesting that their ectopic expression could be restricted to the malignant context. Likewise, lncRNAs only present in thymocytes may be implicated in differentiation of normal T cells which is disrupted in T-ALL cells, or may reflect differential Notch3 activity as the DLL1 ligand, to which $CD34^{+}$ progenitor T cells are exposed in the OP9-DLL1 co-culture system, can activate both Notch1 and Notch3 (Waegemans E, Van de Walle I and Taghon T, unpublished data on preferential Notch receptor-ligand interactions in human, 2011) while both receptors are implicated in modulating human T-cell development.[21] Both subsets of lncRNAs may, therefore, serve as novel therapeutic targets for T-ALL treatment. It is evident that our work, as well as the recent paper by Aifantis and colleagues,[51] strongly favors an important role for lncRNAs in normal T-cell development and T-ALL oncogenesis. Moreover, we show that the Notch1 transcription factor directly controls the

transcription of many of these long non-coding RNAs. Therefore, one can predict that other oncogenic transcription factors and drivers in T-ALL, such as *TAL1*, *TLX1/TLX3*, *LMO1/2* and *HOXA* genes, as well as other transcriptional regulators of normal T-cell development, will also perform similar transcriptional control. Therefore, specific T-ALL subgroups and more distinct subsets of normal immature developing T cells need to be analyzed in human for detection of all lncRNAs. We predict that this will further dramatically expand the lncRNA landscape for T-ALL and thymocyte maturation, and thus provide an important regulatory framework for understanding some of the unique features that control human T-cell biology.[27] Finally, given the central role of oncogenic Notch1 activation in most if not all T-ALLs, and the current limitations of targeted therapy, further exploration of the new therapeutic opportunities offered by these lncRNAs in the context of their specific functionality is strongly recommended.

FUNDING

The authors would like to thank the following funding agencies: the Fund for Scientific Research Flanders ('FWO Vlaanderen' research projects G.0202.09, G.0869.10N, 3G055013N, 3G056413N to FS; 3GA00113N, 3G065614, G.0C47.13N to PVV and G0B2913N, G037514N, 3G002711 to TT; PhD grant to AW; postdoctoral grants to IvdW, TT, PM, PVV, and PR; BP is a senior clinical investigator), IWT Vlaanderen (PhD grant to KD), the Belgian Foundation against Cancer (FS, JVds (SCIE 2010-177)); Ghent University (GOA grant 01G01910 to FS), the Cancer Plan from the Federal Public Service of Health (FS), the Children Cancer Fund Ghent (FS) and the Belgian Program of Interuniversity Poles of Attraction (IUAP P7/03 and P7/07). We also would like to thank Aline Eggermont for excellent technical assistance.

REFERENCES

1. Hori K, Sen A, Artavanis-Tsakonas S. Notch signaling at a glance. J Cell Sci. 2013;126(Pt10):2135-40.
2. Aster JC. In brief: Notch signalling in health and disease. J Pathol. 2014;232(1):1-3.
3. Koch U, Radtke F. Notch signaling in solid tumors. Curr Top Dev Biol. 2010;92:411-55.
4. Ntziachristos P, Lim JS, Sage J, Aifantis I. From fly wings to targeted cancer therapies: a centennial for notch signaling. Cancer Cell. 2014;25(3):318-34.
5. Kumano K, Chiba S, Kunisato A, Sata M, Saito T, Nakagami-Yamaguchi E, et al. Notch1 but not Notch2 is essential for generating hematopoietic stem cells from endothelial cells. Immunity. 2003;18(5): 699-711.
6. Oh P, Lobry C, Gao J, Tikhonova A, Loizou E, Manent J, et al. In vivo mapping of notch pathway activity in normal and stress hematopoiesis. Cell Stem Cell. 2013;13(2):190-204.
7. Radtke F, MacDonald HR, Tacchini-Cottier F. Regulation of innate and adaptive immunity by Notch. Nat Rev Immunol. 2013;13(6):427-37.
8. Sandy AR, Maillard I. Notch signaling in the hematopoietic system. Expert Opin Biol Ther. 2009;9(11):1383-98.
9. Yuan JS, Kousis PC, Suliman S, Visan I, Guidos CJ. Functions of notch signaling in the immune system: consensus and controversies. Annu Rev Immunol. 2010;28: 343-65.
10. De Obaldia ME, Bell JJ, Wang X, Harly C, Yashiro-Ohtani Y, DeLong JH, et al. T cell d evelopment requires constraint of the myeloid regulator C/EBP-alpha by the Notch target and transcriptional repressor Hes1. Nat Immunol. 2013;14(12):1277-84.
11. Hozumi K, Mailhos C, Negishi N, Hirano K, Yahata T, Ando K, et al. Delta-like 4 is indispensable in thymic environment specific for T cell development. J Exp Med. 2008;205 (11):2507-13.
12. Koch U, Fiorini E, Benedito R, Besseyrias V, Schuster-Gossler K, Pierres M, et al. Deltalike 4 is the

essential, nonredundant ligand for Notch1 during thymic T cell lineage commitment. J Exp Med. 2008;205(11): 2515-23.
13. Radtke F, Wilson A, Stark G, Bauer M, van Meerwijk J, MacDonald HR, et al. Deficient T cell fate specification in mice with an induced inactivation of Notch1. Immunity. 1999;10(5):547-58.
14. Wilson A, MacDonald HR, Radtke F. Notch1-deficient common lymphoid precursors adopt a B cell fate in the thymus. J Exp Med. 2001;194(7):1003-12.
15. De Smedt M, Hoebeke I, Reynvoet K, Leclercq G, Plum J. Different thresholds of Notch signaling bias human precursor cells toward B-, NK-, monocytic/dendritic-, or T-cell lineage in thymus microenvironment. Blood. 2005;106(10):3498-506.
16. Wolfer A, Wilson A, Nemir M, MacDonald HR, Radtke F. Inactivation of Notch1 impairs VDJbeta rearrangement and allows pre- TCR-independent survival of early alpha beta Lineage Thymocytes. Immunity. 2002;16(6):869-79.
17. Ciofani M, Knowles GC, Wiest DL, von Boehmer H, Zuniga-Pflucker JC. Stage-specific and differential notch dependency at the alphabeta and gammadelta T lineage bifurcation. Immunity. 2006;25(1):105-16.
18. Garbe AI, von Boehmer H. TCR and Notch synergize in alphabeta versus gammadelta lineage choice. Trends Immunol. 2007;28(3): 124-31.
19. Taghon T, Yui MA, Pant R, Diamond RA, Rothenberg EV. Developmental and molecular characterization of emerging beta- and gammadelta-selected pre-T cells in the adult mouse thymus. Immunity. 2006;24(1):53-64.
20. Van de Walle I, De Smet G, De Smedt M, Vandekerckhove B, Leclercq G, Plum J, et al. An early decrease in Notch activation is required for human TCR-alphabeta lineage differentiation at the expense of TCRgammadelta T cells. Blood. 2009;113(13):2988-98.
21. Van de Walle I, Waegemans E, De Medts J, De Smet G, De Smedt M, Snauwaert S, et al. Specific Notch receptor-ligand interactions control human TCR-alphabeta/gammadelta development by inducing differential Notch signal strength. J Exp Med. 2013;210(4):683-97.
22. Ciofani M, Zuniga-Pflucker JC. Notch promotes survival of pre-T cells at the betaselection checkpoint by regulating cellular metabolism. Nat Immunol. 2005;6(9):881-8.
23. Maillard I, Tu L, Sambandam A, Yashiro-Ohtani Y, Millholland J, Keeshan K, et al. The requirement for Notch signaling at the beta-selection checkpoint in vivo is absolute and independent of the pre-T cell receptor. J Exp Med. 2006;203(10):2239 45.
24. Taghon T, Van de Walle I, De Smet G, De Smedt M, Leclercq G, Vandekerckhove B, et al. Notch signaling is required for proliferation but not for differentiation at a well-defined beta-selection checkpoint during human T-cell development. Blood. 2009;113(14):3254-63.
25. Van Vlierberghe P, Ferrando A. The molecular basis of T cell acute lymphoblastic leukemia. J Clin Invest. 2012;122(10):3398-406.
26. Weng AP, Ferrando AA, Lee W, Morris JPt, Silverman LB, Sanchez-Irizarry C, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science. 2004;306(5694):269-71.
27. Taghon T, Waegemans E, Van de Walle I. Notch signaling during human T cell development. Curr Top Microbiol Immunol. 2012;360:75-97.
28. Kopan R, Ilagan MX. The canonical Notch signaling pathway: unfolding the activation mechanism. Cell. 2009;137(2):216-33.
29. Dohda T, Maljukova A, Liu L, Heyman M, Grander D, Brodin D, et al. Notch signaling induces SKP2 expression and promotes reduction of p27Kip1 in T-cell acute lymphoblastic leukemia cell lines. Exp Cell Res. 2007;313(14):3141-52.
30. Joshi I, Minter LM, Telfer J, Demarest RM, Capobianco AJ, Aster JC, et al. Notch signaling mediates G1/S cell-cycle progression in T cells via cyclin D3 and its dependent kinases. Blood. 2009;113(8):1689-98.
31. Rao SS, O'Neil J, Liberator CD, Hardwick JS, Dai X, Zhang T, et al. Inhibition of NOTCH signaling by gamma secretase inhibitor engages the RB pathway and elicits cell cycle exit in T-cell acute lymphoblastic leukemia cells. Cancer Res. 2009;69(7):3060-8.
32. Gonzalez-Garcia S, Garcia-Peydro M, Martin-Gayo E, Ballestar E, Esteller M, Bornstein R, et al. CSL-MAML-dependent Notch1 signaling controls T lineage-specific IL-7R{alpha} gene expression in early human thymopoiesis and leukemia. J Exp Med. 2009;206(4):779-91.
33. Reizis B, Leder P. Direct induction of T lymphocyte- specific gene expression by the mammalian Notch signaling pathway. Genes Dev. 2002;16(3):295-300.

34. Wang H, Zang C, Taing L, Arnett KL, Wong YJ, Pear WS, et al. NOTCH1-RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. Proc Natl Acad Sci USA. 2014;111(2):705-10.
35. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. Science. 2005;309(5740):1559-63.
36. Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447(7146):799-816.
37. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. Nature. 2012;489(7414):101-8.
38. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense transcription in the mammalian transcriptome. Science. 2005;309(5740): 1564-6.
39. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature. 2002;420(6915):563-73.
40. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011;25(18): 1915-27.
41. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009;458(7235):223-7.
42. Gutschner T, Diederichs S. The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol. 2012;9(6):703-19.
43. Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol. 2013;14(11):699-712.
44. Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. Genetics. 2013;193(3):651-69.
45. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell. 2013;154(1):26-46.
46. Mavrakis KJ, Van Der Meulen J, Wolfe AL, Liu X, Mets E, Taghon T, et al. A cooperative microRNA-tumor suppressor gene network in acute T-cell lymphoblastic leukemia (TALL). Nat Genet. 2011;43(7):673-8.
47. Clappier E, Gerby B, Sigaux F, Delord M, Touzri F, Hernandez L, et al. Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. J Exp Med. 2011; 208(4):653-61.
48. Volders PJ, Helsens K, Wang X, Menten B, Martens L, Gevaert K, et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res. 2013;41(Database issue):D246-51.
49. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207-10.
50. Palomero T, Lim WK, Odom DT, Sulis ML, Real PJ, Margolin A, et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. Proc Natl Acad Sci USA. 2006;103(48):18261-6.
51. Trimarchi T, Bilal E, Ntziachristos P, Fabbri G, Dalla-Favera R, Tsirigos A, et al. Genomewide Mapping and Characterization of Notch-Regulated Long Noncoding RNAs in Acute Leukemia. Cell. 2014;158(3):593-606.
52. Vilimas T, Mascarenhas J, Palomero T, Mandal M, Buonamici S, Meng F, et al. Targeting the NF-kappaB signaling pathway in Notch1-induced T-cell leukemia. Nat Med. 2007;13(1):70-7.
53. Mets E, Van Peer G, Van der Meulen J, Boice M, Taghon T, Goossens S, et al. MicroRNA-128-3p is a novel oncomiR targeting PHF6 in T-cell acute lymphoblastic leukemia. Haematologica. 2014;99(8):1326-33.

SUPPLEMENTARY METHODS

Cell lines and compound treatment

HPB-ALL, TALL-1 and ALL-SIL cells were obtained from the DSMZ cell line repository, CUTLL1 cells were a kind gift of H.G. Wendel (Memorial Sloan Kettering Cancer Center, New York, USA). Cells were maintained in RPMI-.-1640 medium (Life Technologies, 52400-025), supplemented with 10% or 20 % (ALL-SIL and CUTLL1) fetal bovine serum (Biochrom AG, S0615), 1 % L-glutamin (Life Technologies, 1514-148) and 1% penicillin/streptomycin (Life Technologies, 15160-047). CUTLL1, HPB-ALL, TALL-1 and ALL-SIL cells were seeded at a density of $1*10^6$ cells/ml and treated with either DMSO or 1 µM of Compound E (Enzo Life Sciences). Cells were harvested 12 and 48 hours after treatment.

RNA-isolation, cDNA synthesis and RT-qPCR

Total RNA was harvested with the miRNeasy minikit (Qiagen) with DNase treatment on-column. RNA-concentrations were measured by means of spectrophotometry (Nanodrop). cDNA was synthesized using the iScript cDNA synthesis kit (Biorad, 170-8891) according to the manufacturers' protocol, starting with 500ng of RNA, followed by RT-qPCR using the Light Cycler 480 (Roche). Finally, qPCR data was analyzed using the qBasePLUS software (Biogazelle) according to the ΔΔCt-method.

Target	Forward primer	Reverse primer
c-MYC	GCCACGTCTCCACACATCAG	TGGTGCATTTTCGGTTGTTG
HES1	TGTCAACACGACACCGGATAAA	CCATAATAGGCTTTGATGACTTTCTG
DTX1	ACGAGAAAGGCCGGAAGGT	GGTGTTGGACGTGCCGATAG
HPRT	TGACACTGGCAAAACAATGCA	GGTCCTTTTCACCAGCAAGCT
HMBS	GGCAATGCGGCTGCAA	GGGTACCCACGCGAATCAC
TBP	CACGAACCACGGCACTGATT	TTTTCTTGCTGCCAGTCTGGAC
B2M	TGCTGTCTCCATGTTTGATGTATCT	TCTCTGCTCCCCACCTCTAAGT

Reaction conditions for RT-qPCR

Components	Amount
sSo Advanced 2x mastermix	2,5 µl
Forward Primer (5 µM)	0,25 µl
Reverse Primer (5 µM)	0,25 µl
cDNA (2,5 ng/µl)	2 µl

Thermocycling parameters

Step	Temperature	Time	Cycles
Enzyme activation	95 °C	2 min	1
Amplification	95 °C	5 sec	44
	60 °C	30 sec	
	72 °C	1 sec	
Melting cyclus	95 °C	5 sec	1
	60 °C	1 min	
	95 °C	continuous	
Cooling	37 °C	3 min	1

Western blotting

Total protein isolation was performed with RIPA-lysis buffer, supplemented with protease inhibitors and SDS-PAGE was performed according to standard protocols. For immunoblotting, the rabbit polyclonal antibody to cleaved NOTCH1 (Val 1744) (Cell Signaling, 2421S) was used in a 1:500 dilution in BSA.

Protein coding potential calculation

We used PhyloCSF to identify putative protein coding transcripts in the unannotated, novel putative lncRNA loci obtained by RNA-seq. This algorithm employs codon substitution frequencies in whole-genome multi-species alignments to distinguish between coding and non-coding loci. Whole-genome alignments of 46 species are obtained from the UCSC website and processed using the PHAST package (version 1.3) to obtain the required input format for PhyloCSF.

To validate our workflow, we benchmarked PhyloCSF with transcripts annotated in Ensembl (version 75). Transcripts with biotype 'lincRNA' or 'antisense' (20,320 transcripts) serve as a negative set while transcripts with biotype 'protein coding' and an annotated coding sequence (36,959 transcripts) serve as a positive set. Using these sets, we have determined 41.2019 as an optimal treshold for the PhyloCSF score (precision of 95% and sensitivity of 91%).

Micro-array based gene expression profiling

RNA samples from T-ALL cells treated with GSI, $CD34^+$ thymocytes cultured on the OP9-GFP/DLL1 system, sorted T-cell subsets ($CD34^+$ and $CD4^+CD8^+$) and two primary T-ALL patient cohorts of which one cohort including samples with known *NOTCH1* mutation status (n=15) (all *FBXW7* wild type) and a larger cohort (n=64) were profiled on a custom designed Agilent micro-array covering all protein coding genes and 12,000 lncRNAs (23,042 unique lncRNA probes) as described by Volders et al.[46] Profiling was performed according to the manufacturers protocol (One-color Microarray-Based Gene Expression Analysis, Low Input Quick Amp Labeling, Agilent Technologies), with 100 ng RNA as input. Normalization of the expression data was performed with the VSN-package (BioConductor release 2.12) in R. Expression values were further subjected to background subtraction by selecting those probes detecting a 10 % higher expression

than the negative control probes of the array design in at least one treatment. Differential expression analysis was performed in R using Limma. A multifactorial design was used to control for batch effects.

Correlation analysis for functional annotation of selected lncRNAs

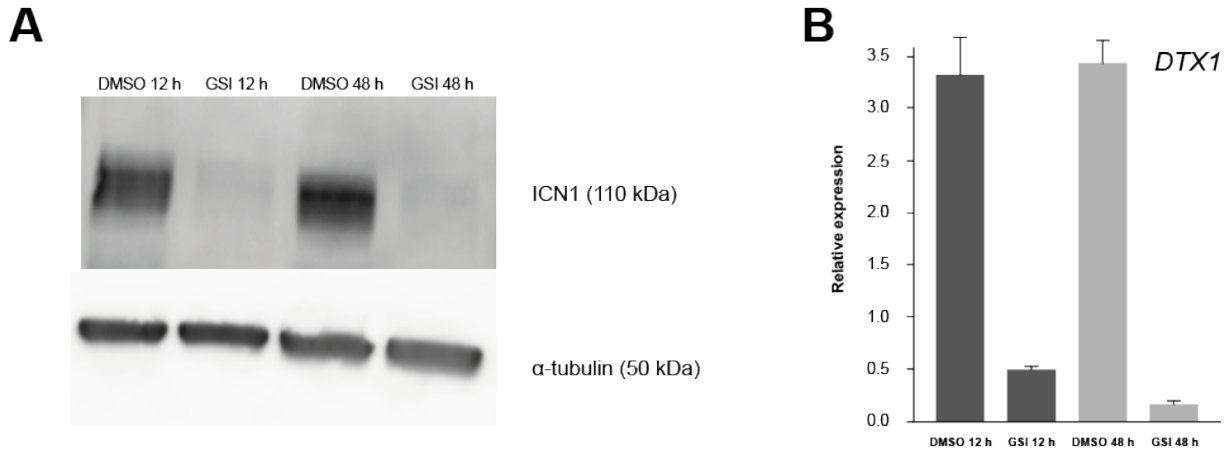
Normalized micro-array based gene expression profiles were generated for all samples of the primary T-ALL patient cohort (n=64). Spearman's rho values were calculated between 15 out of the set of 27 overlapping annotated lncRNAs (Figure 2E) (and also for *LUNAR1* (*lnc-IGF1R-1*)) for which a probe on the custom micro-array[46] was available. This output was used to generate a ranked (.rnk) file and used as an input for a GSEA pre-ranked analysis using the c2v3.1 MsigDB collection as geneset database. The output files were subsequently loaded into Cytoscape. By means of the Cytoscape plug-in 'enrichment mapping' (Isserling et al., F1000Research, 2014), enrichment maps were built representing functional gene set clusters that were significantly correlated (red nodes) or anti-correlated (blue nodes) with the lncRNA-of-interest.

SUPPLEMENTARY TABLES

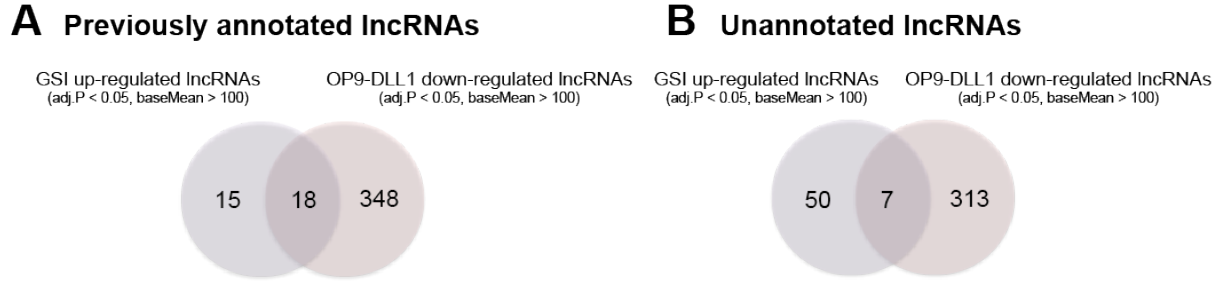
The supplementary tables can be found in the online version of the paper:

http://www.haematologica.org/content/99/12/1808.long

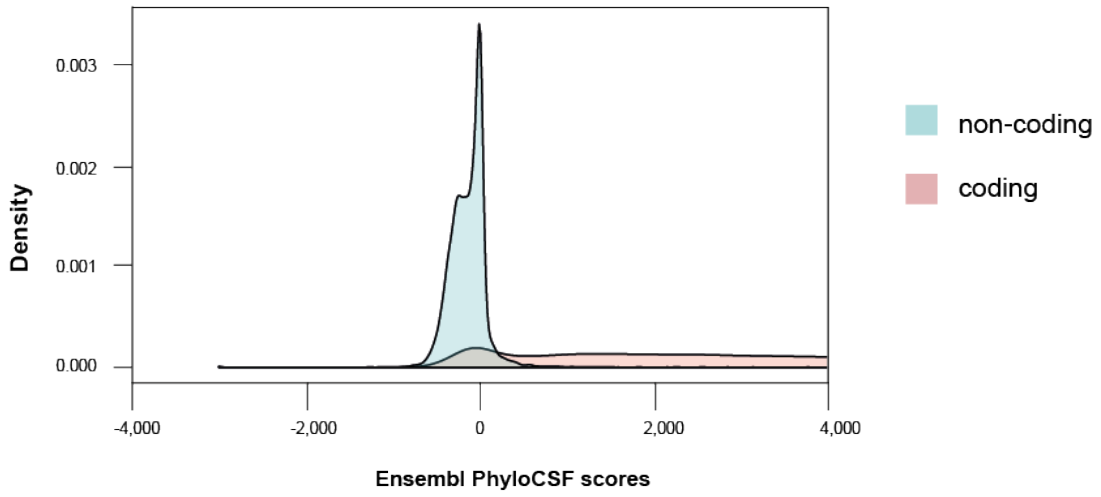
SUPPLEMENTARY FIGURES



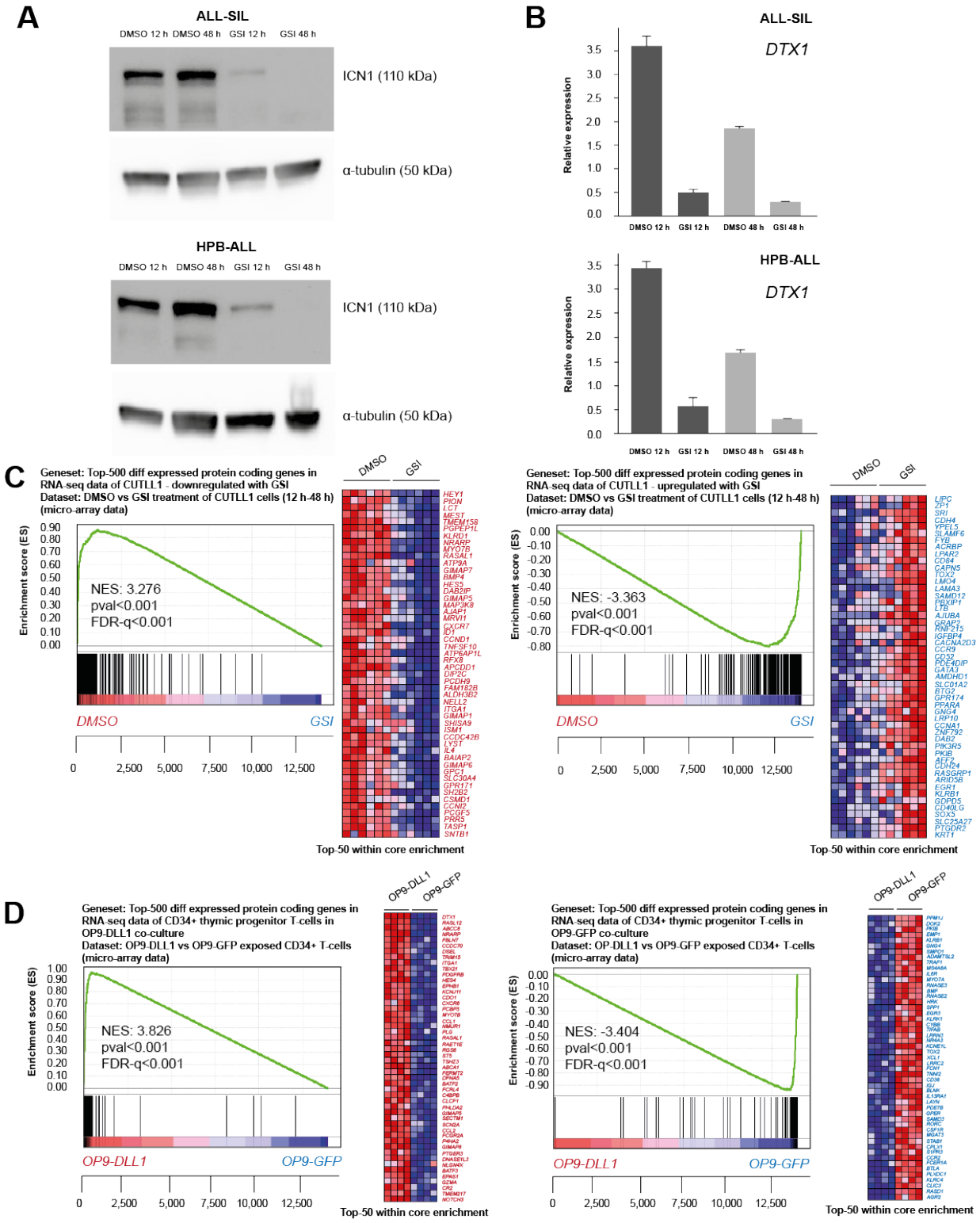
Supplementary Figure 1. Validation of the pharmacological Notch inhibition model in CUTLL1. (**A**) Down-regulation of ICN1 protein levels upon GSI treatment for 12h and 48h in CUTLL1 was validated by Western blot analysis. (**B**) RT-qPCR could confirm down-regulation of *DTX1* expression upon GSItreatment of CUTLL1 T-ALL cells.



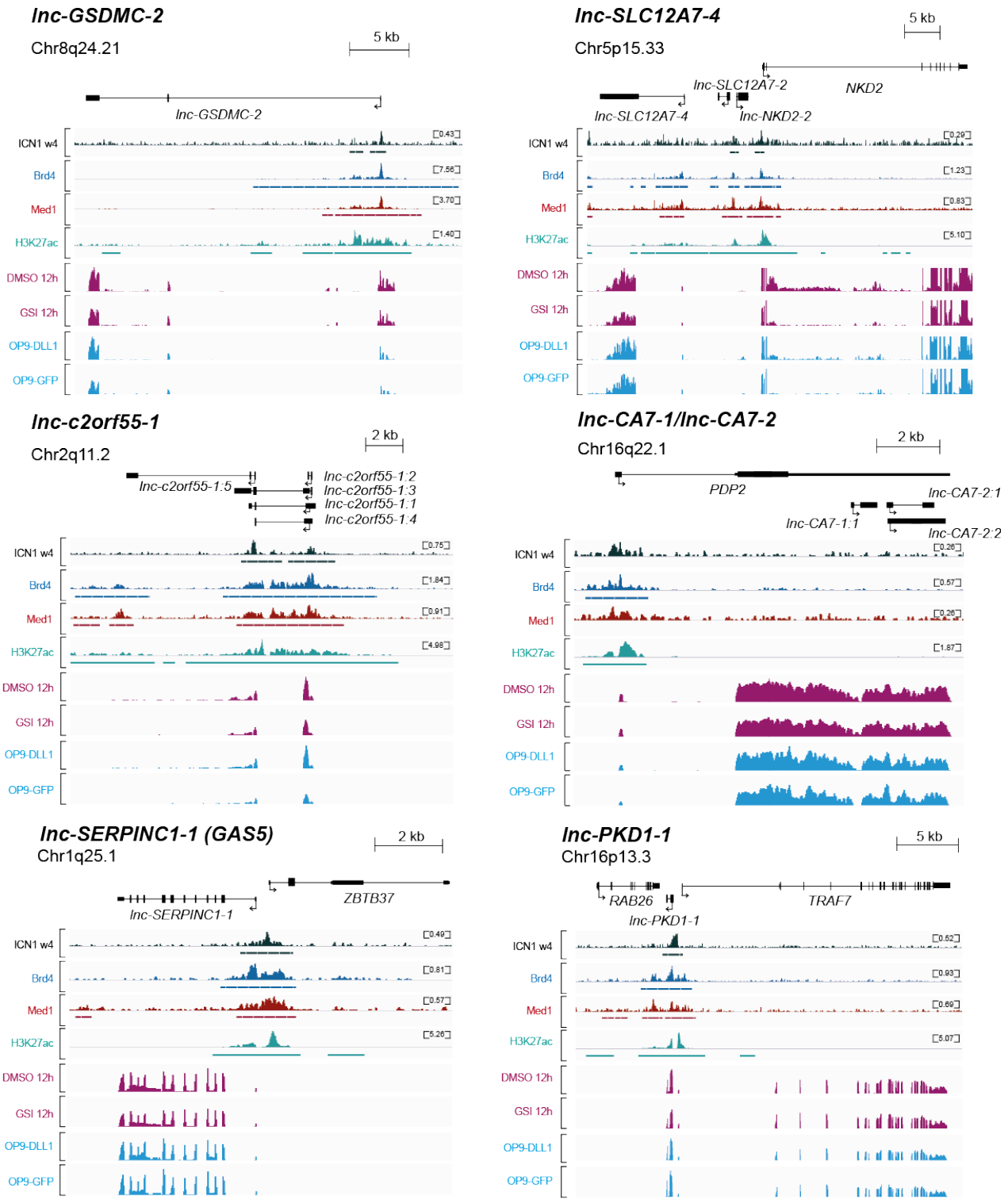
Supplementary Figure 2. Overlap between lncRNAs that are negatively correlated with Notch signaling in GSI treatment of CUTLL1 cells and co-culturing of CD34$^+$ thymocytes on the OP9-DLL1 feeder layer. (**A**) Venn diagram depicting the overlap between previously annotated lncRNAs that are up-regulated upon GSI treatment of the CUTLL1 cell line and down-regulated upon co-culturing of CD34+ thymocytes on the OP9-DLL1 feeder layer. (**B**) Venn diagram depicting the overlap between previously unannotated lncRNAs that are upregulated upon GSI treatment of the CUTLL1 cell line and downregulated upon co-culturing of CD34$^+$ thymocytes on the OP9-DLL1 feeder layer.



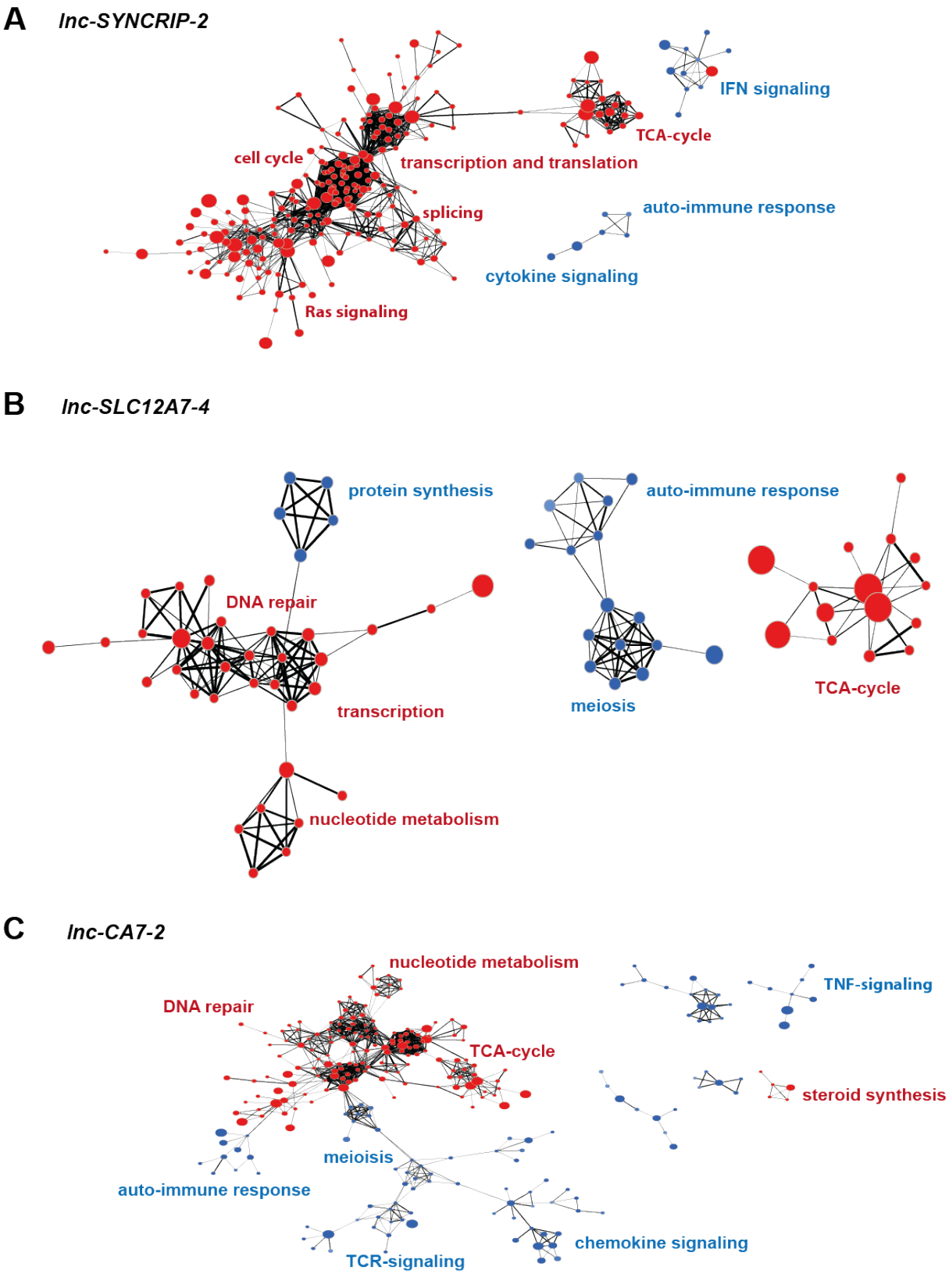
Supplementary Figure 3. PhyloCSF density plot to evaluate the protein coding potential of novel, unannotated lncRNA loci by RNA-seq. The putative protein coding potential of unannotated lncRNA loci in CUTLL1 T-ALL cells and $CD34^{+}$ T-cell progenitors cultured on an OP9 stromal feeder layer was calculated using the PhyloCSF algorithm. The optimal treshold for the PhyloCSF score was determined as 41.2019 to obtain a precision of 95% and sensitivity of 90%.



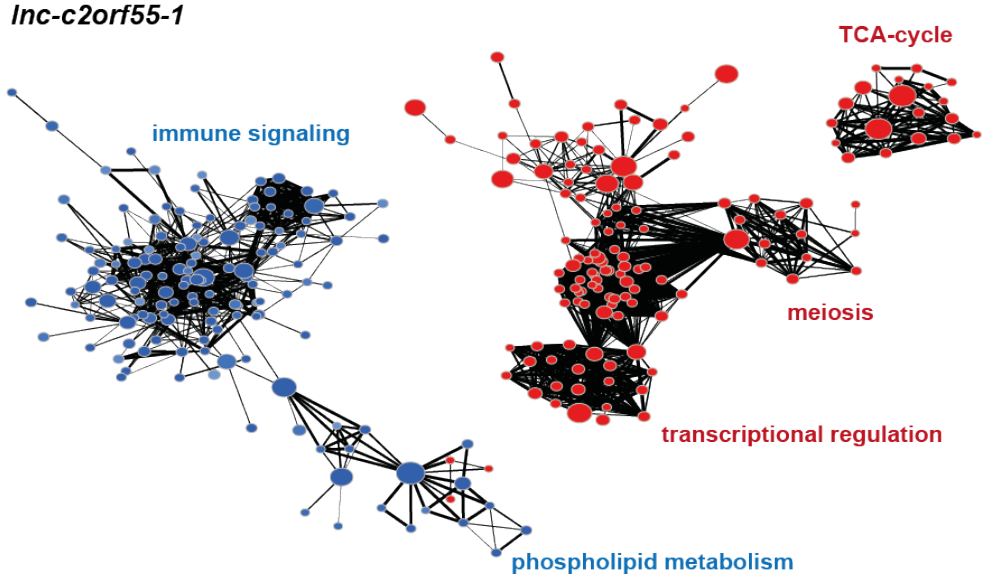
Supplementary Figure 4. Validation of Notch regulated lncRNAs in other model systems. (**A**) Western blot analysis confirms down-regulation of ICN1 in HPB-ALL and ALL-SIL cells upon GSI treatment for 12 h and 48 h. (**B**) RT-qPCR shows *DTX1* down-regulation upon GSI treatment of HPB-ALL and ALL-SIL T-ALL cells. GSEA shows significant overlap for differentially expressed protein coding genes found by RNA-seq and micro-array data of (**C**) GSI-treated CUTLL1 cells and (**D**) $CD34^{+}$ T-cell progenitor on OP9-DLL1/GFP co-cultures.



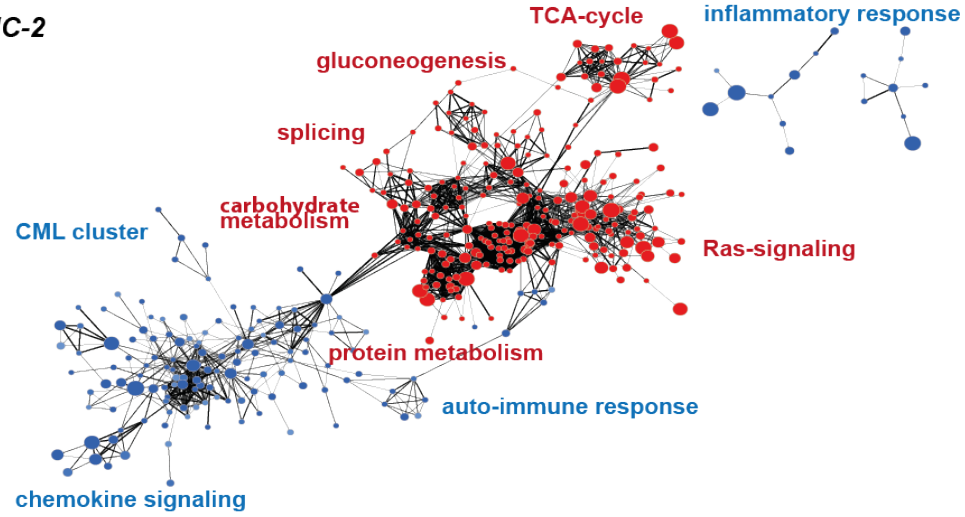
Supplementary Figure 5. Validation of direct regulation of selected lncRNAs by Notch1. ChIP-seq tracks in CUTLL1 cells of Notch1, Brd4, Med1 and H3K27ac are depicted for the selection of annotated lncRNAs that were identified as overlapping Notch1 driven lncRNAs in CUTLL1 cells and $CD34^{+}$ progenitor cells cultured on OP9 stromal cells. Representative RNA-seq tracks are shown for both *in vitro* model systems.


A
lnc-SYNCRIP-2
IFN signaling
TCA-cycle
cell cycle
transcription and translation
splicing
auto-immune response
cytokine signaling
Ras signaling
B
lnc-SLC12A7-4
protein synthesis
auto-immune response
DNA repair
transcription
meiosis
TCA-cycle
nucleotide metabolism
C
lnc-CA7-2
nucleotide metabolism
TNF-signaling
DNA repair
TCA-cycle
steroid synthesis
meioisis
auto-immune response
chemokine signaling
TCR-signaling

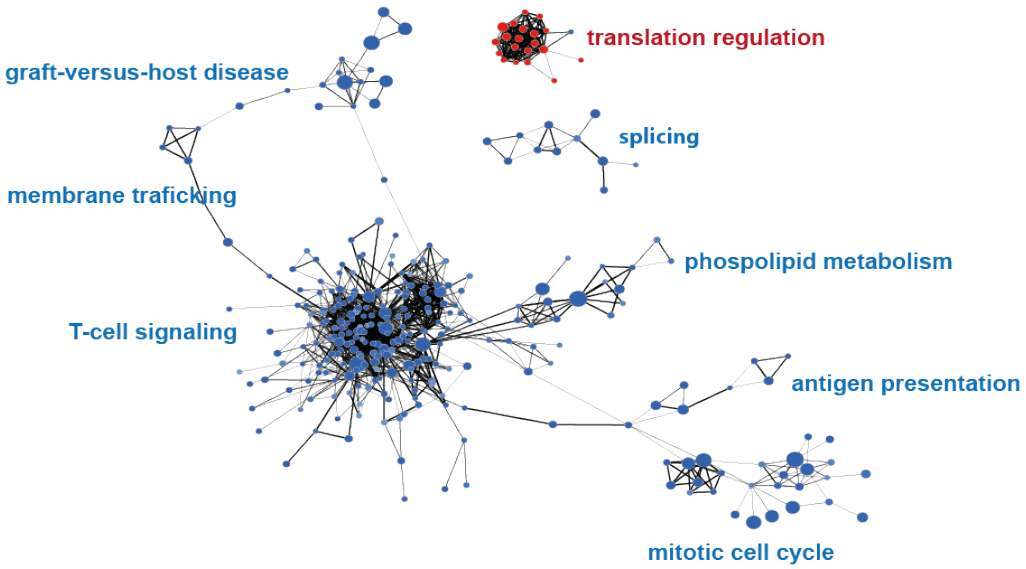
D *lnc-c2orf55-1*



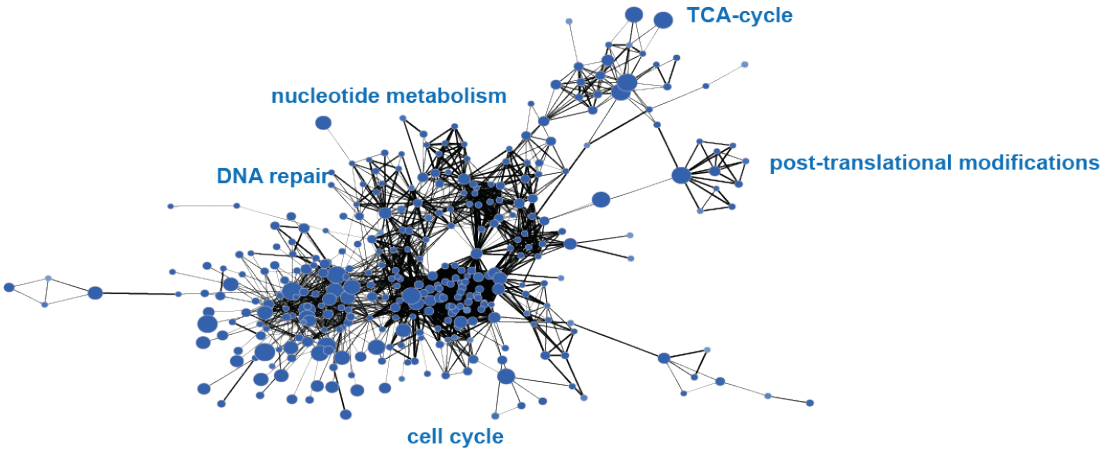
E *lnc-GSDMC-2*



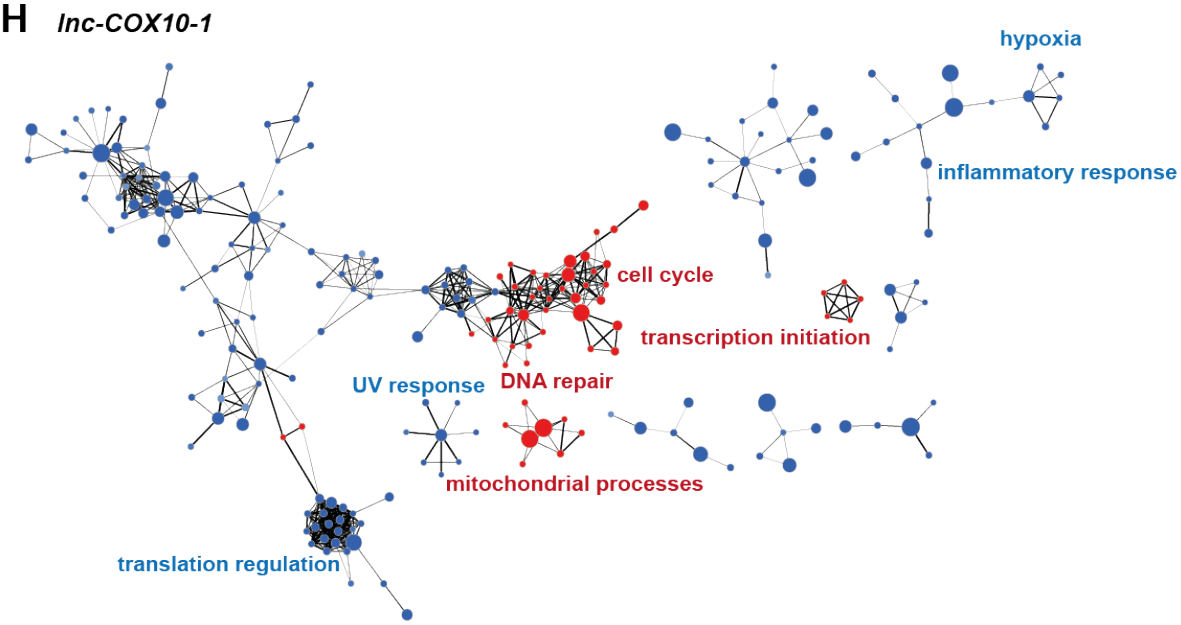
F *lnc-PGBD5-2*



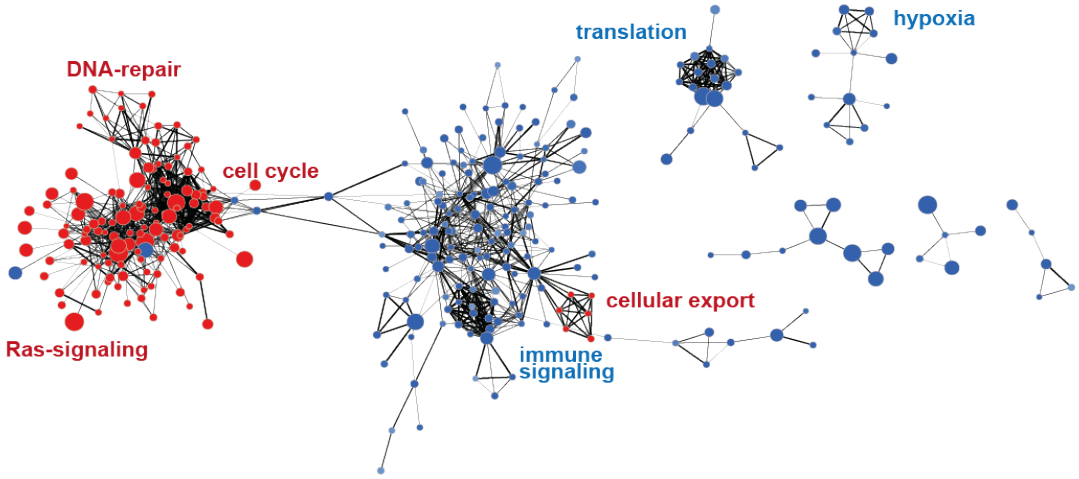
G ***lnc-FAM120AOS-1***



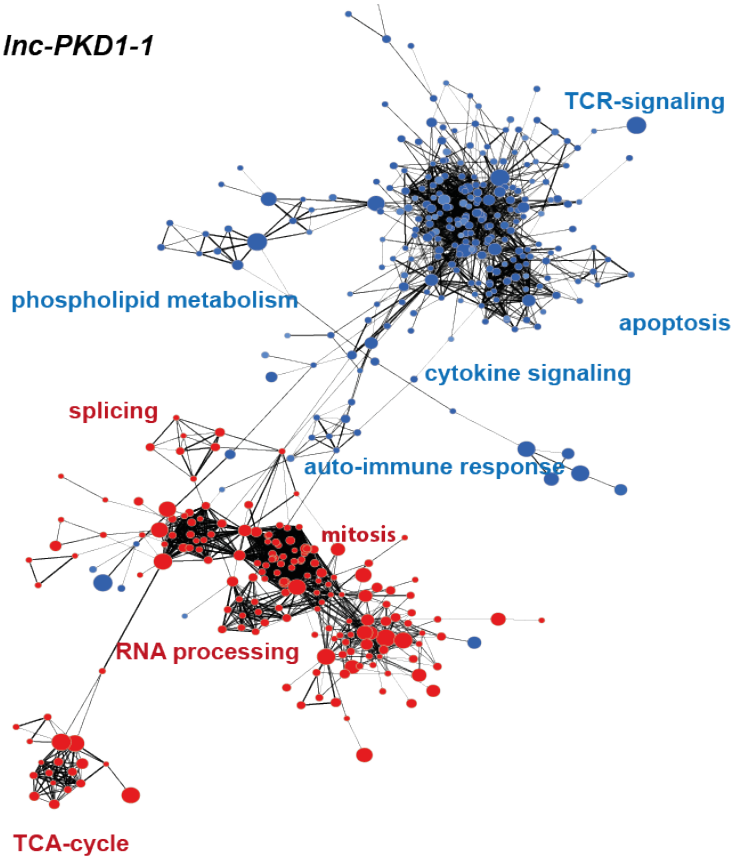
H ***lnc-COX10-1***



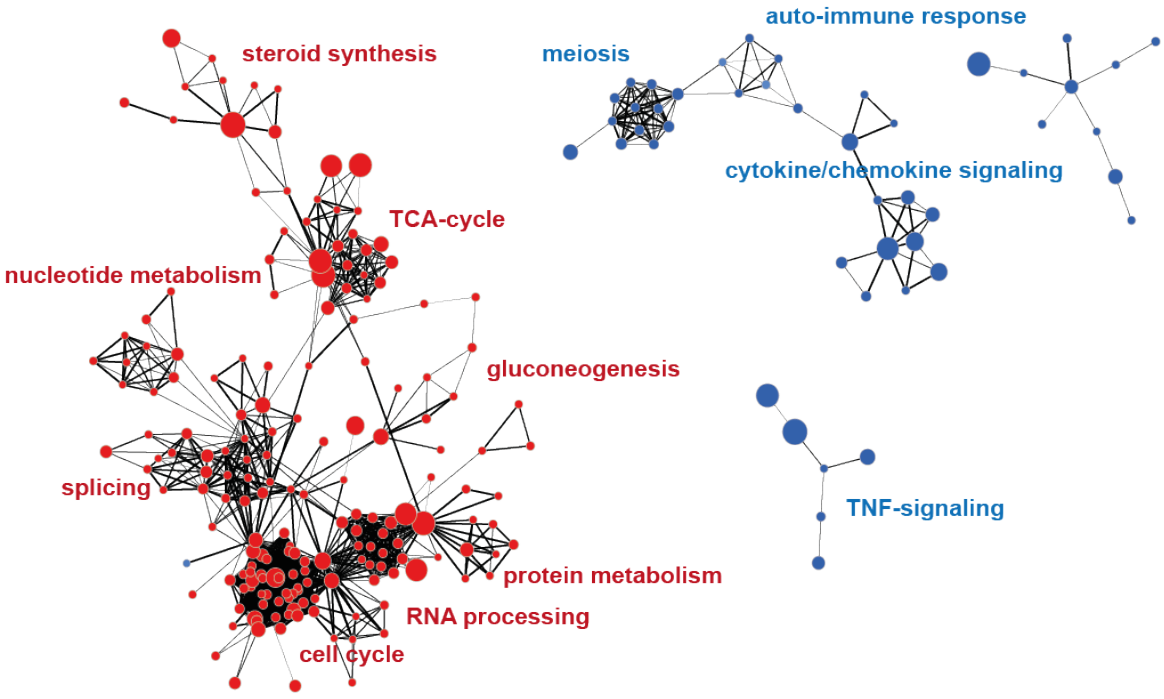
I ***lnc-c9orf163-2***

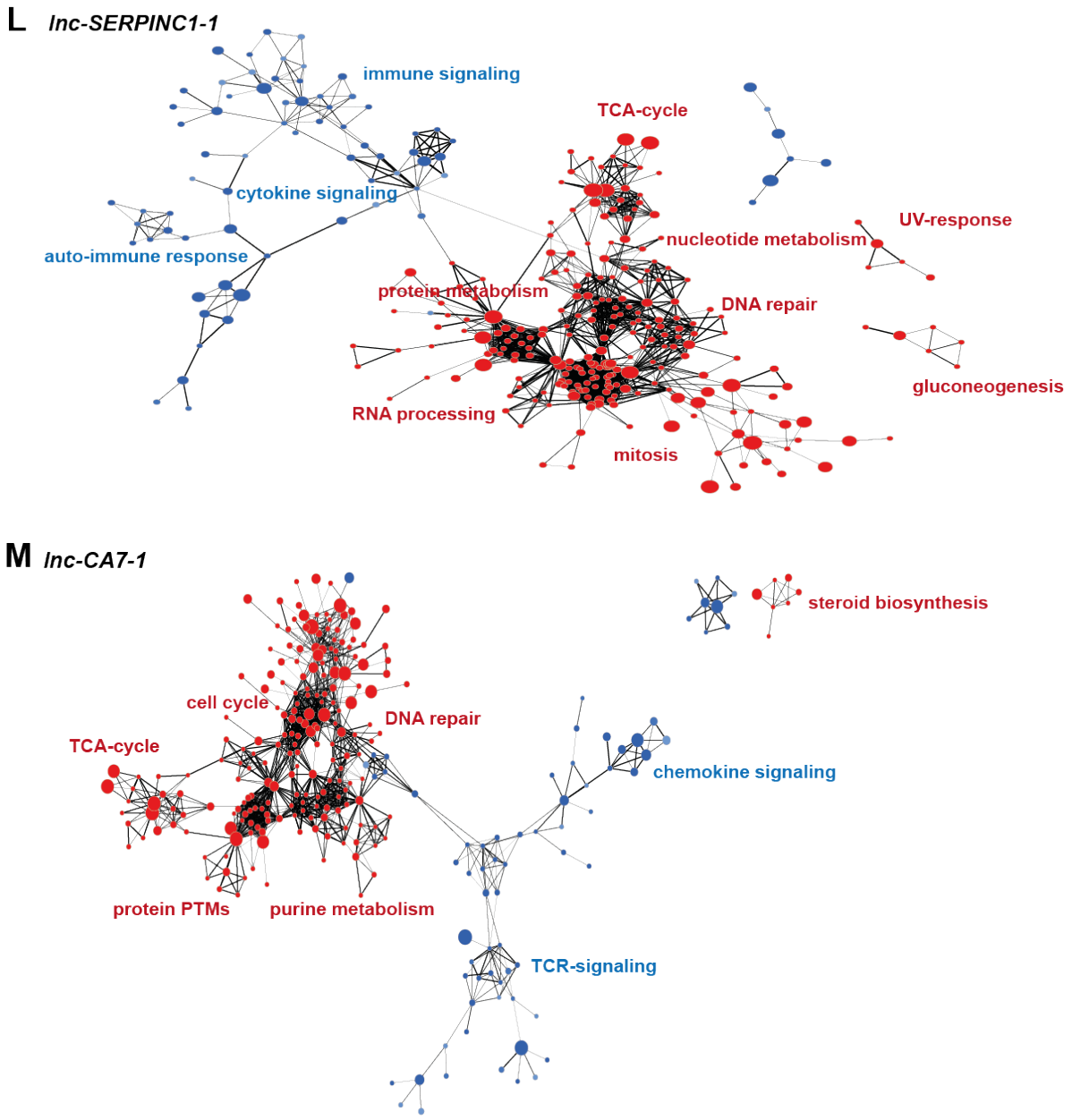


J *lnc-PKD1-1*



K *lnc-GAN-1*





Supplementary Figure 6. Functional annotation of candidate Notch driven lncRNAs in CUTLL1 T-ALL cells and CD34+ thymic progenitor T-cells through enrichment mapping. (**A-M**) Pairwise Spearman's rho correlations were calculated between the selected Notch driven lncRNAs (with probes on the custom array) and all protein coding genes and used for functional annotation of each of the candidate lncRNAs by GSEA. Followingly, enrichment maps were generated in Cytoscape for all selected Notch driven lncRNAs and indicates potential clusters of functionalities linked to each of the candidate lncRNAs such as involvement in the TCA-cycle, meiosis, TCR-signaling. The thickness of the edges represent the overlap of genes between gene sets (nodes).

2 UNRAVELING LONG NONCODING RNA EXPRESSION PATTERNS IN THE DIFFERENT T-ALL SUBTYPES

PAPER 2

Long noncoding RNA signatures define oncogenic subtypes in T-cell acute lymphoblastic leukemia

PAPER 3

RNA-sequencing profiling across T-ALL and thymocyte subsets identifies candidate oncogenic lncRNAs in T-ALL

Long noncoding RNA signatures define oncogenic subtypes in T-cell acute lymphoblastic leukemia

Annelynn Wallaert[1], Kaat Durinck[1], Wouter Van Loocke[1], Inge Van de Walle[2], Filip Matthijssens[1], Pieter-Jan Volders[1], Francisco Avila Cobos[1], Dries Rombaut[1], Pieter Rondou[1], Pieter Mestdagh[1], Jo Vandesompele[1], Bruce Poppe[1], Tom Taghon[2], Jean Soulier[3], Pieter Van Vlierberghe[1*] and Frank Speleman[1*].

[*] *These authors contributed equally to this work.*

[1] *Center for Medical Genetics Ghent, Ghent University, Ghent, Belgium*
[2] *Department of Clinical Chemistry, Microbiology and Immunology, Ghent University, Ghent, Belgium*
[3] *Genome Rearrangements and Cancer Laboratory, U944 INSERM, University Paris Diderot and Hematology Laboratory, Saint-Louis Hospital, Paris, France*

Letter to the Editor

Adapted from Leukemia. 2016 Sep;30(9):1927-30

T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive hematologic malignancy of immature developing precursor T-cells and can be classified into different molecular genetic subgroups based on the aberrant activation of particular transcription factor oncogenes. In addition, these distinct molecular entities display specific gene expression signatures and can be linked to certain stages of T-cell development[1-4]. Most genetic subtypes of human T-ALL are named after the transcription factor oncogene that is predominantly altered in these hematological tumors, i.e. *TAL/LMO*, *TLX1* (*HOX11*), *TLX3* (*HOX11L2*) or *HOXA*. However, immature T-ALLs, which generally lack a unifying molecular genetic alteration, are also considered as a separate T-ALL entity with putative clinical relevance.

Long noncoding RNAs (lncRNAs) are a novel class of untranslated RNAs that are at least 200 nucleotides in size and are implicated in a wide variety of cellular functions and specific developmental processes. Notably, recent studies have shown that lncRNAs can drive tumor development[5, 6] and, in the context of T-ALL, it has been described that NOTCH1 regulates the expression of several lncRNAs[7, 8]. One interesting example is *LUNAR1*, which enhances the expression of IGF1R, leading to sustained IGF1 signaling[8]. Thus far, lncRNAs remained unexplored as genetic markers for distinct T-ALL subtypes.

In this study, we defined the pattern of lncRNA expression in different molecular genetic subtypes of human T-ALL. Integration of these signatures with lncRNA expression in T-ALL cell lines and specific stages of normal human T-cell development provides a resource for the identification of oncogenic or tumor suppressive lncRNAs in the context of T-cell transformation.

To study the role of lncRNAs in the molecular pathogenesis of human T-ALL, we profiled a cohort of 64 T-ALL patient samples[9] using a custom microarray platform detecting 13 000 lncRNAs in addition to the protein coding mRNAs[10]. This cohort consisted of RNA of 15 immature, 17 *TLX1/TLX3*, 25 *TAL-R* and 7 *HOXA* positive T-ALL patients. After VSN-normalization, background subtraction was executed by retaining probes that showed a 10 % higher expression level as compared to the negative control probes in at least 60 % of the samples of one particular molecular genetic subgroup. In keeping with previous reports[1, 2], distinct molecular genetic subclasses showed a unique mRNA gene expression signature (Figure 1A, left panel). Next, we used the 5190 lncRNAs that showed expression in the T-ALL patient cohort to define subtype specific lncRNA signatures of human T-ALL. The top 50 differentially expressed lncRNAs for each subgroup are represented in Figure 1A (right panel). Of note, it was not possible to select 50 differentially expressed lncRNAs for the *HOXA* subgroup, probably due to the small amount of patient samples (n=7). Forty percent of all subtype specific lncRNAs were characterized as 'lincRNA' (long intergenic noncoding RNAs) in Ensembl, whereas 33 % were denoted as 'antisense' (25 % had no Ensembl ID or no biotype was specified). Roughly half of the lncRNAs showed positive correlation in expression levels with one or more protein coding gene(s) in their vicinity (500 kb up- or downstream of the transcription start site; Spearman rho value > 0.5; see Supplemental Table). In contrast, only three lncRNAs were

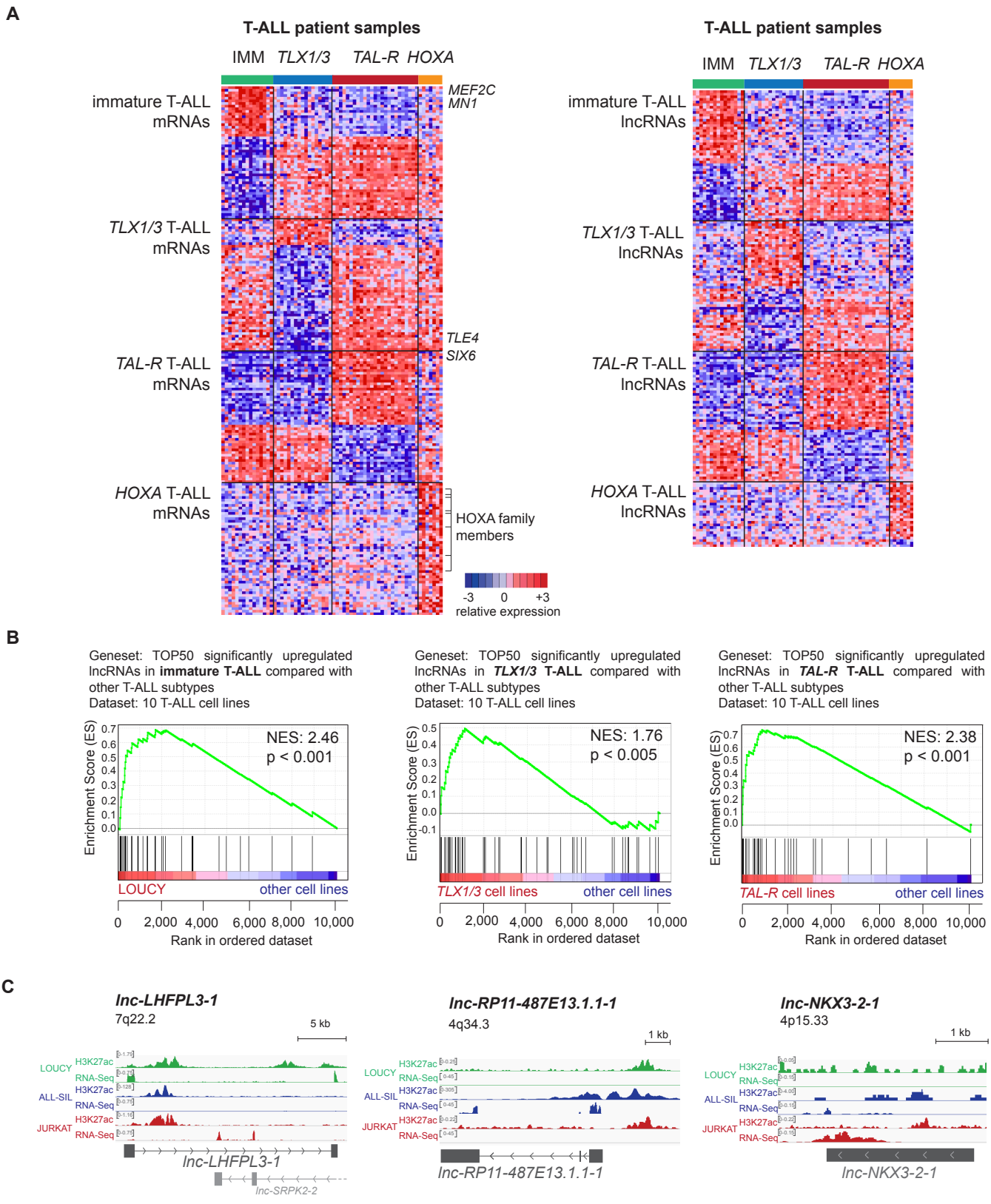


Figure 1. T-ALL subgroups can be characterized by a specific lncRNA expression pattern. (A) Heatmaps showing the top 50 most differentially expressed mRNAs (left) and lncRNAs (right) for each subgroup (adjusted p-value < 0.05). **(B)** Gene set enrichment analysis using the top 50 lncRNAs upregulated in the respective genetic T-ALL subgroups as gene sets within a panel of T-ALL cell lines. **(C)** For each subgroup, one lncRNA from the core enrichment of the cell lines is visualized with RNA-sequencing and H3K27ac ChIP-sequencing tracks for LOUCY, ALL-SIL and JURKAT.

negatively correlated with a neighboring protein coding gene.

In order to identify suitable *in vitro* model systems to study the functional role of lncRNAs in human T-ALL, we also performed lncRNA expression profiling in a subset of 10 human T-ALL cell lines. These tumor lines genetically resemble most of the different molecular subtypes of human T-ALL and included immature (LOUCY), *TLX1/TLX3* positive (ALL-SIL, DND41 and HPB-ALL) and *TAL-R* (PF-382, JURKAT, KE-37 and CCRF-CEM) tumor lines. Next, we used gene set enrichment analysis[11] to evaluate the potential overlap between the lncRNA signatures from primary T-ALL patient samples and the established tumor cell lines (Figure 1B and Supplemental Figure 1). LncRNAs highly expressed in immature T-ALL patients, show a significant enrichment in the immature T-ALL cell line LOUCY as compared to the other cell lines. Similarly, *TLX1/3* specific lncRNAs were enriched in ALL-SIL, DND-41 and HPB-ALL, whereas *TAL-R* associated lncRNAs were predominantly found in PF-382, JURKAT, KE-37 and CCRF-CEM. Similar results were obtained for downregulated lncRNAs and for protein coding genes (Supplemental Figure 1 and 2). For each subgroup, a representative lncRNA is visualized by means of RNA-Seq and H3K27ac ChIP-Seq data of the LOUCY, ALL-SIL and JURKAT T-ALL cell lines (Figure 1C)[12-14]. Furthermore, the publically available RNA-sequencing of 8 T-ALL cell lines (LOUCY, DND-41, HPB-ALL, ALL-SIL, PF-382, JURKAT, KE-37 and CCRF-CEM)[12] has been remapped for lncRNAs and subsequently GSEA has been performed for the top selected mRNAs and lncRNAs that were catalogued in the Ensembl database. This revealed that, except the *TLX1/3* upregulated lncRNAs, all subgroup specific gene sets were enriched in the RNA-sequencing data of the corresponding T-ALL cell lines of that particular subtype (Supplemental Figure 3), validating our microarray-based data by RNA-sequencing. We hereby conclude that lncRNA signatures in T-ALL cell lines mimic those observed in primary leukemia samples indicating that these cell lines represent reliable *in vitro* model systems to study the role of these lncRNAs in the pathogenesis of this disease.

To identify putative oncogenic lncRNAs in T-ALL development, we collected sorted subsets of healthy thymocytes (Supplemental Methods). Normal thymopoiesis is a carefully regulated process with each of the intermediate stages being characterized by a distinct transcriptional profile and a combination of surface markers allowing for selection of these maturing T-cell subsets[15]. During this maturation process, several oncogenic lesions can occur, blocking differentiation and driving these immature thymocytes into an uncontrolled expansion towards full-blown T-ALL. The above-described T-ALL subgroups largely resemble an arrest of thymocytes at specific stages of T-cell development. More specifically, the immature T-ALL patients have an arrest in the early stages of T-cell development, i.e. $CD34^+$ thymocytes. In contrast, the leukemic cells of *TAL-R* patients resemble later stages of T-cell development, typically at the $CD4^+CD8^+$ double positive stage. Given that the transcriptomes of the different genetic T-ALL subtypes strongly recapitulate those of normal immature thymocytes, we also explored the lncRNAome of 6 $CD34^+$ thymocyte populations and 7 $CD4^+CD8^+$ thymocyte samples obtained from 4 independent donors.

Interestingly, 24 out of the top 50 upregulated lncRNAs in immature T-ALL (Supplemental

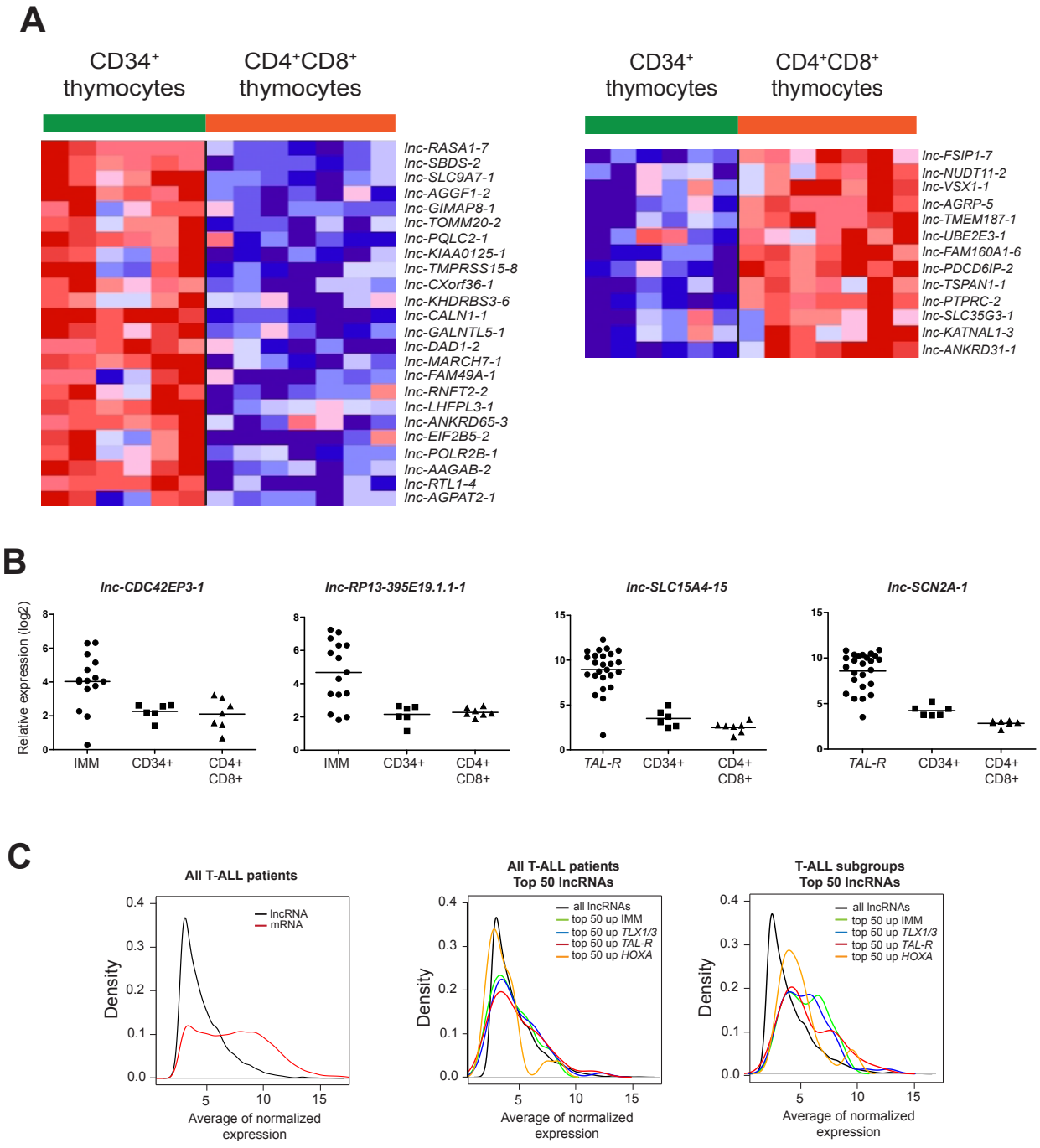


Figure 2. Comparison with healthy thymocytes reveals lncRNAs with a role in T-cell development and potentially oncogenic lncRNAs. (A) (left) Heatmap representing lncRNAs of the top 50 lncRNAs upregulated in the immature T-ALL patients that are also significantly higher expressed in $CD34^+$ thymocytes compared with $CD4^+CD8^+$ thymocytes (adjusted p-value < 0.05). (right) Heatmap representing lncRNAs of the top 50 upregulated lncRNAs in the *TAL-R* group that are significantly higher expressed in $CD4^+CD8^+$ thymocytes compared with $CD34^+$ thymocytes (adjusted p-value < 0.05). **(B)** Boxplots representing 2 lncRNAs that are ectopically expressed in immature T-ALL patients compared with healthy thymocytes and 2 lncRNAs that are ectopically expressed in *TAL-R* T-ALL patients compared with healthy thymocytes. **(C)** Density plots representing the average of the normalized expression of all lncRNAs and mRNAs detected by the array in all patients (left), the top 50 upregulated lncRNAs for each subgroup in all patients (middle) and the 50 upregulated lncRNAs for each subgroup in patients of that particular subgroup (right).

Figure 4A) were also significantly higher expressed in the $CD34^+$ thymocytes as compared to $CD4^+CD8^+$ T-cells (Figure 2A, left), providing a link between immature T-ALLs and their cell of origin. In contrast, 18 of the top 50 showed significant higher expression levels in immature T-ALL patients as compared to $CD34^+$ thymocytes, suggesting that these lncRNAs might act as putative oncogenes in the biology of this disease (Figure 2B and Supplemental Figure 5A). Ten of these 18 lncRNAs were also in the core enrichment of the immature T-ALL cell line LOUCY, making them top candidates for further in depth study.

A similar analysis was performed for the *TAL-R* subgroup in relation to $CD4^+CD8^+$ healthy donor counterparts. Thirteen lncRNAs of the top 50 upregulated in *TAL-R* T-ALLs (Supplemental Figure 4B) were significantly higher expressed in the $CD4^+CD8^+$ thymocytes as compared to $CD34^+$ thymocytes, reflecting a role during normal T-cell development (Figure 2A, right). On the other hand, 29 lncRNAs of the selection are potentially oncogenic, because they are signifycantly higher expressed in the *TAL-R* T-ALL patients compared with the $CD4^+CD8^+$ thymocytes (Figure 2B and Supplemental Figure 5B), of which 9 are also in the core enrichment of the *TAL-R* T-ALL cell lines.

We could also detect possible tumor suppressive lncRNAs by performing the reverse analysis. Sixteen lncRNAs of the top 50 downregulated lncRNAs in the immature patients were significantly higher expressed in the $CD34^+$ thymocytes, whereas 15 lncRNAs of the top 50 downregulated lncRNAs in the *TAL-R* patients were significantly higher expressed in the $CD4^+CD8^+$ thymocytes (Supplemental Figure 6). The same analyses have been performed for protein coding genes, which led to similar conclusions. These results can be found in Supplemental Figures 7 – 10.

LncRNAs are known to have a lower expression level as compared to mRNAs, this also holds true for our data as can be seen in the distribution plot in Figure 2C (left). The average normalized expression of the top 50 upregulated lncRNAs for each subgroup follows the pattern of all lncRNAs detected by the array (Figure 2C, middle). However, the expression patterns of the top 50 upregulated lncRNAs for each subgroup in the patients from that specific subgroup is shifted to higher levels, suggesting that these lncRNAs might be truly functional in the context of that particular subgroup (Figure 2C, right).

In conclusion, this study provides a first landscaping of lncRNAs that can discriminate the distinct T-ALL subgroups and provides a resource for further biomarker discovery and novel targets for patient-stratified therapy.

SUPPLEMENTAL DATA

The supplemental data contain a detailed description of the materials and methods that were used, extra figures and tables of the performed differential analysis and correlation analysis.

ACCESS TO DATA

The cell line expression data and ChIP-Seq data of the LOUCY cell line has been deposited in GEO (GSE74312).

ACKNOWLEDGEMENTS

The authors would like to thank the following funding agencies: the Fund for Scientific Research Flanders ('FWO Vlaanderen' research

projects 3G056413N and 3G055013N to F.S. G065614, 3GA00113N and G.0C47.13N to P.V.V. and G037514N and G0B2913 to T.T.; 'Kom op tegen Kanker' project G0E3514N to F.S.; PhD grant to A.W.; postdoctoral grants to I.v.d.W., P.R. and P.M.), IWT Vlaanderen (PhD grant to D.R.), the Belgian Foundation against Cancer (F.S.), Ghent University Research Fund (GOA GENT-Lung (F.A.C.)), the Cancer Plan from the Federal Public Service of Health (F.S.), The Children Cancer Fund Ghent (F.S.), the Flemish League against Cancer (PhD grant to K.D., postdoctoral grant to F.M.), the Belgian Program of Interuniversity Poles of Attraction (IUAP P7/03) and the Multidisciplinary Research Partnership 'Bioinformatics: From Nucleotides to Networks' Project of Ghent University (01MR0310W to P.V.). The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, the Hercules Foundation and the Flemish Government department EWI. We also would like to thank Aline Eggermont for excellent technical assistance.

AUTHORSHIP

Contribution: A.W. performed and analyzed experiments and wrote the paper; K.D., W.V.L., I.V.d.W., F.M., F.A.C. and D.R. performed experiments; P.V., P.M. and J.V. designed the array and set-up the analysis; T.T. and J.S. provided samples; P.R., B.P., P.V.V. and F.S. designed experiments, directed research and wrote the paper. All the authors read and edited the manuscript.

Conflict-of-interest disclosures: The authors declare no competing financial interests.

Correspondence: Frank Speleman; e-mail: franki.speleman@ugent.be; Annelynn Wallaert; e-mail: annelynn.wallaert@ugent.be

REFERENCES

1. Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC, et al. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer Cell. 2002;1(1):75-87.
2. Soulier J, Clappier E, Cayuela JM, Regnault A, Garcia-Peydro M, Dombret H, et al. HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). Blood. 2005;106(1):274-86.
3. Van Vlierberghe P, Pieters R, Beverloo HB, Meijerink JP. Molecular-genetic insights in paediatric T-cell acute lymphoblastic leukaemia. Br J Haematol. 2008;143(2):153-68.
4. Meijerink JP. Genetic rearrangements in relation to immunophenotype and outcome in T-cell acute lymphoblastic leukaemia. Best Pract Res Clin Haematol. 2010;23(3):307-18.
5. Maass PG, Luft FC, Bahring S. Long non-coding RNA in health and disease. J Mol Med (Berl). 2014;92(4):337-46.
6. Gutschner T, Diederichs S. The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol. 2012;9(6):703-19.
7. Durinck K, Wallaert A, Van de Walle I, Van Loocke W, Volders PJ, Vanhauwaert S, et al. The Notch driven long non-coding RNA repertoire in T-cell acute lymphoblastic leukemia. Haematologica. 2014;99(12):1808-16.
8. Trimarchi T, Bilal E, Ntziachristos P, Fabbri G, Dalla-Favera R, Tsirigos A, et al. Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. Cell. 2014;158(3):593-606.
9. Clappier E, Gerby B, Sigaux F, Delord M, Touzri F, Hernandez L, et al. Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. J Exp Med. 2011;208(4):653-61.
10. Volders PJ, Helsens K, Wang X, Menten B, Martens L, Gevaert K, et al. LNCipedia: a database for annotated human lncRNA transcript

sequences and structures. Nucleic Acids Res. 2013;41(Database issue):D246-51.
11. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.
12. Atak ZK, Gianfelici V, Hulselmans G, De Keersmaecker K, Devasia AG, Geerdens E, et al. Comprehensive analysis of transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. PLoS Genet. 2013;9(12):e1003997.
13. Durinck K, Van Loocke W, Van der Meulen J, Van de Walle I, Ongenaert M, Rondou P, et al. Characterization of the genome-wide TLX1 binding profile in T-cell acute lymphoblastic leukemia. Leukemia. 2015;29(12):2317-27.
14. Kwiatkowski N, Zhang T, Rahl PB, Abraham BJ, Reddy J, Ficarro SB, et al. Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. Nature. 2014;511(7511):616-20.
15. Taghon T, Waegemans E, Van de Walle I. Notch signaling during human T cell development. Curr Top Microbiol Immunol. 2012;360:75-97.

SUPPLEMENTARY DATA

Study design

Bone marrow lymphoblast and blood samples from T-ALL patients were collected after informed consent according to the Declaration of Helsinki from Saint-Louis Hospital, Paris, France. This study was approved by the Institut Universitair d'Hématologie Institutional Review Board and the Ethical committee of Ghent University Hospital. This primary T-ALL cohort was previously investigated[1] and the high-quality RNA samples were used for lncRNA micro-array based expression profiling[2]. Subgroup annotation of the T-ALL patient samples can be found in the table below.

Thymocytes were purified from healthy donors. CD34 MACS purified thymocytes were labeled with CD34, CD1 and CD4 to sort CD34+CD1-CD4- uncommitted and CD34+CD1+CD4- committed early thymocytes. A total thymus suspension was labeled with CD4, CD8 and CD3 to sort CD4+CD8+CD3- and CD4+CD8+CD3+ double positive thymocytes.

The T-ALL cell lines LOUCY, KARPAS-45, PEER, DND-41, ALL-SIL, HPB-ALL, PF-382 and JURKAT were purchased from DSMZ. KE-37 and CCRF-CEM were a kind gift of respectively the Cools Lab and the Barrata lab.

Total RNA from T-ALL patient samples, sorted human thymocytes and T-ALL cell lines was isolated using the miRNeasy mini kit (Qiagen) and the RNA quality was evaluated on Experion (Bio-Rad). RNA samples (100 ng) were profiled on a custom designed microarray platform covering all protein coding genes and 13,000 lncRNAs as described by Volders et al.[3] and according to the manufacturers instructions with oligo-dT primers for cDNA synthesis. Normalization of the data was performed with the VSN-package (BioConductor release 3.1) in R. Background subtraction in the T-ALL patient cohort was executed by retaining probes that showed a 10 % higher expression level as compared to negative control probes in at least 60 % of the samples of one subgroup. Differential expression analysis was performed in R using Limma. Correlation analysis between protein coding genes and lncRNAs was performed by means of Spearman correlation. Expression profiling data have been deposited in the NCBI Gene Expression Omnibus[4] and are accessible through GEO accession numbers GSE61866 (T-ALL patient samples[2]), GSE61873 (sorted human thymocytes[2]) and GSE74312 (T-ALL cell lines). The results of the differential expression analysis and the correlation analysis can be found in supplemental tables.

T-ALL subgroup annotation for the samples used in Durinck et al., 2014 (GSE61866) [2]

Sample	Subgroup	Sample	Subgroup	Sample	Subgroup
TALL_JS_1	TAL-R	TALL_JS_23	HOXA	TALL_JS_45	TLX
TALL_JS_2	IMM	TALL_JS_24	TAL-R	TALL_JS_46	IMM
TALL_JS_3	TLX	TALL_JS_25	TAL-R	TALL_JS_47	TAL-R
TALL_JS_4	IMM	TALL_JS_26	TLX	TALL_JS_48	TAL-R
TALL_JS_5	IMM	TALL_JS_27	IMM	TALL_JS_49	TAL-R
TALL_JS_6	HOXA	TALL_JS_28	TLX	TALL_JS_50	TAL-R
TALL_JS_7	IMM	TALL_JS_29	IMM	TALL_JS_51	TAL-R
TALL_JS_8	TAL-R	TALL_JS_30	IMM	TALL_JS_52	TAL-R
TALL_JS_9	TLX	TALL_JS_31	TLX	TALL_JS_53	TAL-R
TALL_JS_10	TLX	TALL_JS_32	TAL-R	TALL_JS_54	TAL-R
TALL_JS_11	HOXA	TALL_JS_33	IMM	TALL_JS_55	TLX
TALL_JS_12	TAL-R	TALL_JS_34	TAL-R	TALL_JS_56	TLX
TALL_JS_13	TAL-R	TALL_JS_35	HOXA	TALL_JS_57	TLX
TALL_JS_14	HOXA	TALL_JS_36	TAL-R	TALL_JS_58	TLX
TALL_JS_15	TAL-R	TALL_JS_37	TLX	TALL_JS_59	TLX
TALL_JS_16	TAL-R	TALL_JS_38	IMM	TALL_JS_60	TLX
TALL_JS_17	IMM	TALL_JS_39	TLX	TALL_JS_61	HOXA
TALL_JS_18	IMM	TALL_JS_40	HOXA	TALL_JS_62	TLX
TALL_JS_19	TLX	TALL_JS_41	TAL-R	TALL_JS_63	IMM
TALL_JS_20	IMM	TALL_JS_42	TAL-R	TALL_JS_64	TAL-R
TALL_JS_21	TAL-R	TALL_JS_43	TAL-R		
TALL_JS_22	IMM	TALL_JS_44	TAL-R		

H3K27ac ChIP-sequencing in LOUCY

The LOUCY cell line was obtained from the DSMZ cell line repository. Cells were maintained in RPMI-1640 medium supplemented with 10 % fetal bovine serum, 1% of L-glutamine and 1 % of penicillin/streptomycin. The ChIP-protocol has been performed as described by Durinck et al.[5] In brief, 10 million cells were cross-linked with 1,1% formaldehyde for 10 min at room temperature. The cross-linking reaction was quenched with glycine (125 mM final concentration). Nuclei were isolated and the chromatin was purified by chemical lysis. Next, purified chromatin was fragmented by sonication to 200-300 bp fragments (Covaris). Chromatin immunoprecipitation was performed by incubation of the chromatin fraction overnight with 100 µl of protein-A coated beads (Thermo-Scientific, catalog number 53139) and 10 µg of H3K27ac-specific antibody (Abcam, ab4729). The next day, beads were washed to remove non-specific binding events and the enriched chromatin fragments were eluted from the beads, followed by reverse cross-linking by overnight incubation at 65°C. DNA was purified by phenol/chloroform extraction, assisted by phase lock gel tubes (5Prime). DNA obtained from the ChIP-assays was adaptor-ligated, amplified and analyzed by Illumina Hiseq 2000. Raw sequencing data was mapped to the human reference genome (GRCh37/h19) using Bowtie. Peak calling was performed using MACS. ChIP seq data has been deposited in the GEO database (GSE74312).

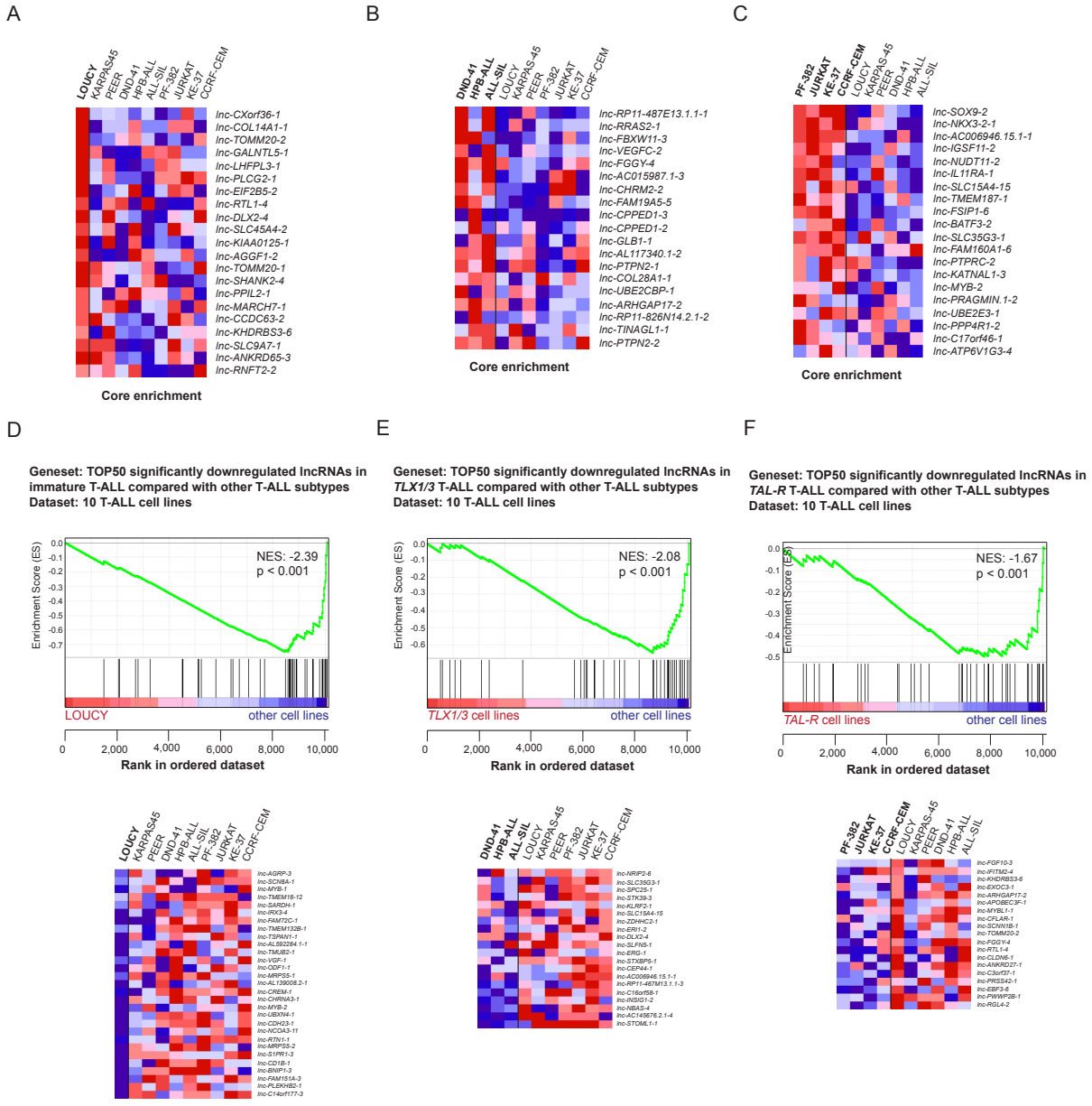
Supplemental tables

Supplementary tables can be downloaded through the Leukemia website or directly via: http://www.nature.com/leu/journal/vaop/ncurrent/suppinfo/leu201682s1.html?url=/leu/journal/vaop/ncurrent/full/leu201682a.html

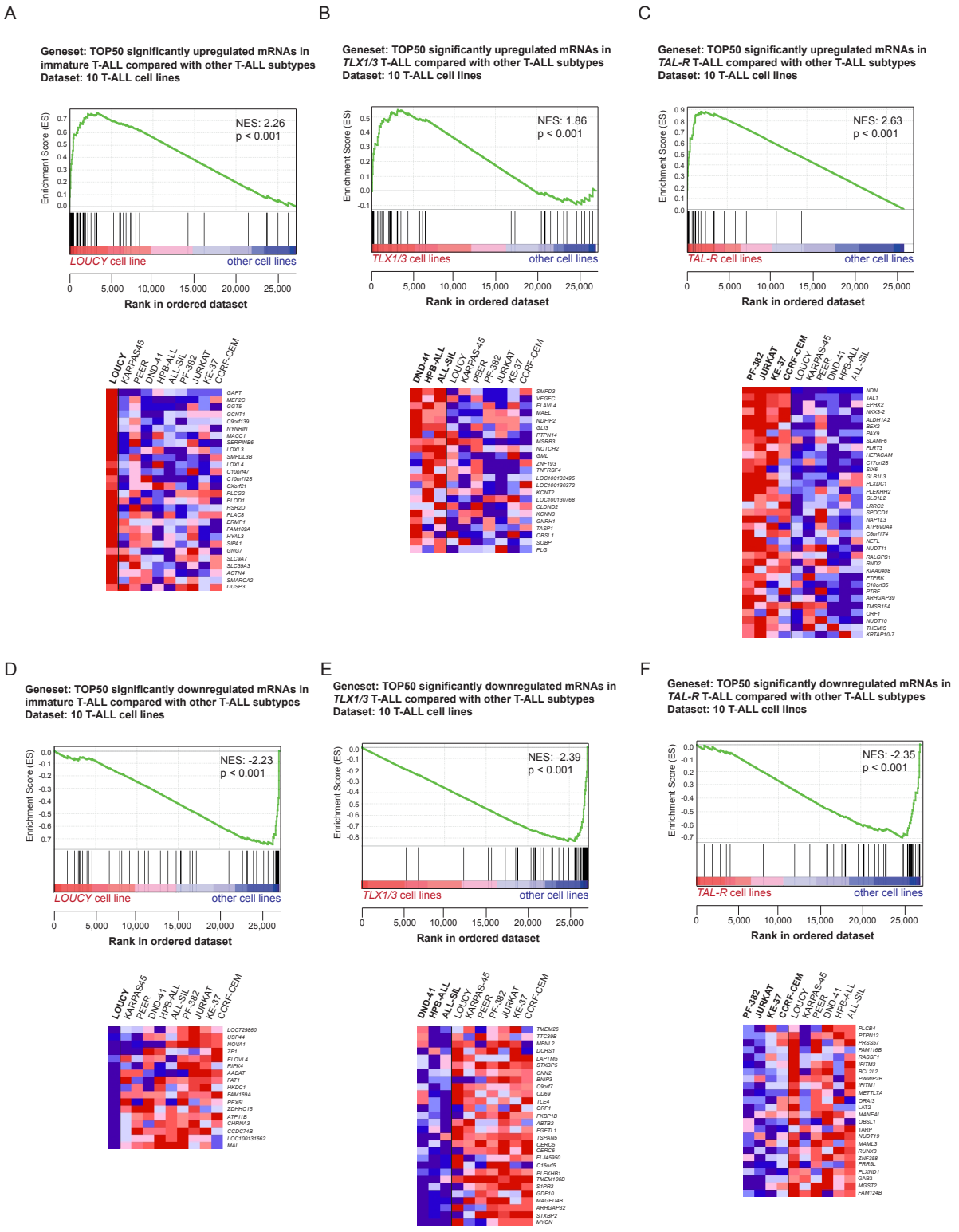
Supplemental references

1. Clappier E, Gerby B, Sigaux F, et al. Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. *J Exp Med*. 2011;**208**(4):653-61.
2. Durinck K, Wallaert A, Van de Walle I, et al. The Notch driven long non-coding RNA repertoire in T-cell acute lymphoblastic leukemia. *Haematologica*. 2014;**99**(12):1808-16.
3. Volders PJ, Helsens K, Wang X, et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res*. 2013;**41**(Database issue):D246-51.
4. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;**30**(1):207-10.
5. Durinck K, Van Loocke W, Van der Meulen J, et al. Characterization of the genome-wide TLX1 binding profile in T-cell acute lymphoblastic leukemia. *Leukemia*. 2015; e-pub ahaed of print 25 June 2015; doi:10.1038/leu.2015.162

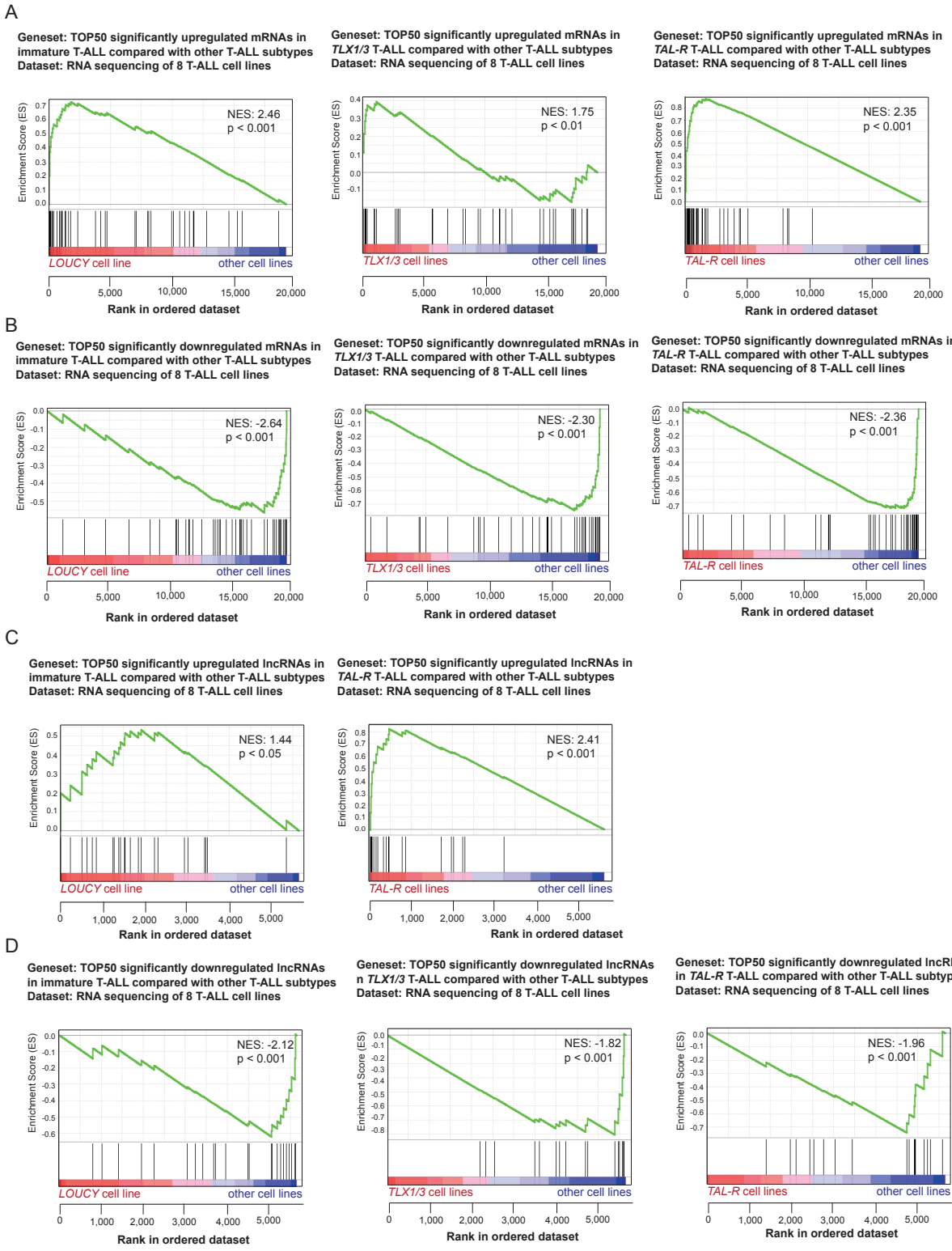
SUPPLEMENTAL FIGURES



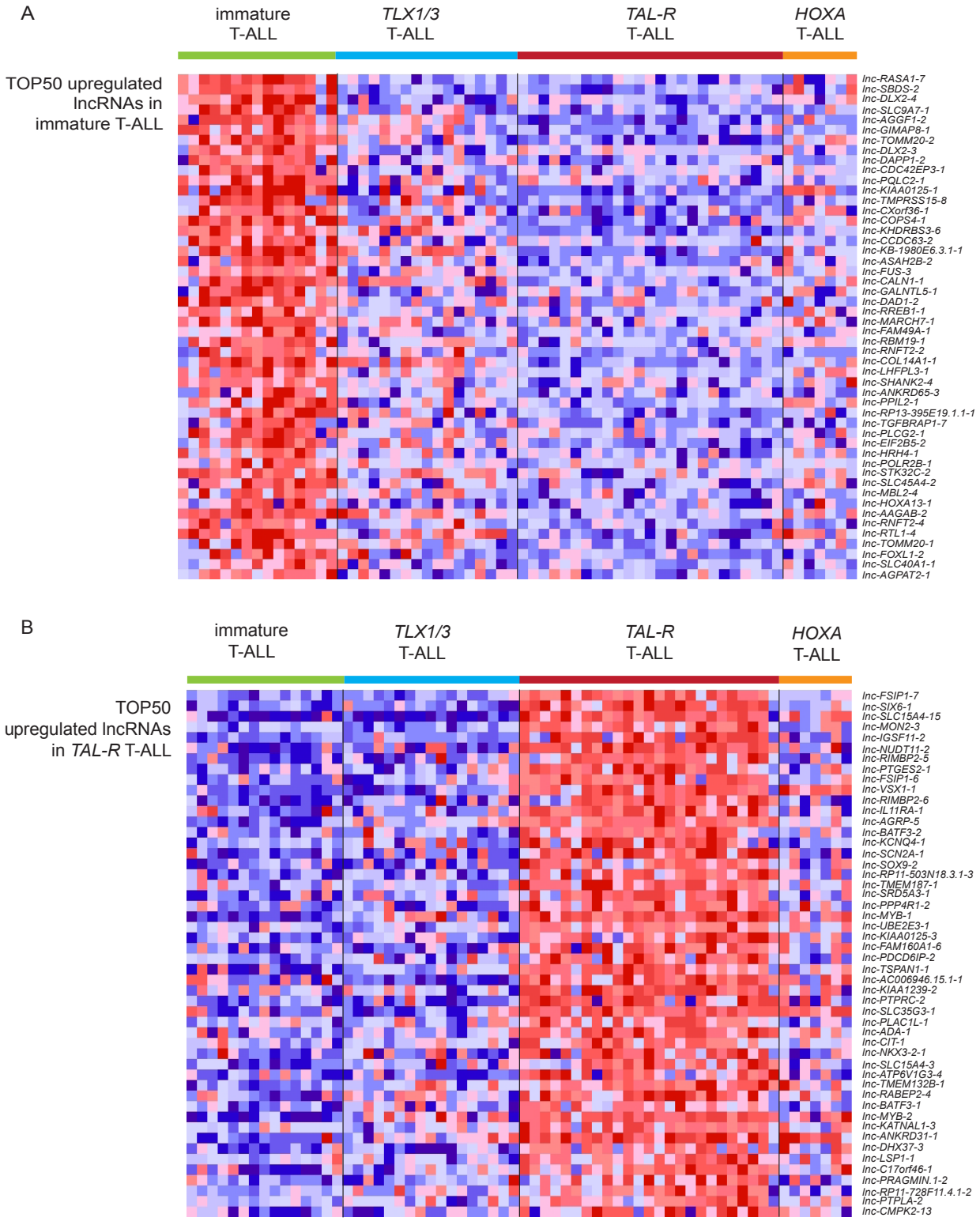
Supplemental Figure 1. T-ALL subgroup lncRNA signatures are mirrored in T-ALL cell lines. (A-C) Heatmap representing the core enrichment of the gene set enrichment analysis performed in figure 1B. **(D-F)** Gene set enrichment analysis of the T-ALL subgroup specific down regulated lncRNAs in the T-ALL cell lines with a heatmap of the core enrichment. **(A+D)** LOUCY as a representative cell line for the immature T-ALL subgroup. **(B+E)** DND-41, HPB-ALL and ALL-SIL as representative cell lines for the TLX1/3 T-ALL subgroup. **(C+F)** PF-382, JURKAT, KE-37 and CCRF-CEM as representative cell lines for the TAL-R subgroup.



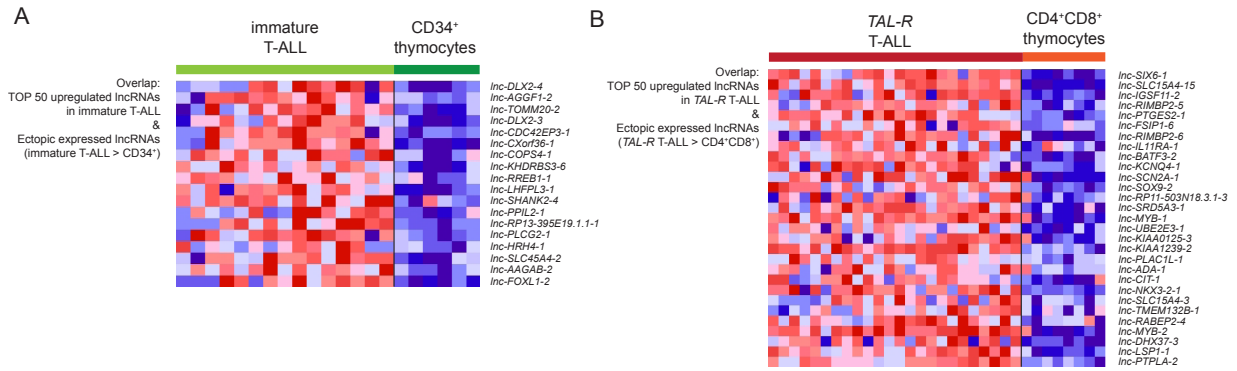
Supplemental Figure 2. T-ALL subgroup mRNA signatures are mirrored in T-ALL cell lines. (A-C) Gene set enrichment analysis of the T-ALL subgroup specific upregulated mRNAs in the T-ALL cell lines with a heatmap of the core enrichment. **(D-F)** Gene set enrichment analysis of the T-ALL subgroup specific downregulated mRNAs in the T-ALL cell lines with a heatmap of the core enrichment. **(A+D)** LOUCY as a representative cell line for the immature T-ALL subgroup. **(B+E)** DND-41, HPB-ALL and ALL-SIL as representative cell lines for the TLX1/3 T-ALL subgroup. **(C+F)** PF-382, JURKAT, KE-37 and CCRF-CEM as representative cell lines for the TAL-R subgroup.



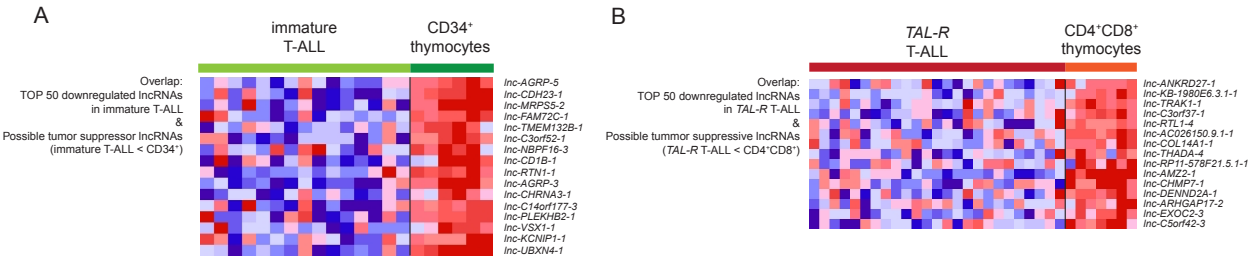
Supplemental Figure 3. T-ALL subgroup signatures defined by microarray are reflected in RNA-sequencing data. **(A)** Gene set enrichment analysis of the T-ALL subgroup specific upregulated mRNAs selected by microarray analysis in RNA-sequencing data of 8 TALL cell lines. **(B)** Gene set enrichment analysis of the T-ALL subgroup specific downregulated mRNAs selected by microarray analysis in RNA-sequencing data of 8 TALL cell lines. **(C)** Gene set enrichment analysis of the T-ALL subgroup specific upregulated lncRNAs selected by microarray analysis in RNA-sequencing data of 8 TALL cell lines. **(D)** Gene set enrichment analysis of the T-ALL subgroup specific downregulated lncRNAs selected by microarray analysis in RNA-sequencing data of 8 TALL cell lines. **(A-D)** LOUCY as a representative cell line for the immature T-ALL subgroup. DND-41, HPB-ALL and ALL-SIL as representative cell lines for the TLX1/3 T-ALL subgroup. PF-382, JURKAT, KE-37 and CCRF-CEM as representative cell lines for the TAL-R subgroup.



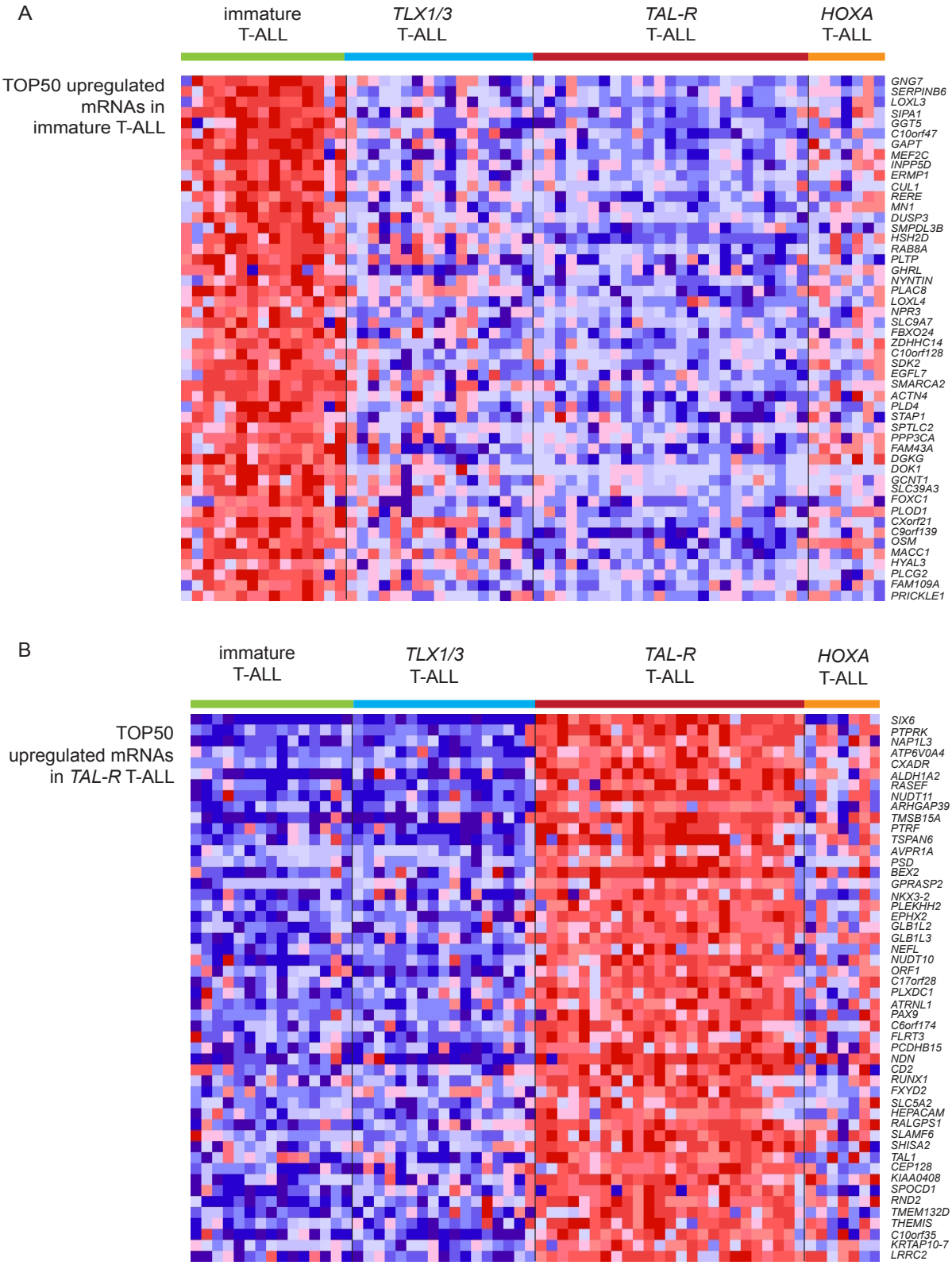
Supplemental Figure 4. Top 50 upregulated lncRNAs in the immature and the TAL-R T-ALL subgroups.(A) Heatmap showing the top 50 upregulated lncRNAs in the immature T-ALL subgroup compared with the other T-ALL subgroups. **(B)** Heatmap showing the top 50 upregulated lncRNAs in the TAL-R T-ALL subgroup compared with the other T-ALL subgroups.



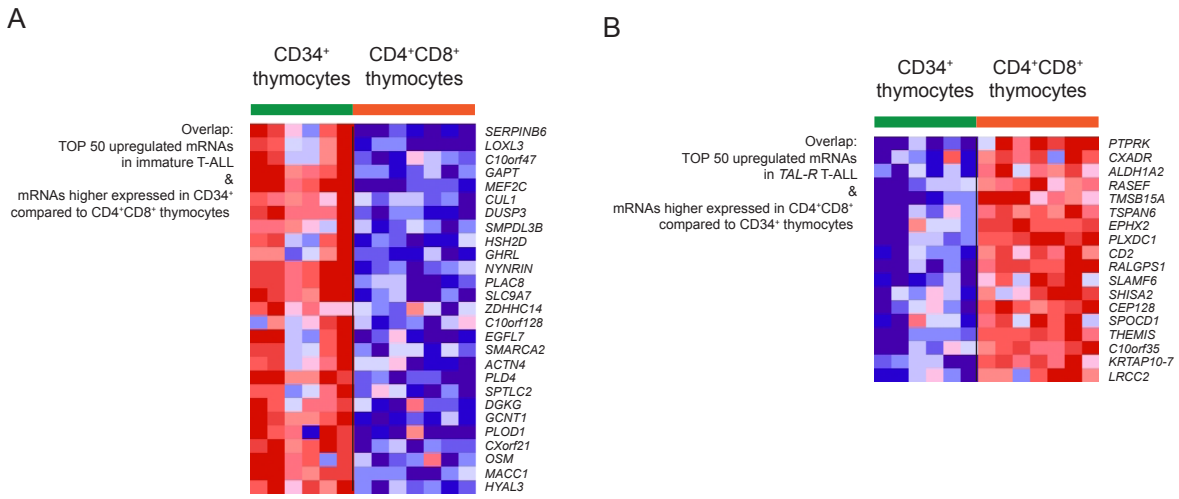
Supplemental Figure 5. Comparative analysis of immature thymocytes and the T-ALL lncRNAome reveals potential oncogenic lncRNAs. (A) Heatmap representing lncRNAs from the top 50 upregulated lncRNAs in the immature T-ALL subgroup that are also significantly higher expressed in the immature T-ALL patients compared with the CD34+ thymocytes (p.adj < 0.05). **(B)** Heatmap representing lncRNAs from the top 50 upregulated lncRNAs in the TAL-R T-ALL subgroup that are also significantly higher expressed in the TAL-R T-ALL patients compared with the CD4+CD8+ thymocytes (p.adj < 0.05).



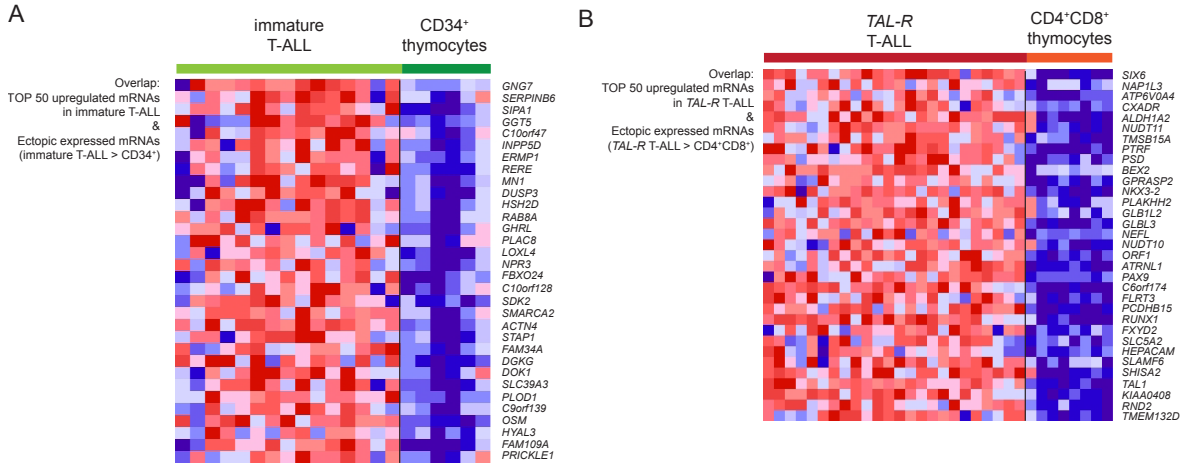
Supplemental Figure 6. Comparative analysis of immature thymocytes and the T-ALL lncRNAome reveals potential tumor suppressive lncRNAs. (A) Heatmap representing lncRNAs from the top 50 downregulated lncRNAs in the immature T-ALL subgroup that are significantly lower expressed in the immature T-ALL patients compared with the CD34+ thymocytes (p.adj < 0.05). **(B)** Heatmap representing lncRNAs from the top 50 downregulated lncRNAs in the TAL-R T-ALL subgroup that are also significantly lower expressed in the TAL-R T-ALL patients compared with the CD4+CD8+ thymocytes (p.adj < 0.05).



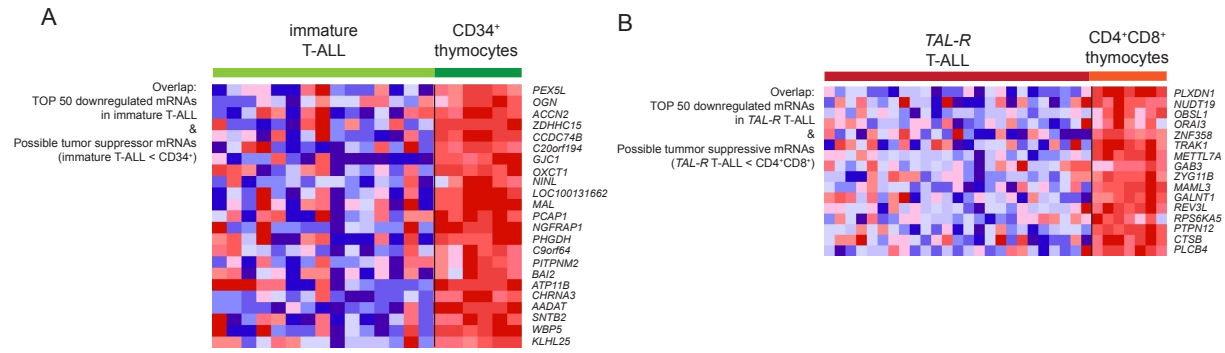
Supplemental Figure 7. Top 50 upregulated mRNAs in the immature and the TAL-R T-ALL subgroups. (A) Heatmap showing the top 50 upregulated mRNAs in the immature T-ALL subgroup compared with the other T-ALL subgroups. **(B)** Heatmap showing the top 50 upregulated mRNAs in the TAL-R T-ALL subgroup compared with the other T-ALL subgroups.



Supplemental Figure 8. Comparison with healthy thymocytes reveals mRNAs with a role in T-cell development. (A) Heatmap representing mRNAs of the top 50 upregulated mRNAs in the immature T-ALL patients that are also significantly higher expressed in CD34+ thymocytes compared with CD4+CD8+ thymocytes (p.adj < 0.05). (B) Heatmap representing mRNAs of the top 50 upregulated mRNAs in the TAL-R group that are significantly higher expressed in CD4+CD8+ thymocytes compared with CD34+ thymocytes (p.adj < 0.05).



Supplemental Figure 9. Comparative analysis of immature thymocytes and the T-ALL mRNAome reveals potential oncogenic mRNAs. (A) Heatmap representing mRNAs from the top 50 upregulated mRNAs in the immature T-ALL subgroup that are also significantly higher expressed in the immature T-ALL patients compared with the CD34+ thymocytes (p.adj < 0.05). **(B)** Heatmap representing mRNAs from the TOP 50 upregulated mRNAs in the TAL-R T-ALL subgroup that are also significantly higher expressed in the TAL-R T-ALL patients compared with the CD4+CD8+ thymocytes (p.adj < 0.05).



Supplemental Figure 10. Comparative analysis of immature thymocytes and the T-ALL mRNAome reveals potential tumor suppressive mRNAs. (A) Heatmap representing mRNAs from the top 50 downregulated mRNAs in the immature T-ALL subgroup that are significantly lower expressed in the immature T-ALL patients compared with the CD34+ thymocytes (p.adj < 0.05). **(B)** Heatmap representing mRNAs from the top 50 downregulated mRNAs in the TAL-R T-ALL subgroup that are also significantly lower expressed in the TAL-R T-ALL patients compared with the CD4+CD8+ thymocytes (p.adj < 0.05).

RNA-sequencing profiling across T-ALL and thymocyte subsets identifies candidate oncogenic lncRNAs in T-ALL

Annelynn Wallaert[1,2], Kaat Durinck[1,2], Wouter Van Loocke[1], Lucie Hernandez[3], Matthias De Decker[4], Tom Taghon[2,4], Jean Soulier[3], Pieter Van Vlierberghe[1,2] and Frank Speleman[1,2]

[1] *Center for Medical Genetics Ghent, Ghent University, Ghent, Belgium*
[2] *Cancer Research Institute Ghent, Ghent, Belgium*
[3] *University Paris Diderot and Hospital Saint-Louis, U944 INSERM, Paris, France*
[4] *Department of Clinical Chemistry, Microbiology and Immunology, Ghent University, Ghent, Belgium*

In preparation

ABSTRACT

Long noncoding RNAs are emerging as important players in cancer. In order to identify candidate oncogenic lncRNAs, we have performed a broad and in depth comparative landscaping of the lncRNAomes of all genetic T-ALL subtypes and normal counterpart T-cell subsets. First, we validated our previous lncRNA T-ALL subgroup classification and showed that poly(A) RNA-sequencing allows more robust classification as compared to array-based analysis. Next, we identified ectopically expressed candidate oncogenic lncRNAs in immature and TAL-R T-ALL, which represent valid targets for functional studies. In addition, using total RNA-sequencing on a smaller subset of samples, we investigated the full spectrum of long noncoding RNAs and identified a set of novel and previously unannotated lncRNAs in T-ALL. Taken together, our study offers a unique resource of data on lncRNAs expressed in T-ALL and normal thymocytes for further data mining and functional exploration.

INTRODUCTION

T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive hematological malignancy arising from a differentiation arrest during T-cell development. T-ALL patients can be classified into different genetic subtypes based upon their aberrant expression of specific transcription factor oncogenes (*TAL*, *TLX1*, *TLX3* or *HOXA*) that demarcates a developmental arrest at a specific stage of T-cell differentiation[1-4]. More than a decade ago, a unique mRNA expression signature characterizing these genetic subgroups was reported[1, 2]. More recently, we showed that long noncoding RNAs also exhibit genetic subtype specific expression patterns[5].

Previous studies describing T-ALL signatures were based on the profiling of the T-ALL samples by micro-array platforms, with a custom designed micro-array platform for the detection of lncRNAs[5, 6]. Currently, RNA-sequencing offers a more powerful platform for sensitive detection of lncRNAs including those who have not been previously annotated.

Here, we analyzed poly(A) RNAseq mRNA and lncRNA profiles of 60 primary T-ALL patient samples from four different genetic T-ALL subtypes (immature, *TLX1/3*, *TAL-R* and *HOXA* overexpressing T-ALL) by poly(A) RNA-seq. We validated our dataset through subgroup classification and show more robust separation by RNA-seq as compared to array-data. Next, we identified ectopically expressed candidate oncogenic lncRNAs through comparison with normal thymocytes. Finally, we detect novel T-ALL lncRNAs, in particular in a subset of 25 primary matching T-ALL patient samples analyzed by total RNA-seq.

MATERIALS AND METHODS

Primary human T-ALL patient samples

Blood samples and bone marrow lymphoblast from T-ALL patients were collected after informed consent according to the Declaration of Helsinki from Saint-Louis Hospital, Paris, France. This study was approved by both the Institut Universitair d'Hématologie Institutional Review Board and the Ethical Committee of Ghent University Hospital. Total RNA was isolated using the miRNeasy mini

kit (Qiagen). These samples are part of a cohort previously investigated by mRNA[8] and lncRNA profiling[5, 9].

Thymocyte subset selection

Thymus tissue was derived from a child undergoing cardiac surgery (UZ Gent) and was obtained and used according to the guidelines of the Medical Ethical Commission of Ghent University Hospital (Ghent, Belgium). Immature $CD34^+$ thymocytes were purified based on MACS purification using CD34 microbeads (Miltenyi Biotec)[10] and were subsequently sorted into a CD1 negative and positive subpopulation. On the other hand, the $CD4^+CD8^+$ double positive subset was isolated using CD4 and CD8 labeled antibodies and was further divided into the CD3 negative and the more mature CD3 positive subpopulations. All sorts were performed using a FACSAriaIII (BDBiosciences).[11] The purity of each subset was at least 98%. Total RNA was isolated using the miRNeasy mini kit (Qiagen).

Poly(A) and total RNA-sequencing

For poly(A) RNA-seq, libraries were prepared using the TruSeq Stranded mRNA sample prep kit (Illumina). 100 ng of total RNA was enriched using the oligodT bead system (Illumina). The isolated mRNA was subsequently fragmented using enzymatic fragmentation. Libraries for total RNA-seq were prepared using the TruSeq Stranded total RNA (with RiboZero Gold) sample prep kit (Illumina). 100 ng of total RNA was depleted of rRNAs using Ribo-Zero Gold magnetic bead based capture-probe system (Illumina). The remaining RNA was subsequently purified (RNAcleanXP) and fragmented using enzymatic fragmentation.

First and second strand synthesis was performed and double stranded cDNA was purified (Agencourt AMPure XP). The cDNA was end-repaired, 3'adenylated and Illumina sequencing adapters were ligated followed by purification. The mRNA stranded libraries were pre-amplified with PCR and purified (Agencourt AMPure XP). The quality of the libraries was inspected on the 2100 Bioanalyzer (high sensitivity DNA chip, Agilent). High quality libraries were quantified using the Qubit Fluorometer (Life Technologies). After concentration normalization, single-end sequencing was performed on the NextSeq500 instrument (Illumina) according to manufacturer instructions.

RNA-seq data-processing

Reads were mapped to the reference genome GRCh38 with STAR v2.4.2a. STAR was also used for gene expression quantification on Gencode v23 GTF. For each sample, transcript assembly was performed with StringTie. Subsequently Cuffmerge was used to merge all transcript assemblies into one superset containing all transcripts present in at least one sample. Expression levels are visualized as Transcripts Per Kilobase Million (TPM), which is calculated by first dividing all read counts by their gene length in kilobases (RPK, reads per kilobase) and subsequently by the scaling factor per sample. This scaling factor was determined by summing up all RPK values per sample and dividing this by 1 million.

To identify unannotated transcripts, the merged GTF file was compared with different annotation GTF files (Gencode23,Refseq78, Ensembl81 and lncipedia3.1) using Cuffcompare.

Differential expression analysis

Gencode mRNA and lncRNA expression data was filtered with a background correction that only retained the genes detected by at least 4 reads in at least 60% of samples from one T-ALL subgroup or in at least all samples from one thymocyte subset. Differential expression analysis was performed using the DESeq2 algorithm in R[12].

RESULTS

Validation of the poly(A) RNA-sequencing dataset of T-ALL patient samples through T-ALL genetic subtype classification

Poly(A) RNA-sequencing was performed for 60 primary T-ALL patient samples with an average sequencing depth of 51 million reads per sample, of which 65% could be mapped to the Gencode v23 database[13,14]. The 60 samples are divided over the four main T-ALL subtypes: 13 immature, 17 *TLX1/*3 overexpressing, 23 *TAL*-rearranged and 7 *HOXA* overexpressing cases (Figure 1A). To correct for background detection, only genes that were detected by at least four reads in 60% of the samples of one particular genetic subtype were retained (14,447 protein coding genes and 3,264 lncRNAs ('lincRNA or 'antisense' biotype)). The expression level of the lncRNAs is, as expected, much lower than for protein coding genes, as can be appreciated from the density plot (Figure 1B).

Principal Component Analysis (PCA) shows a clustering of the samples according to subtype for both the mRNA (Supplementary Figure 1A) and lncRNA (Figure 1C) expression, with the only exception for the *HOXA* samples which are broadly dispersed. Hierarchical clustering of the samples, using the top 1000 mRNAs or top 1000 lncRNAs with the largest standard deviation, also shows a separation of most of the immature, *TAL-R* and *TLX1/3* patient samples to the expected subtype based on clinical data (Supplementary Figure 2). Notably, lncRNA expression profiles yield a similar qualitative separation of T-ALL subtypes as hierarchical clustering based on protein coding gene expression profiles.

To identify the subtype specific lncRNAs in our poly(A) RNA-seq dataset, we performed DESeq2 differential expression analysis[12] using iterations of one-to-three comparisons. We validated this method based on the protein coding gene expression and detected several specific mRNAs that were already identified by previous groups (Supplementary Figure 1B, Supplementary Table 1)[1,2]. The heatmap in Figure 1D represents the top 50 most differentially expressed lncRNAs per subgroup. Again, this pattern is less distinct for the *HOXA* group as only 48 *HOXA*-specific lncRNAs could be identified. For the other subtypes, at least 550 lncRNAs were significantly up or downregulated (p.adj < 0.05) (Supplementary Table 2).

T-ALL subtype specific lncRNAs in T-ALL cell lines

T-ALL cell lines are extensively used for functional tests in genetic studies. As a prelude to further functional analysis of candidate oncogenic lncRNAs, we therefore analyzed lncRNA profiles of commonly used T-ALL cell lines in a publicly available poly(A) RNA-seq data[15]. The LOUCY cell line is representative for the immature T-ALL patients, whereas HPB-ALL (*TLX3+*), DND-41 (*TLX3*+) and ALL-SIL (*TLX1+)* represent the *TLX1/3* T-ALL subtype. The *TAL-R* subtype can be linked to the cell lines JURKAT, PF-382, KE-37 and CCRF-CEM. In Figure 2 and Supplementary Figure 3, the density plots represent the lncRNA expres-

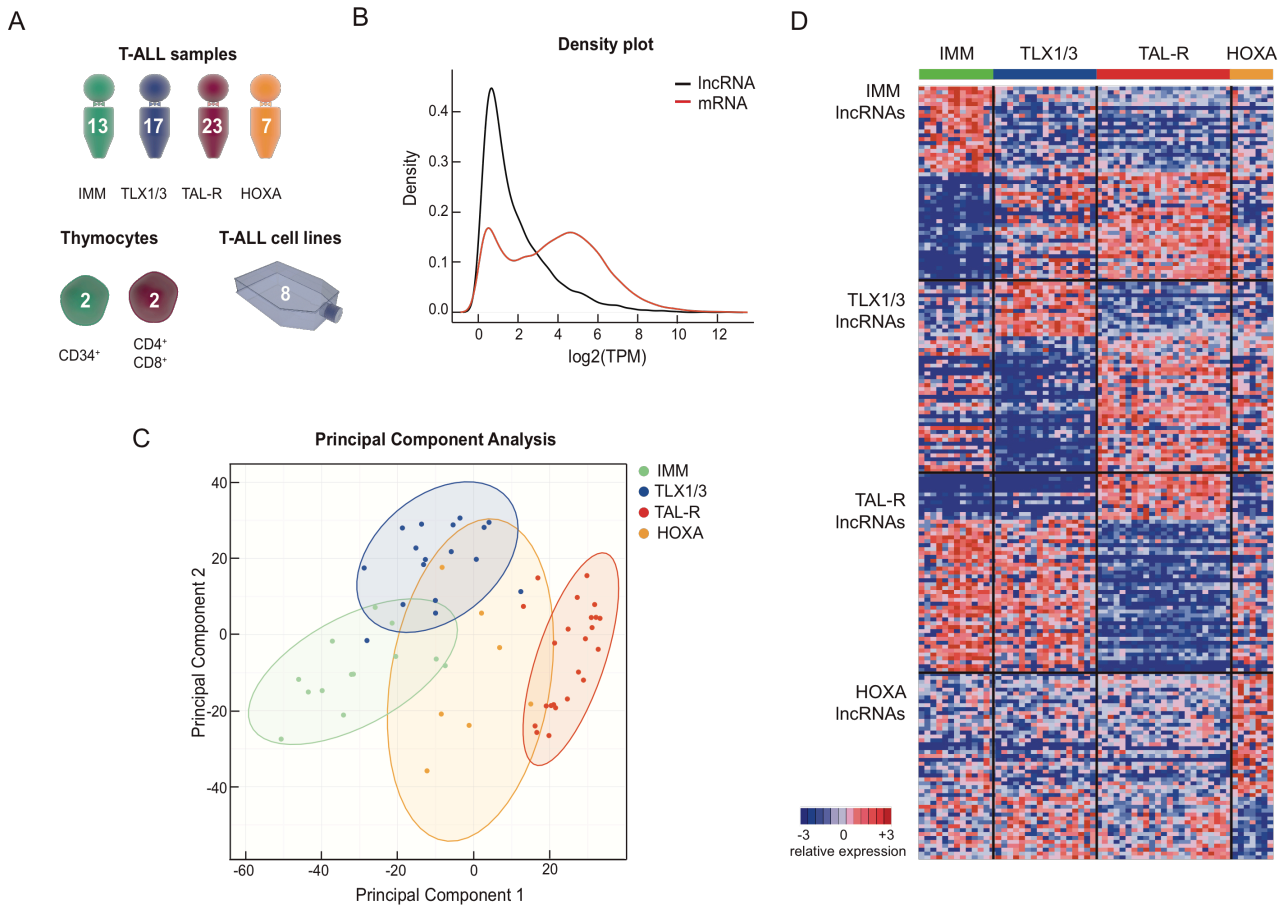


Figure 1. Poly(A) RNA-seq reveals a T-ALL subtype specific lncRNA expression profile. (A) Overview of the samples used in this study. **(B)** Density plot representing the expression level of mRNAs and lncRNAs in the T-ALL patient samples. **(C)** Principal Component Analysis for the lncRNAs detected by poly(A) RNA-seq. **(D)** Heatmap representing the top 50 most differentially expressed lncRNAs for each subtype in comparison to the others (p.adj < 0.05).

sion levels in these cell lines, with the expression of the top 10 highest expressed and upregulated lncRNAs for the immature, *TLX1/3* or *TAL-R* subgroups represented by the vertical black bars. We therefore conclude that nearly all of selected overexpressed lncRNAs are also detected at relatively high levels in the representative cell lines validating their use for further study of the oncogenic effect.

To further explore the T-ALL subtype specific lncRNAs in T-ALL cell lines, we performed Gene Set Enrichment Analysis (GSEA) with the top 50 up- or downregulated lncRNAs per subtype as genesets (Supplementary Figure 4). All genesets containing the 50 downregulated lncRNAs per subtype were significantly enriched in the cell lines not resembling that subtype. For the upregulated lncRNAs per subtype, there was only a significant enrichment for the *TAL-R* lncRNAs in the *TAL-R* cell lines. For the immature and *TLX1/3* upregulated lncRNAs, several were also higher expressed in their representative cell lines as compared to the others, but the full geneset could not be significantly linked to the cell lines.

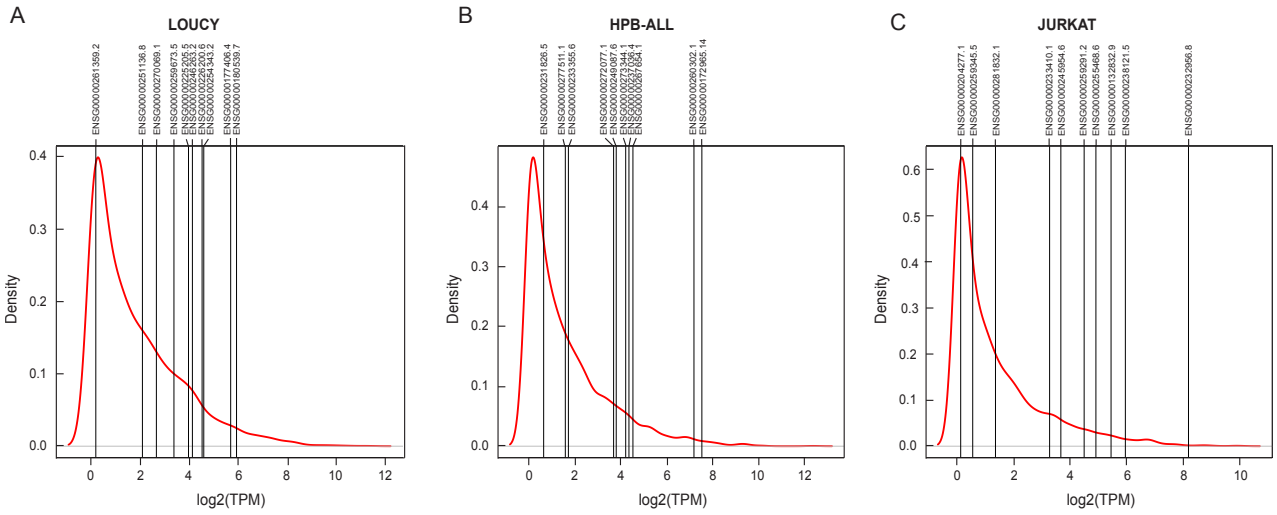


Figure 2. Poly(A) RNA-seq of T-ALL cell lines. Density plots of lncRNA expression in the LOUCY cell line, a cell line representative for IMM T-ALL **(A)**, the HPB-ALL cell line, for the *TLX1/3* subgroup **(B)** and the JURKAT cell line for the *TAL-R* subgroup **(C)**. The vertical bars represent the top 10 lncRNAs that are significantly higher expressed in the representative subgroup compared to the others.

Poly(A) RNA-sequencing detects subtype specific lncRNAs more robustly than micro-array profiling

We previously published the subtype specific mRNA and lncRNA signatures of these patients based on micro-array expression profiling[5]. If we compare both signatures for protein coding genes, we find 50-80% of significant subtype specific mRNAs detected by poly(A) RNA-seq also in our microarray dataset (Supplementary Figure 5A). This overlap is however smaller (less than 30%) for the *HOXA*-specific mRNAs, but again, this might be due to the low number of patient samples. However, with poly(A) RNA-seq, much more *HOXA*-specific mRNAs were detected. Remarkably, if we compared the mRNAs that were subtype specific by both methods, the detected fold change was significantly (Paired students t-test; $p < 0,01$) larger for the micro-array profiling of the immature, *TLX1/3* and *HOXA* signatures (Supplementary Figure 5B).

To compare the lncRNA signatures obtained from both methods, we had to cope with the different annotation of the micro-array probes compared to our poly(A) RNA-seq dataset. As a consequence, we could only convincingly compare 860 Gencode defined lncRNAs ('lincRNA' and 'antisense' biotype) with probes on the micro-array and took only these lncRNAs into account for further analysis. For the T-ALL subtype specific lncRNAs, we detected an overlap of about 50% between both methods (Figure 3A). In comparison to the protein coding genes, the fold change of the subtype specific lncRNAs in the overlap between the 2 methods was significantly higher for the RNA-sequencing dataset as compared to the micro-array data (Paired student's t-test; $p < 0.01$; Figure 3B). In conclusion, our poly(A) RNA-seq dataset gives the advantage over the micro-array data that it detects the same subtype specific lncRNAs, but with a larger difference in expression.

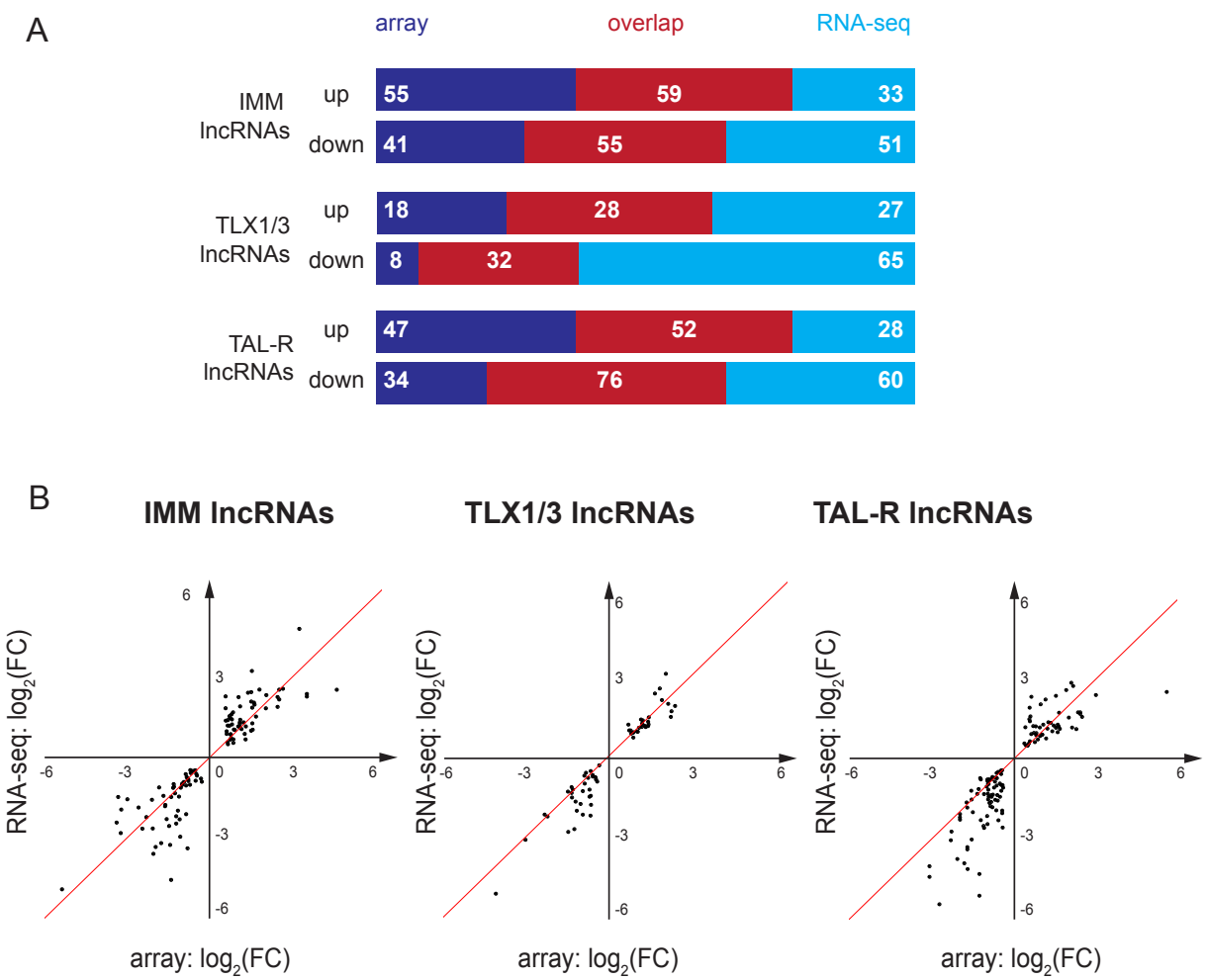


Figure 3. Poly(A) RNA-seq detects subtype specific lncRNAs more robustly than micro-arrays. (A) Overview of the amount of subtype specific lncRNAs detected only by micro-array profiling (dark blue), only by RNA-sequencing (light blue) or by both methods (red). **(B)** Comparison between micro-array profiling and RNA-sequencing of the fold change detected for the significant lncRNAs identified by both methods. The red line represents the distribution if the fold change would be equal (x=y).

Poly(A) RNA-seq detects previously unannotated T-ALL lncRNAs

We next took advantage of the power of RNA-sequencing for the detection of novel lncRNAs. To this end, we aligned our RNA-sequencing reads with the STAR algorithm[16] to the hg38 genome, followed by transcript assembly with StringTie[17]. These transcript assemblies per sample were then combined with Cuff-Compare[18] leading to the in-house establishment of a T-ALL transcriptome (annotated with XLOC codes). This led to the identification of 12,524 genes that were not present in Gencode (v23)[13], Ensembl (v81)[19], RefSeq (v78)[20] and lncipedia (v3.1)[21]. To eliminate very low expressed genes, we performed a background correction in which we only retained the 3303 genes that were detected by at least 4 reads in at least 5 samples. Furthermore, there is more evidence that a gene is truly functional if at least one splicing event occurs, so we filtered out genes with only one exon. Of those 932 genes, 526 genes have no protein coding potential as calculated by the 'Coding Potential

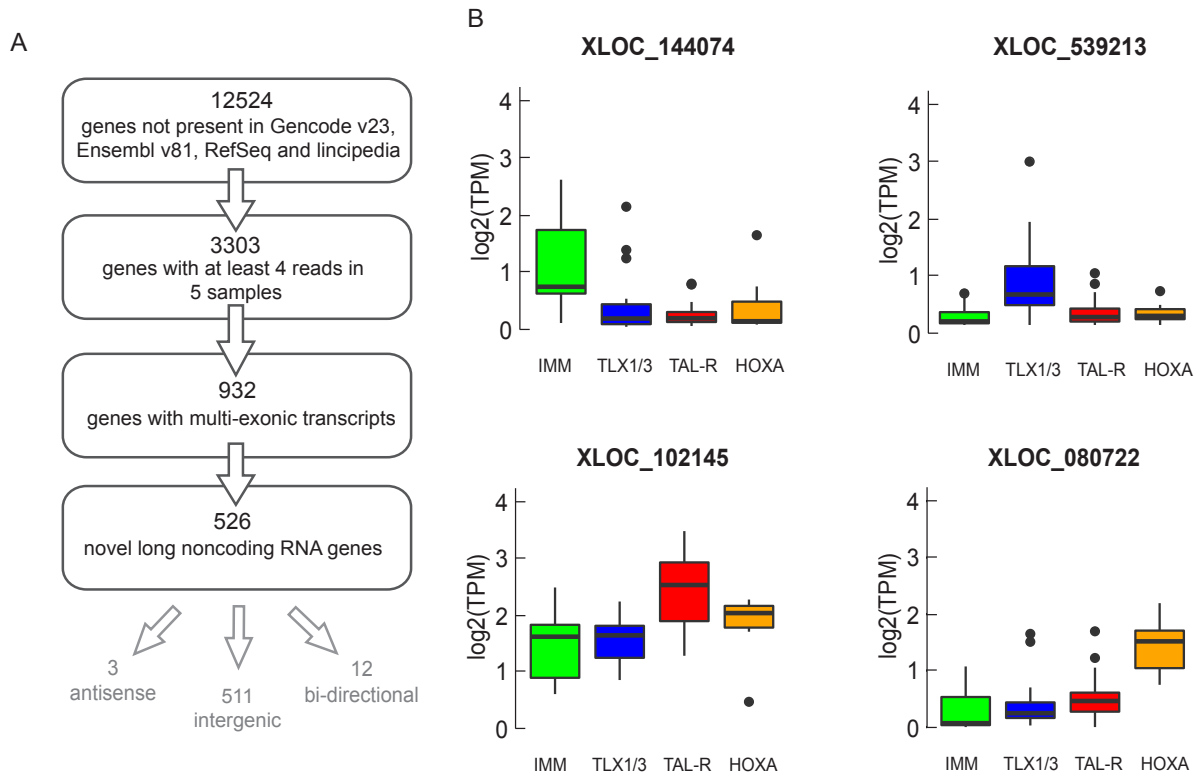


Figure 4. Poly(A) RNA-seq can link previously unannotated lncRNAs to specific T-ALL subtypes. (A) Overview of the selection method to identify novel long noncoding RNA genes. **(B)** Boxplots representing one novel lncRNA per T-ALL subgroup that is significantly upregulated in that subgroup compared to the others.

Calculator' (CPC) algorithm (Figure 4A)[22] and these genes are thus considered as novel lncRNAs (Supplementary table 3). Of these 526 lncRNAs, 511 were intergenic, 3 were antisense and 12 were bi-directional (having their transcription start site closer than 100 bp from the transcription start site of a protein coding gene).

We repeated the differential expression analysis for lncRNAs between the T-ALL subtypes, but this time, we added the novel identified lncRNAs to these already included by Gencode (v23)[13, 14]. Several of these putatively novel lncRNAs were also significantly up- or downregulated in a specific subtype and interestingly, some of these were represented in the top 50 most significantly upregulated lncRNAs. For each subtype, the most significant upregulated novel lncRNA, which was also present in the top 50 of all lncRNAs, is shown in Figure 4B. It should be noted that these novel lncRNAs have very low expression levels, but their subtype specific expression is quite clear. As an interesting example, *XLOC_539213* is upregulated in the *TLX1/3* subtype and detected at quite high rates in the HPB-ALL and DND41 cell lines (*TLX3+*, data not shown). Also the other novel subtype specific lncRNAs are represented in the cell lines, however at a low level.

Comparative analysis of normal thymocyte versus T-ALL lncRNA transcriptomes identifies oncogenic lncRNAs

T-ALL occurs from a malignant transformation of maturing T-cells in the thymus. The lymphoblasts of the immature T-ALL cases show a differentiation arrest very early dur-

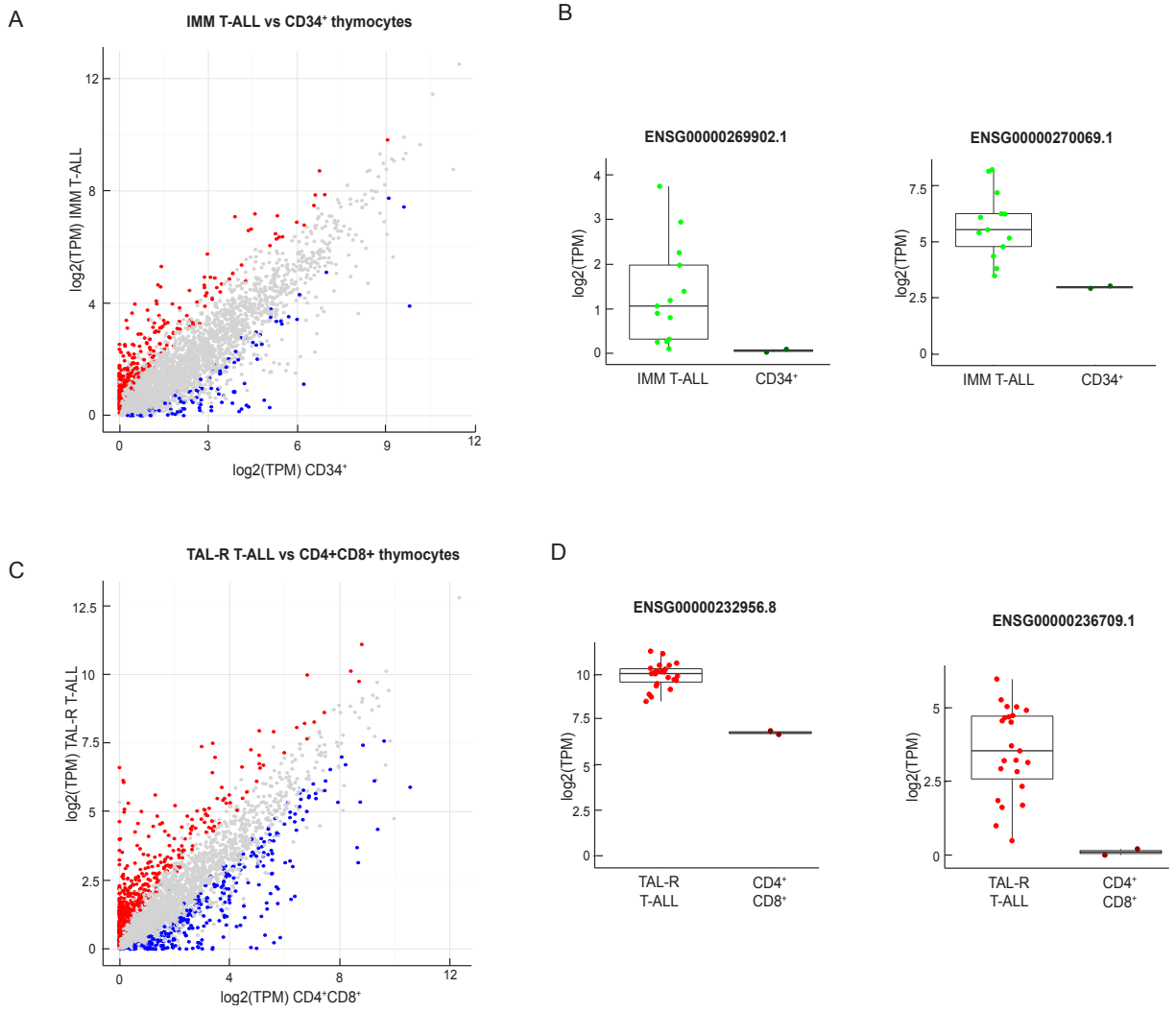


Figure 5. Comparison with thymocyte subsets detects oncogenic subtype specific lncRNAs. (A) Diagonal plot comparing the expression of lncRNAs between IMM T-ALL samples and CD34$^+$ T-cell subsets. Red dots are lncRNAs significantly upregulated in IMM T-ALL and blue dots are the significantly downregulated lncRNAs (p.adj < 0.05). **(B)** Two examples of IMM T-ALL upregulated lncRNAs that are also upregulated in IMM T-ALL as compared to their healthy donor counterparts. **(C)** Diagonal plot comparing the expression of lncRNAs between TAL-R T-ALL samples and CD4$^+$CD8$^+$ T-cell subsets. Red dots are lncRNAs significantly upregulated in TAL-R T-ALL and blue dots are the significantly downregulated lncRNAs (p.adj < 0.05). **(B)** Two examples of TAL-R T-ALL upregulated lncRNAs that are also upregulated in TAL-R T-ALL as compared to their healthy donor counterparts.

ing T-cell development, at the CD34$^+$ stage. The *TAL-R* cases however, have had a differentiation arrest at a later stage during development, from the CD4$^+$CD8$^+$ double positive stage on. To identify oncogenic lncRNAs, we profiled the lncRNA expression of 2 CD34$^+$ and 2 CD4$^+$CD8$^+$ subsets from a healthy donor by poly(A) RNA-sequencing.

By means of differential RNA-seq analysis (DESeq2), we identified 276 lncRNAs that were upregulated in immature T-ALL compared to the CD34$^+$ healthy donor subsets (Figure 5A, Supplementary Table 4). Sixteen of these lncRNAs were also identified in the top 50 upregulated lncRNAs in immature T-ALL compared to the other subtypes (two examples in Figure 5B). These 16 lncRNAs are thus potential oncogenic drivers for immature T-ALL. Furthermore, 475 lncRNAs were upregulated in *TAL-R* T-ALL as compared to

the $CD4^+CD8^+$ subsets (Figure 5C, Supplementary Table 4), with 15 of these also in the top 50 upregulated lncRNAs in *TAL-R* T-ALL (Figure 5D).

Comparison of poly(A) and total RNA sequencing data sets for unannotated lncRNAs

As a last part of this study, we evaluated whether additional information can be obtained on lncRNA expression profiles of T-ALL samples by means of total RNA-seq (with ribodepletion) as compared to poly(A) RNA-seq. A first incentive, is that several lncRNAs do not have a poly(A) tail, so these will, by definition, not be detected by poly(A) RNA-seq. It should however be stated that we detect some known non-poly(A) genes (ex. histone genes and MALAT1 lncRNA) in our poly(A) dataset as some random priming might occur during library preparation, however these genes are only detected at a low ratio. One disadvantage of total RNA-seq is the cost, as the sequencing depth should be increased, to correct for the reads mapping to introns of mRNAs.

We performed a pilot study with 25 samples (5 immature, 10 *TLX1/3*, 5 *TAL-R* and 5 *HOXA* overexpressing T-ALL cases) also profiled by poly(A) RNA-seq. The average read depth of the total RNA-seq was 103 million reads, with 45 million reads mapping to Gencode (v23). To perform an optimal comparison between both technologies, we also selected only the matching 25 samples out of our poly(A) dataset. The same mapping and analysis were performed as described above and this led to the identification of 1108 novel multi-exonic genes with a log2(TPM) expression level of at least 0.1 in 5 samples for the poly(A) dataset and 1288 for the total RNA-seq dataset. Protein coding potential calculation (CPC[22]) further identified 712 novel lncRNAs in the

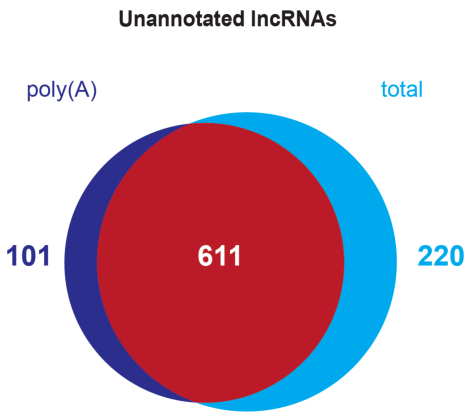


Figure 6. Total RNA-sequencing detects more previously unannotated lncRNAs compared to poly(A) RNA-seq. Venn diagram representing the amount of previously unannotated lncRNAs detected by both methods.

poly(A) data and 831 novel lncRNAs in the total RNA-seq data, with an overlap of 611 lncRNAs (Figure 6, Supplementary Table 5). In this respect, it might be interesting to perform total RNA-seq as it detects more novel lncRNAs than poly(A) RNA-seq. If we also look at the annotated mRNAs and lncRNAs in the Gencode database, we also detect more genes in total RNA-seq after a background correction (more than four reads in at least two samples of one T-ALL subtype). With poly(A) RNA-seq we detected 15,858 protein coding genes and 4,720 lncRNAs and with total RNA-seq there were 16,388 protein coding genes and 5,759 lncRNAs. A pairwise comparison between the samples also revealed that total RNA-seq detected significantly more mRNAs and lncRNAs with 4 reads or more (Students t-test, p-value < 0.001 for both mRNAs and lncRNAs). So total RNA-seq does not only detect more novel genes, but also more known genes.

DISCUSSION

T-ALL genetic subtypes have already been linked in the past to specific mRNA profiles and more recently our research team established subgroup specific lncRNA expression profiles by micro-array technology[1, 2, 5]. In this study, we profiled gene expression of 60 T-ALL patient samples divided over the several subtypes by poly(A) RNA-sequencing. First, we could confirm that patients from a specific subtype cluster together by their gene expression profiles, either for mRNAs or for lncRNAs and defined an updated lncRNA signature for the subtypes (Figure 1). By using a publically available unstranded poly(A) RNA-seq dataset of T-ALL cell lines, we could also verify the expression of the subtype specific lncRNAs in matching T-ALL cell lines. Most of the denoted signatures were linked to the subtype specific cell lines, however there were some exceptions (Figure 2). These exceptions might be due to the different nature of the RNA-sequencing datasets, as our T-ALL patient cohort was a stranded poly(A) dataset and the publicly available T-ALL cell line data was unstranded. Furthermore, using poly(A) RNA-seq data of healthy donor thymocytes we could identify ectopically expressed subtype specific lncRNAs for immature T-ALL and TAL-R T-ALL (Figure 5)

We could compare this dataset with the previously published micro-array data[5], as the same patient cohort was used. For protein coding genes, there was a big overlap between the 2 datasets and for this analysis, poly(A) RNA-sequencing did not immediately give an advantage. However, it might be interesting to check this poly(A) dataset for specific transcripts or splicing variants of mRNAs present in the subtypes, but this falls beyond the scope of this paper. In terms of lncRNAs, the overlap between both datasets was quite difficult and we could only compare a small amount of lncRNAs. Still, there was quite some overlap between both subtype specific signatures. Remarkably, lncRNAs that were denoted as significant by both methods had a larger fold change in the poly(A) RNA-seq dataset. This could imply that poly(A) RNA-seq has the possibility to better determine the differential expression for lncRNAs than micro-arrays have (Figure 3).

Another advantage of poly(A) RNA-sequencing over micro-array data is the possibility to detect novel, unanotated lncRNAs. This can be very interesting, as lncRNAs are known to be tissue or even cell-type specifically expressed at rather low levels. This could imply that lncRNAs with a specific role in T-cell development or T-ALL (subtype specific) oncogenesis have not been detected by any other RNA-seq study. In our poly(A) dataset of 60 T-ALL samples, we detected 526 potential novel interesting lncRNAs, that were not present in Gencode (v23), Ensemble (v81) RefSeq (v78) and lncipedia (v3.1)[13, 14, 19-21]. Other criteria we took into account is the expression level (at least 4 reads in 5 samples), the need for a splicing event and no protein coding potential (calculated with CPC[22]). It should be noted that unspliced, single-exon transcripts might also be interesting, as they can for example be linked to the class of enhancer RNAs[23]. Several novel lncRNAs are also subtype specifically expressed and are even present in the top 50 most significant lncRNAs (annotated and novel) for a subtype (Figure 4). However, it should be interesting to also add CAGE (Cap Analysis Gene Expression) data and histone marks (ex. H3K4me3 for active promoters and H3K27ac for enhancers) to further identi-

fy these novel lncRNAs and to claim that these are independent transcripts.

We also set up a pilot study with 25 patient samples to test if total RNA-sequencing would give an extra benefit over poly(A) RNA-seq to detect lncRNAs in T-ALL. With total RNA-sequencing, more known protein coding genes and lncRNAs were detected and we could also show that more novel previously unannotated lncRNAs were detected. Furthermore, total RNA-seq could be interesting to possibly detect circular RNAs[24] or other non poly-adenylated RNA transcripts (for ex. snoRNAs).

With this study, we updated the lncRNA signatures of T-ALL subtypes by poly(A) RNA-sequencing and could show that poly(A) RNA-seq has an advantage over the previously used micro-array profiling. Next to that, we also identified several novel, previously unannotated lncRNAs that could be of interest for further analysis as they might be lncRNAs that are only expressed in a T-ALL context.

AUTHOR CONTRIBUTION

AW and WV performed data analysis, AW wrote the manuscript, KD, LH and MD performed experiments, TT, JS, PV and FS supervised the project and writing of the manuscript.

ACKNOWLEGDEMENTS

The authors would like to thank the following funding agencies: the Fund for Scientific Research Flanders ('FWO Vlaanderen' research projects 3G056413N and 3G055013N to FP; PhD grant to AW), Ghent University Research Fund (GOA LNCCA), the Belgian Program of Interuniversity Poles of Attraction (IUAP P7/03) and the Childhood Cancer Fund Ghent (Kinderkankerfonds VZW).

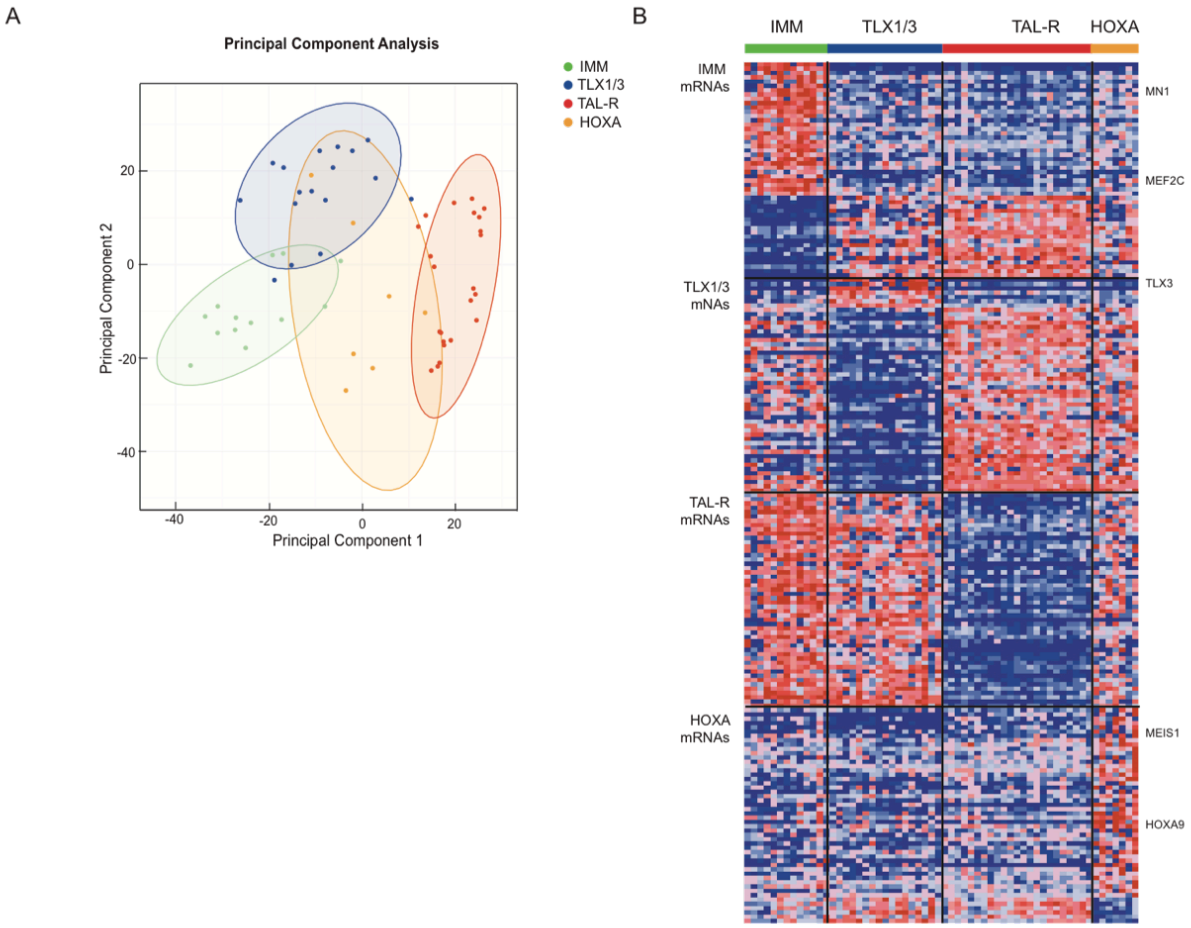
Supplementary tables can be requested via e-mail.

REFERENCES

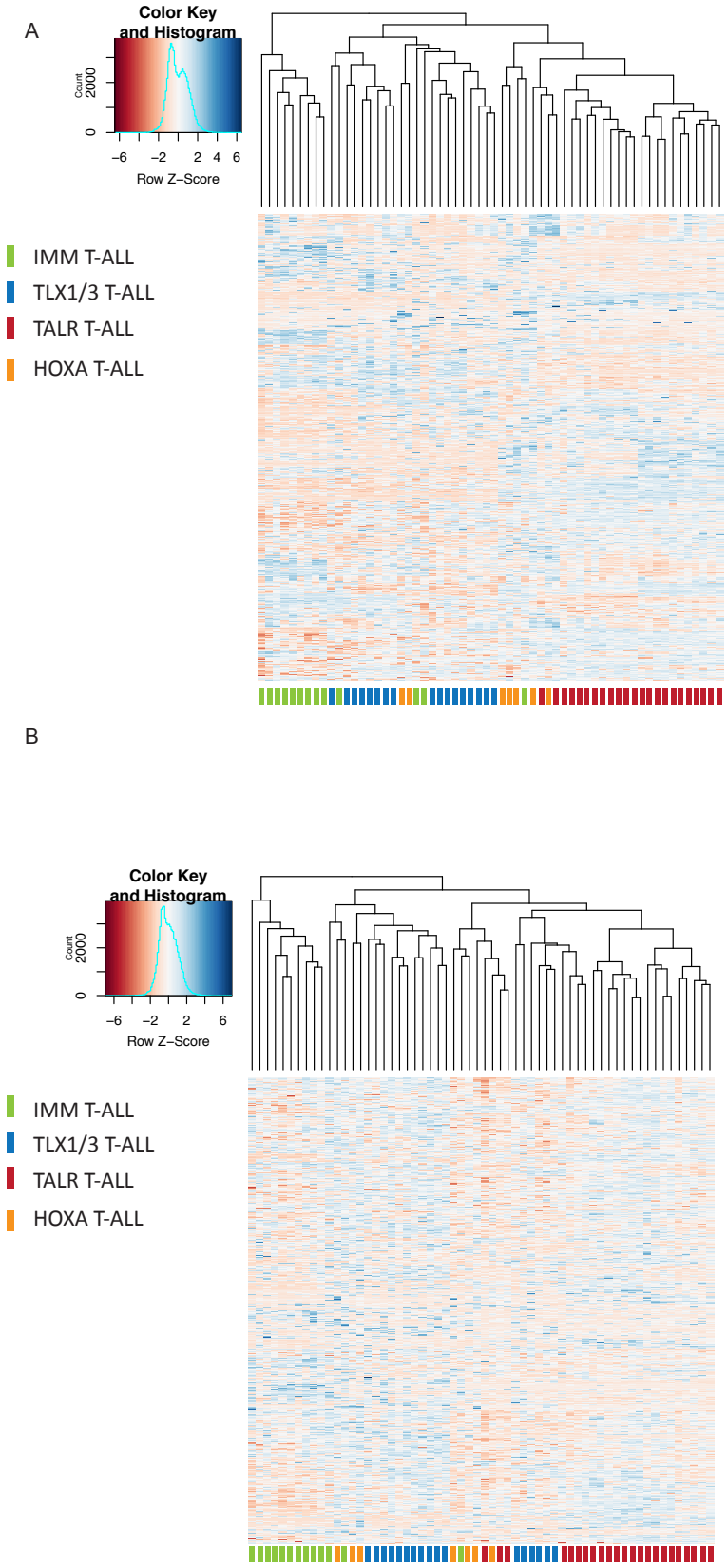
1. Soulier J, Clappier E, Cayuela JM, Regnault A, Garcia-Peydro M, Dombret H, et al. HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). Blood. 2005;106(1):274-86.
2. Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC, et al. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer Cell. 2002;1(1):75-87.
3. Van Vlierberghe P, Pieters R, Beverloo HB, Meijerink JP. Molecular-genetic insights in paediatric T-cell acute lymphoblastic leukaemia. Br J Haematol. 2008;143(2):153-68.
4. Meijerink JP. Genetic rearrangements in relation to immunophenotype and outcome in T-cell acute lymphoblastic leukaemia. Best Pract Res Clin Haematol. 2010;23(3):307-18.
5. Wallaert A, Durinck K, Van Loocke W, Van de Walle I, Matthijssens F, Volders PJ, et al. Long noncoding RNA signatures define oncogenic subtypes in T-cell acute lymphoblastic leukemia. Leukemia. 2016;30(9):1927-30.
6. Volders PJ, Helsens K, Wang X, Menten B, Martens L, Gevaert K, et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res. 2013;41(Database issue):D246-51.
7. Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL. Genomewide characterization of non-polyadenylated RNAs. Genome Biol. 2011;12(2):R16.
8. Clappier E, Gerby B, Sigaux F, Delord M, Touzri F, Hernandez L, et al. Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. J Exp Med. 2011;208(4):653-61.
9. Durinck K, Wallaert A, Van de Walle I, Van Loocke W, Volders PJ, Vanhauwaert S, et al. The Notch driven long non-coding RNA repertoire in T-cell acute lymphoblastic leukemia. Haematologica. 2014;99(12):1808-16.
10. Van de Walle I, Waegemans E, De Medts J, De Smet G, De Smedt M, Snauwaert S, et al. Specific Notch receptor-ligand interactions control human TCR-alphabeta/gammadelta development by inducing differential Notch signal strength. J Exp Med. 2013;210(4):683-97.
11. Taghon T, Waegemans E, Van de Walle I. Notch signaling during human T cell development. Curr Top Microbiol Immunol. 2012;360:75-97.
12. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for

RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
13. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012;22(9):1760-74.
14. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22(9):1775-89.
15. Atak ZK, Gianfelici V, Hulselmans G, De Keersmaecker K, Devasia AG, Geerdens E, et al. Comprehensive analysis of transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. PLoS Genet. 2013;9(12):e1003997.
16. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.
17. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature biotechnology. 2015;33(3):290-5.
18. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562-78.
19. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. Nucleic Acids Res. 2016;44(D1):D710-6.
20. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44(D1):D733-45.
21. Volders PJ, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, et al. An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res. 2015;43(8):4363-4.
22. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007;35(Web Server issue):W345-9.
23. Lai F, Shiekhattar R. Enhancer RNAs: the new molecules of transcription. Curr Opin Genet Dev. 2014;25:38-42.
24. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. Nature biotechnology. 2014;32(5):453-61.

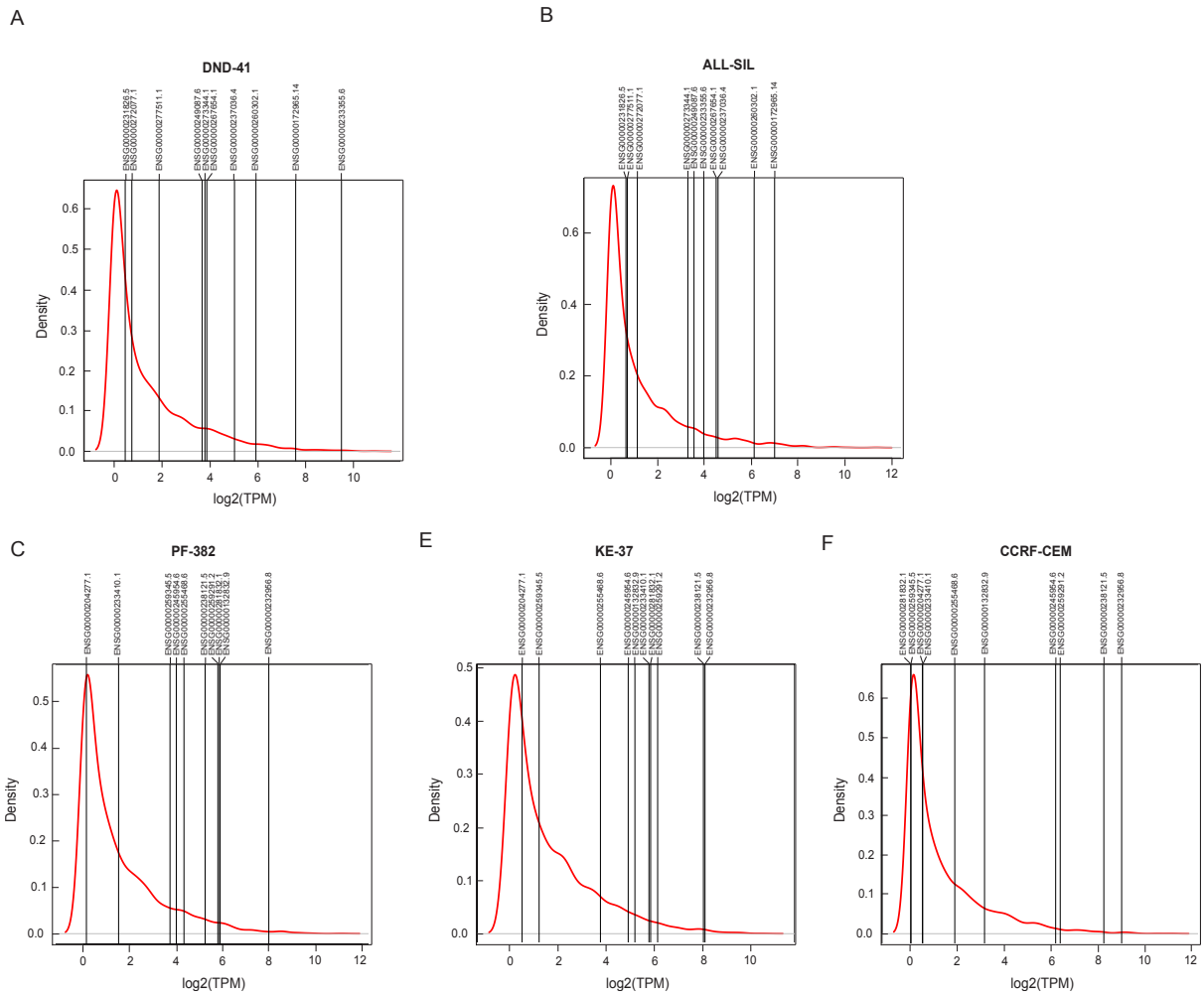
SUPPLEMENTARY FIGURES



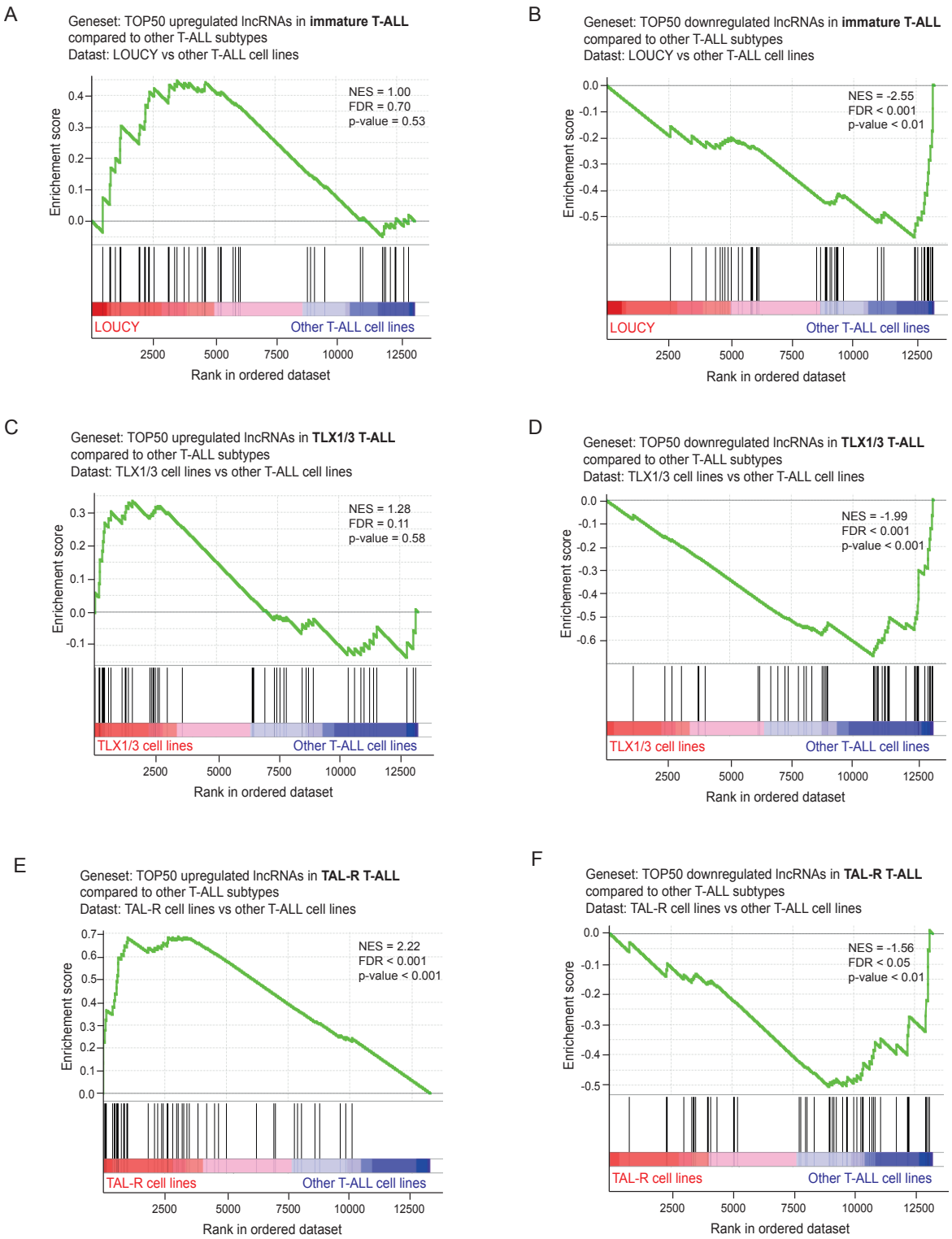
Supplementary Figure 1. Subtype specific expression pattern of mRNAs detected by poly(A) RNA-seq of 60 T-ALL patients samples. (A) Principal Component Analysis for the mRNAs detected by poly(A) RNA-seq. **(B)** Heatmap representing the top 50 most differentially expressed mRNAs for each subtype in comparison to the others (p.adj < 0.05).



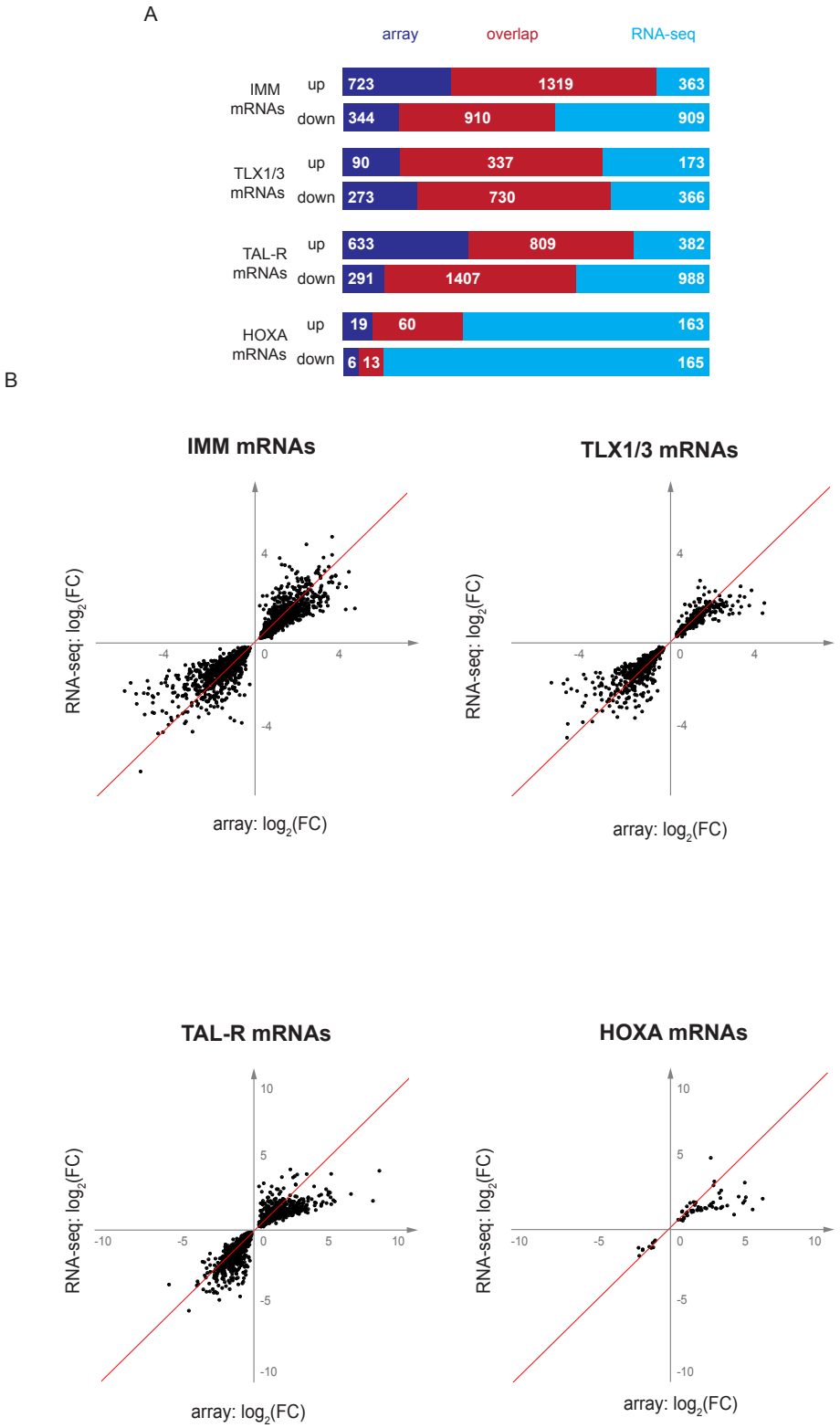
Supplementary Figure 2. Unsupervised clustering of T-ALL patient samples divides the samples according to subtype. (A) Hierarchical clustering of the 1000 mRNAs with the highest standard deviation over all samples. **(B)** Hierarchical clustering of the 1000 lncRNAs with the highest standard deviation over all samples.



Supplementary Figure 3. Poly(A) RNA-seq of T-ALL cell lines. Density plots of lncRNA expression in the DND-41 and ALL-SIL cell line, as cell lines representative for the *TLX1/3* subgroup and the PF-382, KE-37 and CCRF-CEM cell lines for the *TAL-R* subgroup. The vertical bars represent the top 10 lncRNAs that are significantly higher expressed in the representative subgroup compared to the others.



Supplementary Figure 4. Gene set enrichment analysis for the T-ALL cell lines. Gene set enrichment analysis using the top 50 lncRNAs up- or downregulated in the respective genetic T-ALL subgroups as gene sets within a panel of T-ALL cell lines.



Supplementary Figure 5. Poly(A) RNA-seq has no major advantage over micro-array data for subtype specific mRNA detection. (A) Overview of the amount of subtype specific mRNAs detected only by micro-array profiling (dark blue), only by RNA-sequencing (light blue) or by both methods (red). **(B)** Comparison between micro-array profiling and RNA-sequencing of the fold change detected for the significant mRNAs identified by both methods. The red line represents the distribution if the fold change would be equal (x=y).

3 COMPLETING THE T-ALL SUBTYPE SPECIFIC TRANSCRIPTOME BY MICRORNA PROFILING

PAPER 4

Comprehensive miRNA expression profiling in human T-cell acute lymphoblastic leukemia by small RNA-sequencing

Comprehensive miRNA expression profiling in human T-cell acute lymphoblastic leukemia by small RNA-sequencing

Annelynn Wallaert[1,2], Lucie Hernandez[3], Tom Taghon[2,4], Frank Speleman[1,2] and Pieter Van Vlierberghe[1,2]

[1] *Center for Medical Genetics Ghent, Ghent University, Ghent, Belgium*
[2] *Cancer Research Institute Ghent, Ghent, Belgium*
[3] *University Paris Diderot and Hospital Saint-Louis, U944 INSERM, Paris, France*
[4] *Department of Clinical Chemistry, Microbiology and Immunology, Ghent University, Ghent, Belgium*

Submitted to Scientific Reports

ABSTRACT

T-cell acute lymphoblastic leukemia (T-ALL) is a genetically heterogeneous disease that can be classified into different molecular genetic subtypes according to their mRNA gene expression profile. In this study, we applied RNA sequencing to investigate the full spectrum of miRNA expression in primary T-ALL patient samples, T-ALL leukemia cell lines and healthy donor thymocytes. Notably, this analysis revealed that genetic subtypes of human T-ALL also display unique miRNA expression signatures, which are largely conserved in human T-ALL cell lines with corresponding genetic background. Furthermore, small RNA-sequencing also unraveled the variety of isoforms that are expressed for each miRNA in T-ALL and showed that a significant number of miRNAs are actually represented by an alternative isomiR. Finally, comparison of $CD34^+$ and $CD4^+CD8^+$ healthy donor thymocytes and T-ALL miRNA profiles allowed identifying several novel miRNAs with putative oncogenic or tumor suppressor functions in T-ALL. Altogether, this study provides a comprehensive overview of miRNA expression in normal and malignant T-cells and sets the stage for functional evaluation of novel miRNAs in T-ALL disease biology.

INTRODUCTION

T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive hematological malignancy that is classified into different genetic subtypes based upon the aberrant expression of specific transcription factor oncogenes (*TAL*, *TLX1*, *TLX3* or *HOXA*) or the arrest at a specific stage of T-cell differentiation (immature T-ALL).[1-4] Notably, these molecular subgroups are characterized by unique mRNA and long noncoding RNA expression signatures, which partially reflect their putative cell of origin.[1, 2, 5]

MicroRNAs (miRNAs) are short noncoding RNAs that function as post-transcriptional repressors of specific target genes.[6, 7] Several studies have previously described a role for miRNAs in malignant T-cell transformation, including the identification of both an oncogenic (*miR-19b*, *mir-20a*, *miR-26a*, *miR-92* and *miR-223*) [8] as well as a tumor suppressor (miR-150, *miR-155*, *miR-200* and *miR-193b-3p*) [9, 10] miRNA network involved in T-ALL disease biology. However, studies that addressed the role of miRNAs in human T-ALL have largely been focused on previously recognized miRNA molecules as they consistently used RT-qPCR or microarrays as detection platforms.

More recently, small RNA-sequencing emerged as a more comprehensive technology that enables unbiased detection of the full spectrum of small RNA molecules. In addition, it also provides information on specific isoforms that differ from canonical miRNAs by the addition or deletion of one or more nucleotides at the 5' or 3' end of the miRNA.[11, 12] Notably, this heterogeneity in miRNA sequences, which is thought to result from RNA-editing, exonuclease activity or imprecise cleavage by DICER or DROSHA (ribonucleases involved in the miRNA processing) [13], could be functionally relevant as shown for a number of miRNAs.[14-17]

Here, we used small RNA-sequencing to study the full spectrum of miRNAs that are expressed in human T-ALL samples. We demonstrate, for the first time, that molecular genetic subtypes of human T-ALL are characterized by unique miRNA expression

signatures, delineate the pattern of miRNA isoforms that are expressed in malignant T-cells and use small RNA sequencing profiles of normal T-cell subsets to identify novel putative oncogenic or tumor suppressive miRNAs in the context of human T-ALL.

MATERIALS AND METHODS

Study design

Small RNA-sequencing was performed on 48 primary T-ALL patient samples, 7 T-ALL cell lines and 2 $CD34^+$ and 2 $CD4^+CD8^+$ healthy donor thymocyte subsets, to profile the full T-ALL miRNA transcriptome (Figure 1a).

Primary human T-ALL patient samples and T-ALL cell lines

Blood samples and bone marrow lymphoblast from T-ALL patients were collected after informed consent according to the Declaration of Helsinki from Saint-Louis Hospital, Paris, France. This study was approved by both the Institut Universitair d'Hématologie Institutional Review Board and the Ethical Committee of Ghent University Hospital. Total RNA was isolated using the miRNeasy mini kit (Qiagen). These samples are part of a cohort previously investigated by mRNA [18] and lncRNA profiling.[5, 19] The T-ALL cell lines LOUCY, DND-41, HPB-ALL, ALL-SIL, PF-382 and JURKAT were purchased from DSMZ. KE-37 was a kind gift from the Cools lab.

Thymocyte subset selection

Thymus tissue was derived from children undergoing cardiac surgery (UZ Gent) and was obtained and used according to the guidelines of the Medical Ethical Commission of Ghent University Hospital (Ghent, Belgium). Both thymocyte subsets were each purified from two different healthy donors in order to obtain two independent replicates for each subset. Immature $CD34^+$ thymocytes were purified based on MACS purification using CD34 microbeads (Miltenyi Biotec) [20] while CD4 and CD8 labeling was used to sort the $CD4^+CD8^+$ double positive subset using a FACSAriaIII (BDBiosciences).[21] The purity of each subset was at least 98%. Total RNA was isolated using the miRNeasy mini kit (Qiagen).

MicroRNA profiling by small RNA-sequencing

The libraries for small RNA-sequencing were prepared using the TruSeq small RNA library kit from Illumina with 50 ng of total RNA as input for T-ALL samples and 100 ng of total RNA as input for T-ALL cell lines and thymocyte subset samples. According to the manufacturer's use, 3' and 5' RNA adapters were ligated to the RNA followed by reverse transcription and PCR amplification (with barcoded primers). The PCR products were separated using a Pippin Prep System to recover the 147 nt and 157 nt fractions. Sequencing of the small RNA libraries was performed on a NextSeq500 (Illumina), with an average of 14.4 million reads per sample. After read quality control and adapter trimming, reads were mapped to the reference genome (GRCh38) using Bowtie.[22] Raw data files are submitted into the GEO database [23] with accession number GSE89978.

Differential expression analysis

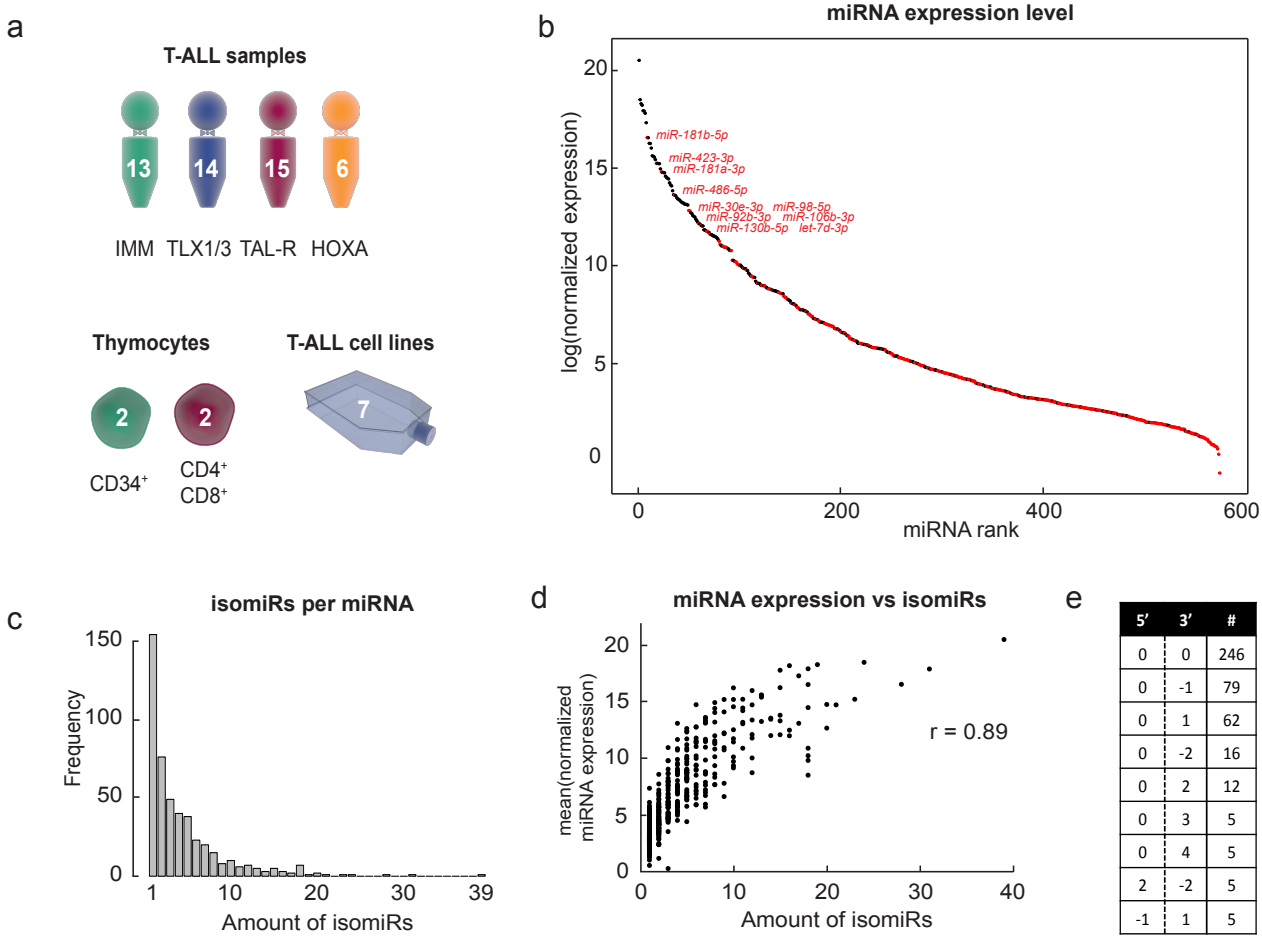
MicroRNA expression data was filtered with a background correction that only retained miRNAs detected by at least 4 reads in at least 60 % of samples from one T-ALL subgroup or in at least all samples from one thymocyte subset. Differential expression analysis was performed using the DESeq2 algorithm in R.[24] The expression was normalized using de Var-

iance Stabilizing Transformation from the DESeq2 algorithm.

RESULTS

Small RNA-sequencing of T-ALL patient samples, healthy donor thymocytes and T-ALL cell lines

To study the full spectrum of miRNAs involved in normal and malignant T-cell development, we performed small RNA-sequencing on 48 primary T-ALL patient samples of different T-ALL subgroups (13 immature, 14 *TLX1*$^{+}$ or *TLX3*$^{+}$, 15 *TAL*-rearranged and 6 *HOXA*-overexpressing T-ALL samples), CD34^{+} and CD4^{+}CD8^{+} normal thymocyte subsets from healthy donors and a panel of 7 T-ALL cell lines (Figure 1a). While 1816 miRNAs were initially detected in the total panel of T-ALL patient samples, we further only consid-



5'	3'	#
0	0	246
0	-1	79
0	1	62
0	-2	16
0	2	12
0	3	5
0	4	5
2	-2	5
-1	1	5

Fig 1. Small RNA-sequencing of T-ALL samples, healthy thymocytes and cell lines detects novel T-ALL miRNAs and isomiRs. (a) Overview of samples profiled by small RNA-sequencing. (b) Dot plot representing the mean normalized expression levels of all 574 miRNAs detected by small RNA-sequencing of the 48 T-ALL patient samples. Each dot represents one miRNA and the miRNAs are ranked from highest to lowest mean expression. Black dots are miRNAs that were already detected by a qRT-PCR platform from previous studies. Red dots are the novel miRNAs detected in T-ALL samples. (c) Bar plot visualizing the distribution of the miRNAs by means of the amount of isomiRs they are represented by. (d) Correlation plot between the mean expression level of the miRNAs and their amount of isomiR forms. (e) Table representing the isomiR form that was represented by the highest expression level for each of the detected miRNAs. The first column denotes the 5' overhangs or deletions, the second column the 3' overhang or deletions of the isomiR in comparison to the canonical miRNA. The third column shows the amount of miRNAs of which the highest expressed miRNA had that isomiR form. (Graphics from www.somersault1824.com)

ered the 574 miRNAs for which four reads were present in at least 60% of patients from one T-ALL subtype (Figure 1b).

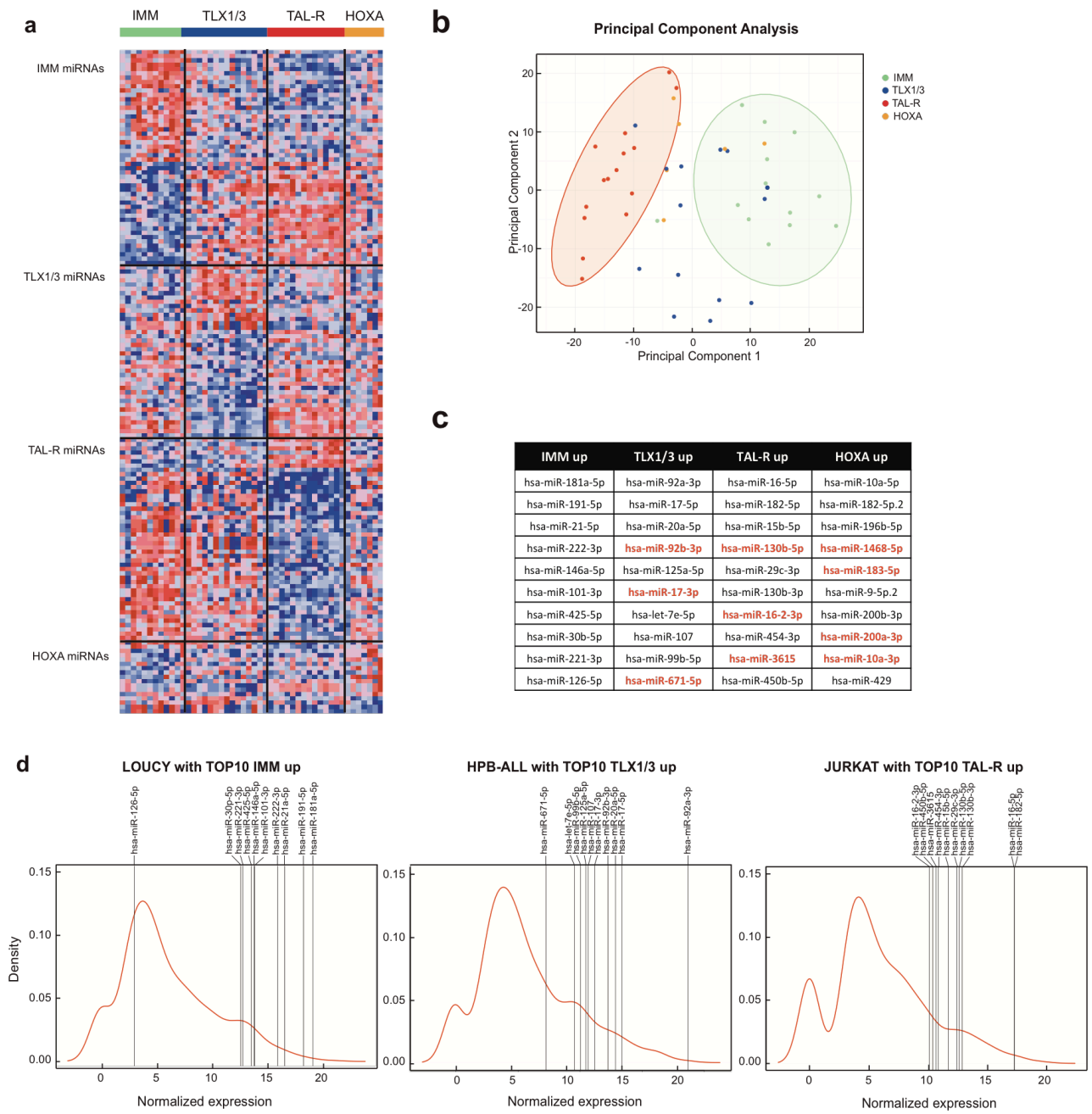
Given that for this same T-ALL patient cohort miRNA expression profiling was done previously using an RT-qPCR approach [10], we were able to directly compare the absolute number of detected miRNAs between the RT-qPCR and small RNA-sequencing platforms. Using a similar background selection as mentioned above, 283 miRNAs were detected by RT-qPCR [10], which reflects about half of the miRNA transcripts that were identified by small RNA-sequencing. Next, we converted the miRNA annotation of the RT-qPCR platform to the most recent annotation in miRBase by using the miRBaseTracker [25], which resulted in 248 unique miRNAs. There was an overlap of 198 miRNAs with small RNA-sequencing, implying that the RT-qPCR also detects 50 miRNAs that were not present in the small RNA-sequencing dataset.

In Figure 1b, the mean expression values for the 574 miRNAs, detected by small RNA-sequencing in primary T-ALL patients, are plotted, with the black dots representing the 280 miRNAs that were previously identified by RT-qPCR and the red dots representing the miRNAs detected for the first time in T-ALL samples. Most of the formerly known miRNAs reside amongst the most highly expressed in human T-ALL (Figure 1b). However, some of the newly T-ALL identified miRNA transcripts also show a very high expression pattern (the top 10 of highest expressed novel miRNAs are annotated in Figure 1b), suggesting that they might possess oncogenic potential in the context of T-ALL disease biology. Nevertheless, the average expression level of most novel T-ALL miRNAs that were exclusively detected by small RNA-sequencing is median to low.

Small RNA-sequencing also enables the detection of so-called isomiRs, i.e. miRNA isoforms that deviate from the canonical sequence by one or a few nucleotide(s).[11, 12] In our small RNA-sequencing dataset, we identified 2139 isomiRs covering 481 different miRNAs, losing some very low expressed miRNAs from the analysis above. Although 154 miRNAs were only represented by one isomiR, some others showed expression of more than 10 different isomiR forms (Figure 1c). We observed a positive correlation between the number of isomiRs detected for a specific miRNA and its expression level in human T-ALL (r = 0.89, Figure 1d). For example, we detected 39 different isomiR forms for *miR-181a-5p*, the miRNA that shows the highest average expression in T-ALL. Remarkably, for 106 out of 481 miRNAs, the canonical miRNA was not expressed in our patient series. In addition, for only half of the miRNAs (246 out of 481), the canonical isoform showed the highest expression level, suggesting that a substantial amount of miRNAs are mainly represented by alternative isomiRs. The distribution of isomiRs that show the highest expression for each miRNA is shown in Figure 1e.

Small RNA-sequencing reveals a subtype specific expression pattern of miRNAs in human T-ALL

Previous studies have convincingly shown that molecular genetic subtypes of human T-ALL display unique mRNA [1, 2] and lncRNA [5] expression signatures. Here, we used the 574 miRNAs detected by small RNA-sequencing to define subtype specific miRNA expression signatures in human T-ALL (adjusted p-value



IMM up	TLX1/3 up	TAL-R up	HOXA up
hsa-miR-181a-5p	hsa-miR-92a-3p	hsa-miR-16-5p	hsa-miR-10a-5p
hsa-miR-191-5p	hsa-miR-17-5p	hsa-miR-182-5p	hsa-miR-182-5p.2
hsa-miR-21-5p	hsa-miR-20a-5p	hsa-miR-15b-5p	hsa-miR-196b-5p
hsa-miR-222-3p	**hsa-miR-92b-3p**	**hsa-miR-130b-5p**	**hsa-miR-1468-5p**
hsa-miR-146a-5p	hsa-miR-125a-5p	hsa-miR-29c-3p	**hsa-miR-183-5p**
hsa-miR-101-3p	**hsa-miR-17-3p**	hsa-miR-130b-3p	hsa-miR-9-5p.2
hsa-miR-425-5p	hsa-let-7e-5p	**hsa-miR-16-2-3p**	hsa-miR-200b-3p
hsa-miR-30b-5p	hsa-miR-107	hsa-miR-454-3p	**hsa-miR-200a-3p**
hsa-miR-221-3p	hsa-miR-99b-5p	**hsa-miR-3615**	**hsa-miR-10a-3p**
hsa-miR-126-5p	**hsa-miR-671-5p**	hsa-miR-450b-5p	hsa-miR-429

Fig 2. Small RNA-sequencing of primary T-ALL samples reveals a subtype specific expression pattern of miRNAs. (a) Heatmap representing the top 50 most significantly up- or downregulated miRNAs per subgroup in comparison to the other subgroups (adjusted p-value < 0.05). (b) PCA-plot showing the distribution of the patient samples. The different colors denote patient samples from a different subgroup. (c) Table representing the selection of 10 miRNAs per subgroup. These were the highest expressed miRNAs that were significantly upregulated in that subgroup compared to the other subgroups. MiRNAs denoted in red were not detected by a previously used qRT-PCR platform. (d) Density plots representing the distribution of the miRNA expression in the LOUCY, HPB-ALL and JURKAT cell line. Vertical bars show the expression level of the top 10 miRNAs selected for the subgroup these cell lines represent. LOUCY represents the immature T-ALL subgroup, HPB-ALL the *TLX1/3* subgroup and JURKAT the *TAL-R* subgroup.

< 0.05; Figure 2a; Supplementary Table 1). Principal Component Analysis confirmed that the most pronounced differences in miRNA expression are observed between immature and *TAL-R* T-ALL patient samples (Figure 2b). The 10 miRNAs that show the highest expression level in each of the genetic subtypes are shown in Figure 2c. The miRNAs, which were not previously detected by the RT-qPCR platform, are depicted in red (Figure 2c).

As mentioned above, small RNA-sequencing was also performed on a panel of 7 human T-ALL cell lines. These *in vitro* model systems reflect most of the different genetic subtypes of human T-ALL and included *TLX1/3* positive (ALL-SIL, DND41 and HPB-ALL), *TAL-R* positive (PF-382, JURKAT and KE-37) and immature/*HOXA* overexpressing (LOUCY) tumor lines. Notably, subtype specific miRNAs, which were identified in primary T-ALL patient samples, also showed high expression in the T-ALL cell lines from their corresponding genetic subtype (Figure 2d and Supplementary Figure 1). Therefore, the subtype specific tumor lines can be used as valuable *in vitro* tools to evaluate the role of specific miRNAs in the pathogenesis of this disease.

MiRNA profiling of normal thymocyte subsets reveals oncogenic subtype specific miRNAs

Small RNA-sequencing was also performed on $CD34^+$ and $CD4^+CD8^+$ normal thymocyte samples from two healthy donor controls. First, DESeq2 analysis revealed 190 miRNAs that show significant differential expression between these $CD34^+$ and $CD4^+CD8^+$ normal T-cell subsets (126 high in $CD34^+$ and 64 high in $CD4^+CD8^+$; adjusted p-value < 0.05; Figure 3a and Supplementary Table 2). The top ten most significant miRNAs are listed in Figure 3b, with the miRNAs depicted in red those that were not covered in previous RT-qPCR analyses. Of note, small RNA-sequencing data

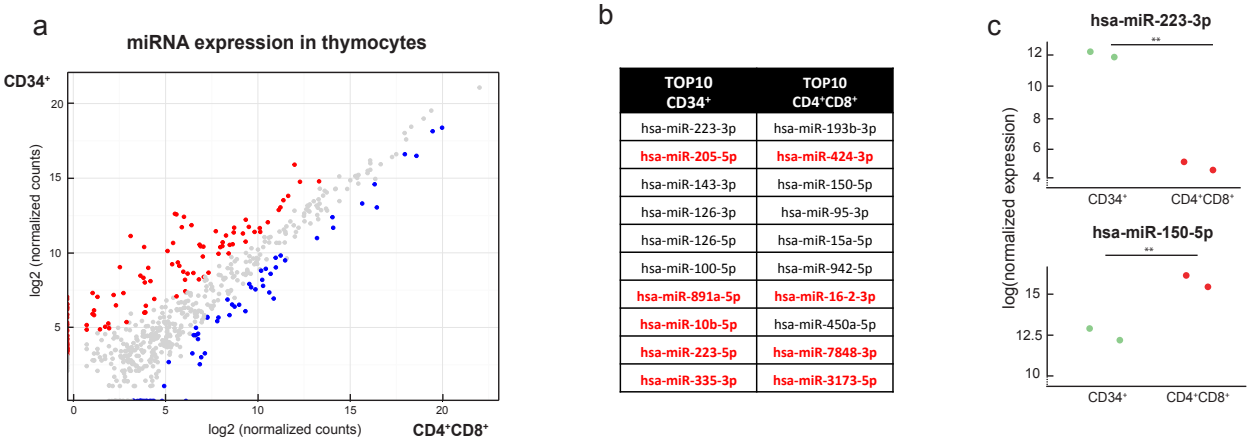


Fig 3. MiRNA profiling of healthy thymocyte subsets reveals different miRNA expression profiles between $CD34^+$ and $CD4^+CD8^+$ subsets. (a) Diagonal plot showing the expression of the miRNAs in the different thymocyte subsets. Red dots represent miRNAs that are significantly higher expressed in the $CD34^+$ subset, blue dots are the miRNAs significantly higher expressed in the $CD4^+CD8^+$ subset. (b) Top 10 most significantly upregulated miRNAs for the $CD34^+$ subset and for the $CD4^+CD8^+$ subset. MiRNAs denoted in red were not detected by a previously used qRT-PCR platform. (c) Dot plots of two representative miRNAs for the subsets. **: significant difference with an adjusted p-value < 0.001.

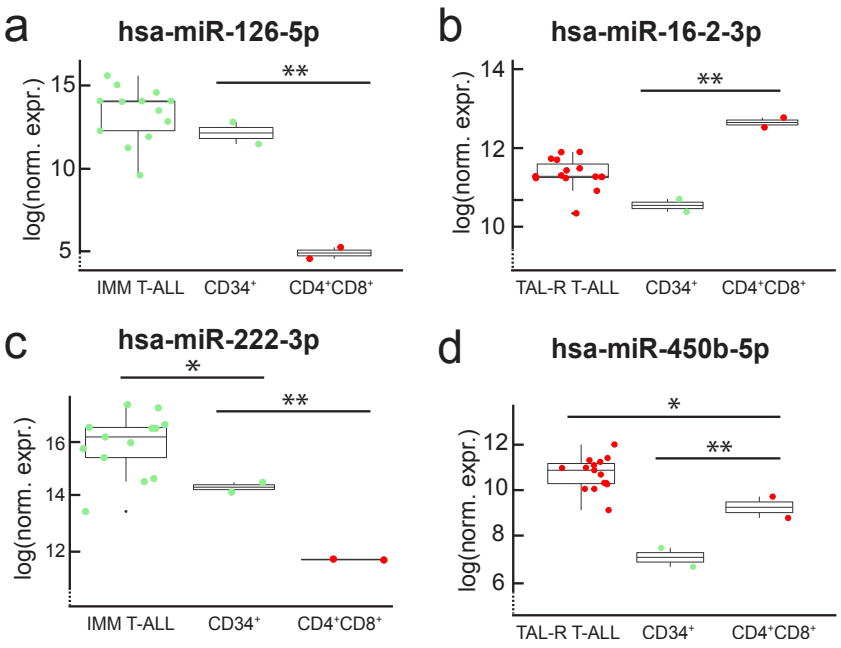
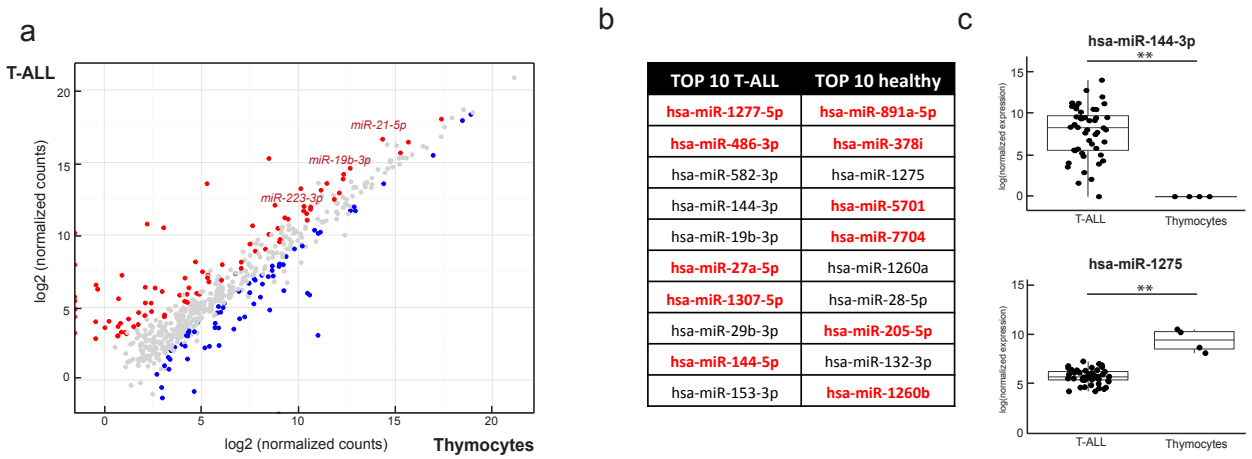


Fig 4. Subtype specific miRNAs can be either oncogenic or representative for the differentiation arrest of lymphoblasts. (a+c) Box plots showing the expression of the immature T-ALL specific miRNAs hsa-miR-126-5p and hsa-miR-222-3p in the immature T-ALL patients and the thymocyte subsets. (b+d) Box plots showing the expression of the *TAL-R* T-ALL specific miRNAs hsa-miR-16-2-3p and hsa-miR-450b-5p in the *TAL-R* T-ALL patients and the thymocyte subsets. *: Significant difference with an adjusted p-value < 0.05; **: Significant difference with an adjusted p-value < 0.001.

were highly concordant between both donors for each subset (Figure 3c and Supplementary Figure 2).

Next, we integrated these miRNA expression data obtained from healthy donors with the subtype specific miRNAs that were identified in the primary T-ALL patient cohort. For example, 4 miRNAs (*hsa-miR-222-3p*, *hsa-miR-146a-5p*, *hsa-mir-221-3p* and *hsa-miR-126-5p*) from the top 10 immature T-ALL specific miRNAs (Figure 2c) also showed significant higher expression in $CD34^+$ vs. $CD4^+CD8^+$ T-cell subsets (Figure 4a). Similarly, three miRNAs (*hsa-miR-16-5p*, *hsa-miR-16-2-3p* and *hsa-miR-450b-5p*) from the top 10 *TAL-R* specific miRNAs (Figure 2c) are significantly upregulated in $CD4^+CD8^+$ T-cell subsets (Figure 4b). Therefore, these miRNAs most probably reflect the specific T-cell maturation arrest associated with these molecular genetic subtypes of T-ALL and their respective cell of origin.

In order to identify miRNAs with true oncogenic potential in specific T-ALL subgroups, we also performed differential expression analysis between immature T-ALLs and $CD34^+$ thymocytes, and between *TAL-R* T-ALLs and $CD4^+CD8^+$ thymocytes, as these thymocyte subsets represent the stage of differentiation arrest during T cell development leading to these specific T-ALL subtypes (Supplementary Tables 3 and 4). From the top ten immature specific miRNAs, three miRNAs (*hsa-miR-21-5p*, *hsa-miR-222-3p* and *hsa-miR-101-3p*) were significantly upregulated in the immature samples compared to the healthy control $CD34^+$ samples. Remarkably, *hsa-miR-222-3p* was also significantly upregulated in the $CD34^+$ subset, but its expression is further increased in immature T-ALLs (Figure 4c). Similarly, three miRNAs (*hsa-miR-182-5p*,



TOP 10 T-ALL	TOP 10 healthy
hsa-miR-1277-5p	hsa-miR-891a-5p
hsa-miR-486-3p	hsa-miR-378i
hsa-miR-582-3p	hsa-miR-1275
hsa-miR-144-3p	hsa-miR-5701
hsa-miR-19b-3p	hsa-miR-7704
hsa-miR-27a-5p	hsa-miR-1260a
hsa-miR-1307-5p	hsa-miR-28-5p
hsa-miR-29b-3p	hsa-miR-205-5p
hsa-miR-144-5p	hsa-miR-132-3p
hsa-miR-153-3p	hsa-miR-1260b

Fig 5. Novel oncogenic miRNAs are detected by small RNA-sequencing. (a) Diagonal plot showing the expression of the miRNAs in T-ALL compared to healthy donor thymocytes. Red dots represent miRNAs that are significantly higher expressed in the T-ALL patient samples, blue dots are the miRNAs significantly higher expressed in thymocyte subsets. (b) Top 10 most significantly upregulated miRNAs in the T-ALL patients and in the healthy donor samples. MiRNAs denoted in red were not detected by a previously used qRT-PCR platform. (c) Dot plots of two representative miRNAs from the table. **: Significant difference with an adjusted p-value < 0.001.

hsa-miR-29c-3p and *hsa-miR-450b-5p*), from the *TAL-R* subtype specific signature, show significant higher expression in the *TAL-R* T-ALLs as compared to their $CD4^+CD8^+$ normal counterparts. Here, *hsa-miR-450b-5p* was already higher expressed in the $CD4^+CD8^+$ double positive subset, but showed a further increase in activity in *TAL-R* rearranged leukemias (Figure 4d).

Small RNA-sequencing reveals putative oncogenic miRNAs in human T-ALL

Finally, we aimed to identify novel T-ALL miRNAs with potential oncogenic activity in human T-ALL irrespective of the genetic subtypes. Differential expression analysis between 48 T-ALL samples and four normal thymocyte samples (Supplementary Table 5) resulted in the identification of 87 significantly upregulated miRNAs and 69 downregulated miRNAs in human T-ALL (Figure 5a). Several miRNAs with a known oncogenic role in T-ALL [8, 26] were recovered from this analysis and are depicted in Figure 5a. However and most notably, this analysis also identified different miRNAs, which were not previously detected in the context of normal and malignant T cell development, and which could potentially act as novel oncomiRs or tumor suppressor miRNAs involved in the biology of this disease (Figure 5b-c).

DISCUSSION

More than a decade ago, Ferrando et al. and Soulier et al. described different T-ALL subtypes according to specific transcriptional profiles.[1, 2] Last year, we were able to show that these molecular genetic subtypes of human T-ALL also display unique long non-coding RNA expression signatures.[5] Here, we performed small RNA-sequencing on 48 T-ALL patient samples to finalize the transcriptional characterization of human T-ALL by a comprehensive analysis of miRNA expression in this disease.

Small RNA-sequencing enabled the detection of twice as many miRNAs as compared to RT-qPCR platforms.[10] Although most of the new-

ly detected miRNAs were expressed at very low levels, we also identified a selection of novel T-ALL miRNAs with very high expression levels in T-ALL patient samples, including *hsa-miR-181b-5p*, *hsa-miR-423-3p*, *hsa-miR-486-5p* and *hsa-miR-92b-3p* (Figure 1b). Interestingly, some of these miRNAs have previously been associated with malignant transformation in different tumor entities. For example, *hsa-miR-181b-5p* is a known oncogene in several cancer types as reviewed by Liu et al..[27] In the context of leukemia, overexpression of *hsa-miR-181b-5p* was shown to enhance proliferation in acute myeloid leukemia by targeting *MLK2*. Furthermore, *hsa-miR-92b-3p* was identified as an oncomiR in glioblastoma by targeting SMAD3 (ref. 28), which is known to be lost in several cases of pediatric T-ALL (ref. 29), and *PTEN* (ref. 30), a well-established T-ALL oncosuppressor.[31, 32] In addition, *hsa-miR-92b-3p* is also specifically higher expressed in *TLX1/3* T-ALL compared to the other T-ALL subtypes.

Interestingly, small RNA-sequencing also detects deviations from canonical miRNA sequences. Indeed, our analysis revealed 2139 different isomiR forms, corresponding to 481 miRNAs. Remarkably, for only half of the miRNAs, the canonical form showed the highest expression (Figure 1e). In addition, most highly expressed isomiRs displayed modifications at their 3' end, suggesting that these alterations would not affect the miRNA seed sequence and, therefore, have no functional effect on target recognition. Nevertheless, a number of studies have shown that these 3' modifications might impact target specificity and stability of the miRNA.[17, 33]

Finally, small RNA sequencing of human $CD34^+$ and $CD4^+CD8^+$ thymocytes enabled the identification of subtype specific and oncogenic miRNAs in the context of human T-ALL. An interesting example is *hsa-miR-486*. *Hsa-miR-486-5p* is one of the highest expressed newly identified miRNAs in T-ALL (Figure 1b) and, together with *hsa-miR-486-3p*, it also shows higher expression in T-ALL samples as compared to healthy donor thymocytes. *Hsa-miR-486-5p* is an oncomiR in Down syndrome myeloid leukemias, where it is regulated by GATA1 (ref. 34). *Hsa-miR-486-3p* has also been linked to erythroid development downstream of MYB, a known oncogene in T-ALL, and targeting *MAF* (ref. 35).

Altogether, this study provides the first comprehensive overview of miRNA expression in molecular-genetic subtypes of human T-ALL. Integration of these signatures with miRNA expression profiles in normal T-cell subsets provides a unique resource to study novel miRNAs that are implicated in T-ALL disease biology.

Supplementary tables can be requested via e-mail.

ACKNOWLEDGEMENTS

The authors would like to thank the following funding agencies: Fund for Scientific Research Flanders ('FWO Vlaanderen' research projects GA00113N, 3G065614, G0C4713N, G022216N and 31500615W to PVV; 3G056413N and 3G055013N to FP; PhD grant to AW), Ghent University Research Fund (GOA LNCCA), the Belgian Program of Interuniversity Poles of Attraction (IUAP P7/03) and the Childhood Cancer Fund Ghent (Kinderkankerfonds VZW)

Authors' contributions: AW analyzed and interpreted the data and wrote the manuscript, LH and TT were responsible for sample collection and manipulation, FS and PV interpreted the data and wrote the manuscript.

All authors read and approved the final manuscript.

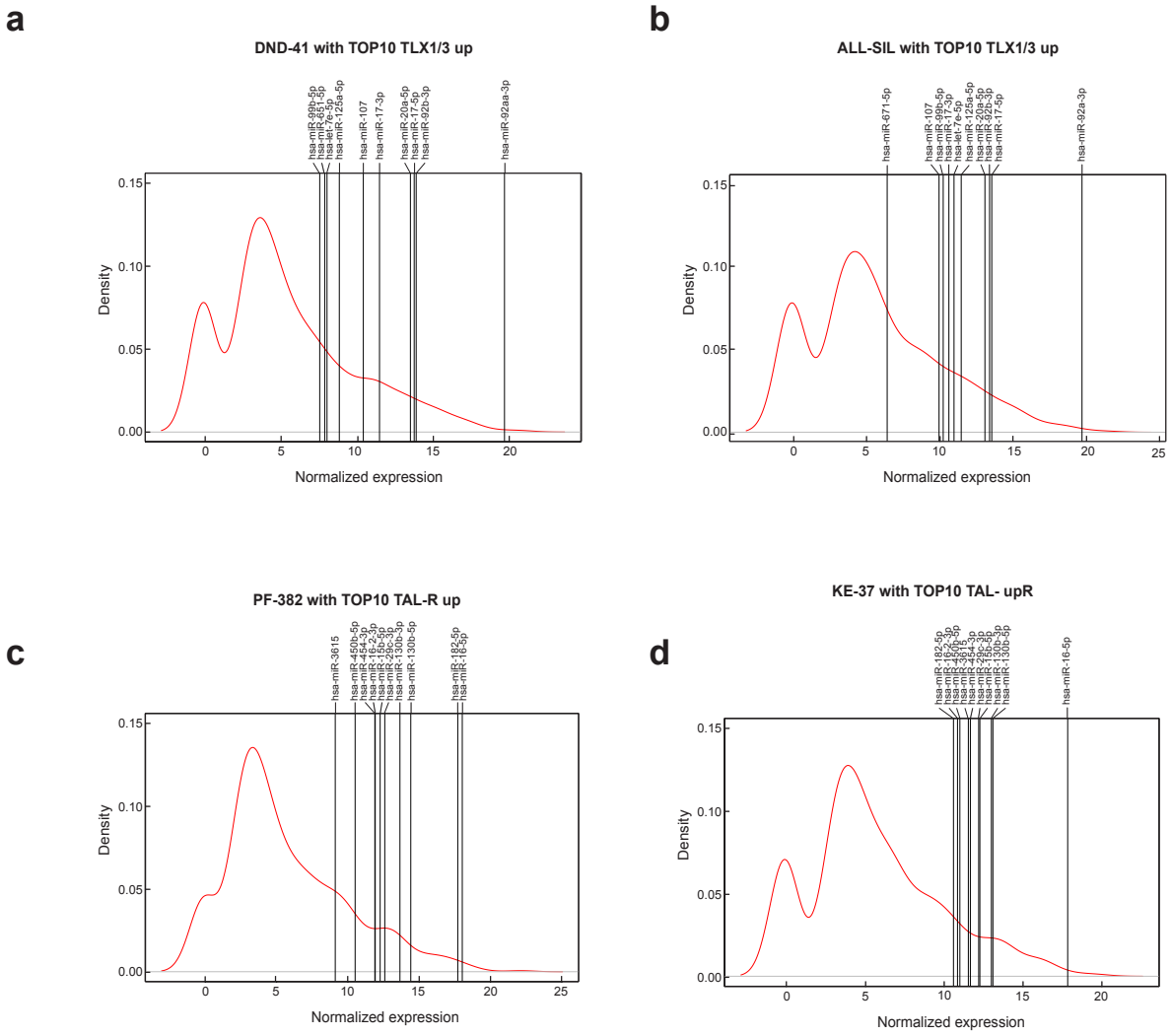
Competing interests: The authors have no competing interests.

REFERENCES

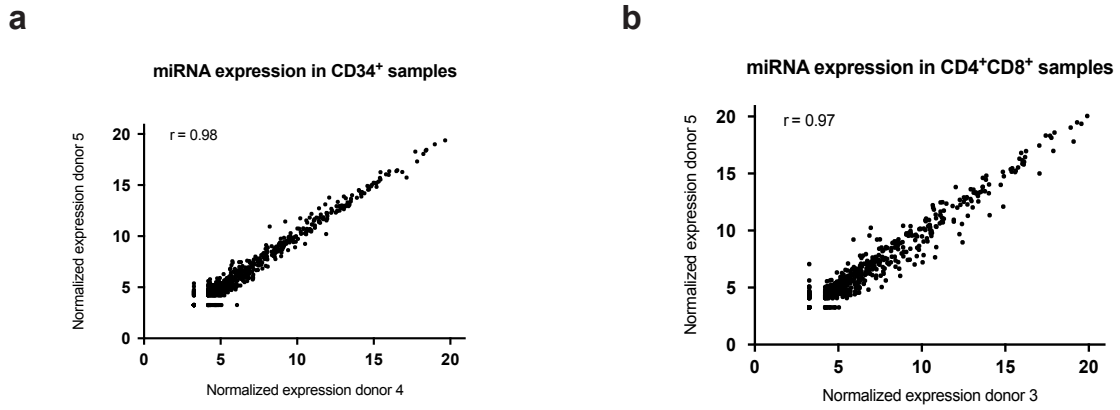
1. Soulier J, Clappier E, Cayuela JM, Regnault A, Garcia-Peydro M, Dombret H, et al. HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). Blood. 2005;106(1):274-86.
2. Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC, et al. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer Cell. 2002;1(1):75-87.
3. Van Vlierberghe P, Pieters R, Beverloo HB, Meijerink JP. Molecular-genetic insights in paediatric T-cell acute lymphoblastic leukaemia. Br J Haematol. 2008;143(2):153-68.
4. Meijerink JP. Genetic rearrangements in relation to immunophenotype and outcome in T-cell acute lymphoblastic leukaemia. Best Pract Res Clin Haematol. 2010;23(3):307-18.
5. Wallaert A, Durinck K, Van Loocke W, Van de Walle I, Matthijssens F, Volders PJ, et al. Long noncoding RNA signatures define oncogenic subtypes in T-cell acute lymphoblastic leukemia. Leukemia. 2016;30(9):1927-30.
6. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer. 2006;6(4):259-69.
7. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014;15(8):509-24.
8. Mavrakis KJ, Van Der Meulen J, Wolfe AL, Liu X, Mets E, Taghon T, et al. A cooperative microRNA-tumor suppressor gene network in acute T-cell lymphoblastic leukemia (T-ALL). Nat Genet. 2011;43(7):673-8.
9. Sanghvi VR, Mavrakis KJ, Van der Meulen J, Boice M, Wolfe AL, Carty M, et al. Characterization of a set of tumor suppressor microRNAs in T cell acute lymphoblastic leukemia. Sci Signal. 2014;7(352):ra111.
10. Mets E, Van der Meulen J, Van Peer G, Boice M, Mestdagh P, Van de Walle I, et al. MicroRNA-193b-3p acts as a tumor suppressor by targeting the MYB oncogene in T-cell acute lymphoblastic leukemia. Leukemia. 2015;29(4):798-806.
11. Neilsen CT, Goodall GJ, Bracken CP. IsomiRs--the overlooked repertoire in the dynamic microRNAome. Trends Genet. 2012;28(11):544-9.
12. Lee LW, Zhang S, Etheridge A, Ma L, Martin D, Galas D, et al. Complexity of the microRNA repertoire revealed by next-generation sequencing. RNA. 2010;16(11):2170-80.
13. Starega-Roslan J, Witkos TM, Galka-Marciniak P, Krzyzosiak WJ. Sequence features of Drosha and Dicer cleavage sites affect the complexity of isomiRs. Int J Mol Sci. 2015;16(4):8110-27.
14. Wyman SK, Knouf EC, Parkin RK, Fritz BR, Lin DW, Dennis LM, et al. Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. Genome Res. 2011;21(9):1450-61.
15. Tan GC, Chan E, Molnar A, Sarkar R, Alexieva D, Isa IM, et al. 5' isomiR variation is of functional and evolutionary importance. Nucleic Acids Res. 2014;42(14):9424-35.
16. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22(9):1775-89.
17. Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Nishibu T, Ukekawa R, et al. A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. Genome Res. 2010;20(10):1398-410.
18. Clappier E, Gerby B, Sigaux F, Delord M, Touzri F, Hernandez L, et al. Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. J Exp Med. 2011;208(4):653-61.
19. Durinck K, Wallaert A, Van de Walle I, Van Loocke W, Volders PJ, Vanhauwaert S, et al. The Notch driven long non-coding RNA repertoire in T-cell acute lymphoblastic leukemia. Haematologica. 2014;99(12):1808-16.
20. Van de Walle I, Waegemans E, De Medts J, De Smet G, De Smedt M, Snauwaert S, et al. Specific Notch receptor-ligand interactions control human TCR-alphabeta/gammadelta development by inducing differential Notch signal strength. J Exp Med. 2013;210(4):683-97.
21. Taghon T, Waegemans E, Van de Walle I. Notch signaling during human T cell development. Curr Top Microbiol Immunol. 2012;360:75-97.

22. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
23. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207-10.
24. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
25. Van Peer G, Lefever S, Anckaert J, Beckers A, Rihani A, Van Goethem A, et al. miRBase Tracker: keeping track of microRNA annotation changes. Database (Oxford). 2014;2014.
26. Mavrakis KJ, Wolfe AL, Oricchio E, Palomero T, de Keersmaecker K, McJunkin K, et al. Genome-wide RNA-mediated interference screen identifies miR-19 targets in Notch-induced T-cell acute lymphoblastic leukaemia. Nat Cell Biol. 2010;12(4):372-9.
27. Liu J, Shi W, Wu C, Ju J, Jiang J. miR-181b as a key regulator of the oncogenic process and its clinical implications in cancer (Review). Biomed Rep. 2014;2(1):7-11.
28. Wu ZB, Cai L, Lin SJ, Lu JL, Yao Y, Zhou LF. The miR-92b functions as a potential oncogene by targeting on Smad3 in glioblastomas. Brain Res. 2013;1529:16-25.
29. Wolfraim LA, Fernandez TM, Mamura M, Fuller WL, Kumar R, Cole DE, et al. Loss of Smad3 in acute T-cell lymphoblastic leukemia. N Engl J Med. 2004;351(6):552-9.
30. Song H, Zhang Y, Liu N, Wan C, Zhang D, Zhao S, et al. miR-92b regulates glioma cells proliferation, migration, invasion, and apoptosis via PTEN/Akt signaling pathway. J Physiol Biochem. 2016;72(2):201-11.
31. Zuurbier L, Petricoin EF, 3rd, Vuerhard MJ, Calvert V, Kooi C, Buijs-Gladdines JG, et al. The significance of PTEN and AKT aberrations in pediatric T-cell acute lymphoblastic leukemia. Haematologica. 2012;97(9):1405-13.
32. Silva A, Yunes JA, Cardoso BA, Martins LR, Jotta PY, Abecasis M, et al. PTEN posttranslational inactivation and hyperactivation of the PI3K/Akt pathway sustain primary T cell leukemia viability. J Clin Invest. 2008;118(11):3762-74.
33. Lu S, Sun YH, Chiang VL. Adenylation of plant miRNAs. Nucleic Acids Res. 2009;37(6):1878-85.
34. Shaham L, Vendramini E, Ge Y, Goren Y, Birger Y, Tijssen MR, et al. MicroRNA-486-5p is an erythroid oncomiR of the myeloid leukemias of Down syndrome. Blood. 2015;125(8):1292-301.
35. Bianchi E, Bulgarelli J, Ruberti S, Rontauroli S, Sacchi G, Norfo R, et al. MYB controls erythroid versus megakaryocyte lineage fate decision through the miR-486-3p-mediated downregulation of MAF. Cell Death Differ. 2015;22(12):1906-21.

SUPPLEMENTARY FIGURES



Supplementary Figure 1. Subtype specific miRNAs are also represented in the T-ALL cell lines. Density plots representing the distribution of the miRNA expression in the T-ALL cell lines DND-41, ALL-SIL, PF-382 and KE-37. Vertical bars show the expression level of the top 10 miRNAs selected for the subgroup these cell lines represent. DND-41 (a) and ALL-SIL (b) for the TLX subgroup and PF-382 (c) and KE-37 (d) for the TAL/LMO subgroup.



Supplementary Figure 2. miRNA expression is highly correlated in 2 different donors. Scatterplots representing the log2 normalized expression of the miRNAs profiled by small RNA-sequencing for 2 different donors of CD34+ thymocytes (a) and CD4+CD8+ thymocytes (b).

CHAPTER 5
DISCUSSION AND FUTURE PERSPECTIVES

DISCUSSION

In the past decades, the genetic basis of T-cell acute leukemia has been extensively investigated, first by cytogenetic and molecular cytogenetic techniques and more recently by next-generation sequencing technologies. This led to the identification of several key oncogenic driver mutations and allowed to identify several T-ALL genetic subtypes. Also, the identification of the genes implicated in T-ALL provided fundamental insights into the process of normal and perturbed thymocyte differentiation and oncogenesis. Importantly, this work also offered insights into novel therapeutic targets. In addition to protein coding genes, the role of noncoding RNAs in T-ALL has also received growing attention. Previous studies of the host lab and others have shed light on the role of multiple miRNAs in T-ALL leukemogenesis. In this PhD thesis, I aimed to further explore the potential role of noncoding RNAs in T-ALL, including a novel class of noncoding RNAs, the so-called long noncoding RNAs (lncRNAs) that recently gained much interest in the field of cancer research.

First, I elucidated the set of lncRNAs driven by NOTCH1-signaling both in T-ALL and normal T-cell development **(paper 1)**. For this study, 64 primary T-ALL patient samples were profiled on a custom designed micro-array detecting the expression of mRNAs and lncRNAs. Secondly, I explored this dataset in more depth to unravel subtype specific expression patterns of lncRNAs in those primary T-ALLs, as it was already described that the distinct T-ALL subgroups are defined by a subtype specific gene expression profile **(paper 2)**. As next-generation sequencing technologies became cheaper in the last years, RNA-sequencing is now the preferred technology over micro-array gene expression profiling. Because of this, I reprofiled 60 samples of the previous dataset by poly(A) RNA-sequencing and 25 samples by total RNA-sequencing to (1) define a novel T-ALL subtype specific lncRNA expression profile and (2) compare technologies **(paper 3)**. We also hypothesized that there should be a subtype specific expression pattern for some miRNAs, as they are described to be regulated by T-ALL genes or to regulate T-ALL genes themselves. However, previous studies could not define these profiles after RT-qPCR profiling of miRNA expression in a primary T-ALL patient cohort. Therefore, I performed small RNA-sequencing on 48 primary patient samples to detect a broader range of miRNAs than previous studies and with this dataset I was then able to find miRNAs with a specific expression in T-ALL patient subgroups **(paper 4)**. An overview of the matching samples used in the different projects can be found in Figure 11.

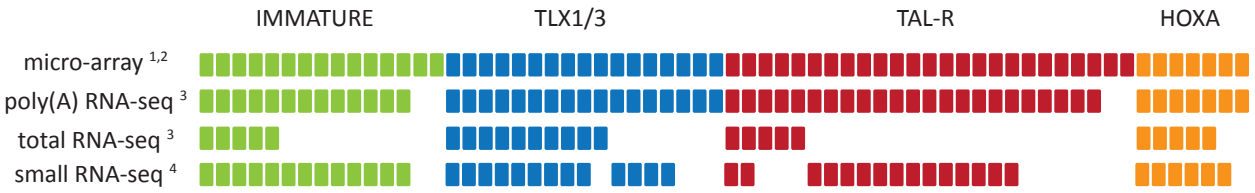


Figure 11. Overview of the primary T-ALL patient samples used in this thesis. The patient samples from the same cohort were used to study the expression of lncRNAs (micro-array, poly(A) and total RNA-seq) and miRNAs (small RNA-seq) in T-ALL subtypes. [1,2,3,4]: papers in which these datasets are used.

Defining the NOTCH1-regulated lncRNAome in T-ALL

The NOTCH1-signaling pathway is of major interest in T-ALL research as it is overactivated in more than half of T-ALL patients, making it a strong candidate therapeutic target[1]. The NOTCH1-receptor is present at the membrane of the immature T-cell and is activated through its ligand(s) in the thymus microenvironment. This leads to a signaling cascade necessary in the first steps towards T-cell commitment and cell proliferation in the early stages of T-cell development[2]. Overactivation of this signaling pathway or ligand independent signaling leads however to the uncontrolled proliferation of these immature thymocytes and aberrant signaling through the NOTCH1 receptor has been shown sufficient to drive T-cell leukemia development *in vivo*[3].

Pharmacological inhibition of hyperactive NOTCH1-signaling has been explored in the past decade by the use of GSIs (γ-secretase inhibitors) and has been tested in the clinic. Although promising results were obtained through *in vitro* evaluation of GSIs, clinical trials have been terminated due to gastro-intestinal toxicity and treatment resistance[4, 5]. This gastro-intestinal toxicity is due to the presence of the NOTCH1 and NOTCH2 receptors on the intestinal epithelium, which are both targeted by GSIs. In this PhD thesis, I used a GSI (Compound E) as a tool to inhibit NOTCH1-signaling *in vitro* in the T-ALL cell line CUTLL1 and evaluated the transcriptional effects on downstream lncRNA expression (see **paper 1**). Next to that, we activated NOTCH signaling in $CD34^+$ T-cells by growing them on a feeder layer containing the DLL1-ligand. The NOTCH1-activated lncRNAs were further studied in different T-ALL cell lines, T-ALL patient samples and T-cell subsets that were profiled by a custom designed micro-array platform detecting the expression of mRNAs and lncRNAs. Publicly available ICN1 (activated NOTCH1) ChIP-sequencing data allowed me to determine the lncRNAs that were directly regulated by NOTCH1, of which almost all also had BRD4 and MED1 binding and H3K27 acetylation, hinting towards a possible enhancer locus. It should however be noted that some recent publications question the function of eRNAs or *cis*-acting lncRNAs, as only the act of transcription at that locus is important and not the transcript itself[6, 7]. The use of this ICN1 ChIP-sequencing data could also filter out lncRNAs that were downregulated after GSI treatment due to the regulation by other NOTCH receptors instead of the NOTCH1 receptor. The NOTCH3 receptor is also present at the T-ALL cell membranes, but at lower levels than NOTCH1. This could however also lead to an extra effect of the GSI, however this is expected to be modest. Also the other pathways that could be targeted by GSIs (eg. Beta-Amyloid) are not expressed in the T-ALL cells.

Further in depth study should give more insights on the real function and importance of these lncRNAs. By the lncRNA and mRNA profiling of 64 T-ALL patient samples, we could perform the so-called 'guilt-by-association' analysis to identify the pathways these lncRNAs are correlated with, showing that several of the NOTCH1-activated lncRNAs might be involved in cancer-related pathways or in T-cell development.

At the same time of our study, the Aifantis lab (New York, USA) published the identification and mechanism-of-action of the T-cell specific lncRNA '*Leukemia Induced Noncoding Activator RNA 1*' (*LUNAR1*), a NOTCH1-activated lncRNA that regulates the expression of *IGF1R* (*Insulin-like*

Growth Factor 1 Receptor) in T-ALL[8]. *IGF1R* was previously identified as regulated by NOTCH1-signaling and pharmacological inhibition of IGF1R could block the growth and viability of T-ALL cells[9]. *LUNAR1* was also found in our dataset as the top-candidate amongst the set of NOTCH1-activated lncRNAs both in the applied *in vitro* models for T-ALL and normal T-cell development as well as in the dataset of the primary T-ALL patient cohort, underscoring the validity of our analyses. The research team of prof. Aifantis could show that there was a physical interaction between the *LUNAR1* promoter and an active enhancer in the last intron of its neighboring gene *IGF1R*. The knockdown of *LUNAR1*, but not its overexpression, could modulate the *IGF1R* expression, showing a *cis*-regulatory effect. Furthermore, it was shown that *LUNAR1* was involved in the recruitment of the Mediator complex and RNAPII to the *IGF1R* enhancer, leading to increased *IGF1R* expression.

Briefly after the discovery of *LUNAR1*, another lncRNA was linked to the NOTCH1-pathway (*NOTCH1-associated lncRNA in T-ALL* or *NALT*). The *NALT* transcript is bi-directionally transcribed from the *NOTCH1* locus[10]. The knockdown of *NALT in vitro* and *in vivo* inhibited cell proliferation, potentially due to *cis*-regulation of *NOTCH1* mRNA transcription as it showed a reduction in the NOTCH1 downstream targets. This study thus further provides another example that lncRNAs might play a crucial role in NOTCH1-driven signaling. *NALT* knockdown might be a novel therapeutical approach, however the expression level and function of *NALT* in other NOTCH1-dependent tissues should be checked to avoid off-target effect.

The identification of lncRNAs with a role in NOTCH1-signaling in T-ALL can also be of interest for other cancer types. NOTCH1 and other NOTCH-receptors have been described as oncogenes in different hematological malignancies, with the exception of acute myeloid leukemia (AML) in which NOTCH1 plays a tumor suppressive role[11]. Also in solid tumors, for example breast cancer and lung adenocarcinoma, *NOTCH1* overactivation has been linked to an increased proliferation and a restricted differentiation[12]. In these other cancer types, several groups have recently also studied the role of lncRNAs in NOTCH1-signaling. For example, in glioma, the '*Taurine Upregulated Gene 1' (TUG1)* lncRNA is activated by NOTCH1 and acts as a sponge for *miR-145*, increasing the expression of stemness related genes *SOX2* and *MYC*. The knockdown of *TUG1* induces apoptosis in glioma stem cells *in vitro* and could inhibit tumorigenesis *in vivo* more strongly than GSIs[13].

In this study, we thus explored lncRNAs with a potential role in the NOTCH1-signaling cascade in malignant and normal T-cell development. This might aid in the identification of novel druggable lncRNAs to inhibit the oncogenic mechanism of NOTCH1 in T-ALL. One example is *LUNAR1* that was also put forward as a valuable target in our study.

Unraveling T-ALL subtype specific expression of long noncoding RNAs

Whereas *NOTCH1* mutations are recurrent in all T-ALL patient subgroups, key driver oncogenes have been described that define the signature of a particular subtype such as *TLX1* or *TLX3* overexpression, *TAL1*-rearrangements or *HOXA* overexpression. These oncogenic events are mutually exclusive. The research teams of prof. Ferrando (NY, USA) and prof. Soulier (Paris,

France) previously published a detailed analysis of protein-coding gene expression in primary T-ALL patient samples by micro-array based profiling and could show that these expression profiles clustered together according to their oncogenic driver event and in the meantime revealed *TLX3* and the *HOXA* genes as novel T-ALL oncogenes[14, 15]. These subtypes do not only have a specific gene expression profile, but also resemble a specific stage of differentiation arrest during T-cell development. In comparison to the previously stated subtypes with a specific oncogenic driver event, one subtype, immature T-ALL, is solely identified by a differentiation arrest very early during T-cell development at the $CD34^+$ stage. Nevertheless, also this group of patients shows a subtype specific clustering of mRNA expression profiles compared to the other groups[14].

We hypothesized that lncRNA expression profiles in T-ALL patients might also allow to discriminate the previously established T-ALL subtypes and that subtype specific lncRNAs might be potent candidates as novel players in T-ALL development. Therefore, I dissected lncRNA expression profiles obtained from 64 primary T-ALL patients by means of a custom in-house designed micro-array platform (**paper 2**). This was the first time that a large T-ALL patient cohort was profiled for lncRNA expression and the first time that a comprehensive study was performed on the lncRNA expression in T-ALL subtypes. Next to that, I also had access to expression profiles of healthy donor T-cells from the $CD34^+$ or $CD4^+CD8^+$ subsets as controls.

First, I validated our dataset from the primary T-ALL cohort by evaluating known subset specific expression patterns of protein coding genes. One disadvantage of our dataset appeared to be the underrepresentation of *HOXA*-overexpressing cases (7 out of 64) making it more difficult to identify genes up- or downregulated for this subtype. Nevertheless, several HOXA-genes were identified as significantly upregulated in the HOXA subtype. Repeating the same analyses for the lncRNAs detected on the micro-array platform, also allowed to detect a subtype specific expression pattern, but with only a small amount of lncRNAs in the HOXA signature. Further, mRNA and lncRNA expression profiles of different T-ALL cell lines, linked to the immature, *TLX1/3* or *TAL-R* subtype were also included allowing to validate in those *in vitro* models the subtype specific lncRNA expression patterns identified from patient samples. Finally, by using the expression profiles of healthy donor T-cell subsets, I could identify specific oncogenic lncRNAs for the immature T-ALL subtype (compared to $CD34^+$ T-cell subsets) and for *TAL-R* T-ALL (compared to $CD4^+CD8^+$ T-cell subsets).

Similar as to our study, also in other leukemias lncRNA expression patterns could be identified by micro-array profiling and linked to a specific subtype. For example, Garzon et al. could couple the expression of lncRNAs to a specific mutation status of AML (acute myeloid leukemia) samples and identified a 48 lncRNA signature that was related to event free survival[16]. Also in AML, Diaz-Beya et al. revealed a subtype specific lncRNA expression profile and could link the transcription factor GATA2 to the lncRNAs of the t(8;16) subtype[17]. In B-ALL, it was shown that the subtypes defined by mRNA profiling had specific lncRNA expression profiles and these lncRNA profiles could also predict the subtype of a validation cohort. Furthermore, they identified *BALR-2* (*'B-ALL associated long RNA*-2') as an lncRNA linked to poor prognosis and reduced prednisone treatment response[18]. Quite remarkably, based on lncRNA expression profiling, it

was shown that the expression of only two lncRNAs could form a risk model in CLL[19]. Also in solid tumors lncRNA expression followed the known subtypes and the lncRNA expression profiles could be linked to prognosis (as recently reviewed by Flippot et al.[20]).

Micro-arrays were the method of choice for gene expression profiling for many years, but more recently RNA-sequencing has become the golden standard as both sequencing and data mining has become robust and reliable. Moreover, important advantages of RNA-sequencing are that low abundant transcripts can be detected more easily as well as splice variants and novel transcripts. These topics are of specific interest for lncRNA research as lncRNA are expressed at a much lower level than most protein coding genes and several novel lncRNAs are being detected as they are only expressed in a specific tissue or disease.

Taking the major advantages of RNA-sequencing for lncRNA research into account, we reprofiled 60 T-ALL samples from the previous cohort by poly(A) RNA-sequencing (see **paper 3**). At the level of protein-coding genes, it was clear that for our analysis both the data generated by micro-array and the poly(A) RNA-sequencing were equally fit for the purpose of our study. With respect to landscaping of lncRNA expression profiles, the comparison between micro-array and RNA-seq based generation of expression profiles was more challenging. However, from the matching lncRNAs I could identify several that were significantly differentially expressed in a T-ALL subtype compared to the others by both methods. Interestingly, the lncRNAs that were subtype specific in both methods, were identified with a larger fold change in the poly(A) RNA-sequencing dataset compared to the micro-array dataset. This is consistent with previous publications that also detected a larger difference with RNA-sequencing technologies. One of the major advantages of RNA-sequencing in lncRNA research is the possibility to detect novel, previously unannotated lncRNAs. In this dataset, I discovered 526 novel multi-exonic lncRNAs without protein coding potential. Some of these novel lncRNAs are also specific for a T-ALL subtype, implying that they might have a very specific role in subtype specific T-ALL oncogenesis. It should however also be noted that these novel lncRNAs are detected at very low levels, which can question the relevance in leukemogenesis. It is also not sure that these detected lncRNAs are independent transcripts; therefore several other experiments should be performed. CAGE-sequencing (Cap Analysis of Gene Expression)[21] of a T-ALL cell line could first of all identify the possible transcription start sites in the cell line and ATAC-seq (Assay for Transposase-Accessible Chromatin, detects open chromatin)[22] and histone marks (ex. H3K4me3 for active promoters, H3K36me3 for active gene bodies and H3K27ac for enhancer regions) can inform on the chromatin status of a given chromosomal locus.

Of the above-mentioned 60 patient samples, we selected 25 for which RNA was remaining to profile RNA expression with total RNA-sequencing for the discovery of novel previously unannotated lncRNAs. Indeed, as it has been estimated that about one third of all lncRNAs do not have a poly(A) tail, a significant additional number of lncRNAs should be detected using this approach. As a side note, it should be mentioned that also some known non-polyadenylated mRNAs and lncRNAs were present in the poly(A) dataset, as also random priming occurs in poly(A) RNA-seq. From our analysis, and in keeping with previous reports, it seems that the major advantage of total RNA-sequencing is indeed the detection of novel lncRNAs. We detected

17 % more novel lncRNAs than poly(A) RNA-seq and next to that, total RNA-seq also detects more annotated lncRNAs. It might also be interesting to check for the presence of circular RNAs (circRNAs) in this total RNA-seq dataset, as these are not detected by poly(A) RNA-seq. Total RNA-seq has however some disadvantages; one example is the detection of pre-mRNA (unspliced mRNA) leading to a lot of reads present from introns. This implies that more reads are necessary to detect the same amount of genes and could lead to more noise in the detection of novel lncRNAs.

With the datasets provided here, a lot of novel insights in T-ALL could be explored. First of all, the subtype specific expression of lncRNAs could be linked to protein coding genes and might reveal some novel T-ALL oncogenic networks. As an example, we are now further exploring the lncRNAs directly regulated by TLX1 and we are also elucidating possible enhancer regions and eRNAs involved in TLX1 driven T-ALL oncogenesis. Using the RNA-sequencing dataset, we can also look for (novel) ectopically expressed lncRNAs with an oncogenic role in several cases. Such lncRNAs could lead to novel therapies with less off-target effects compared to the current treatments. It should also be noted that this is the first time that such a large T-ALL patient cohort is profiled with poly(A) RNA-sequencing and this might also be used to look for specific or novel transcripts of protein coding genes and give more insights in the role of protein coding genes in T-ALL.

The hurdles to take in lncRNA research

In Chapter 1 of this thesis I already elaborated on several questions that arose during the first decade of lncRNA research, such as the lack of conservation of lncRNA loci, the binding of ribosomes to lncRNAs and the fact that some lncRNAs are mere by-products of mRNA transcription. Despite these issues, several lncRNAs are nowadays already identified to have a *bona fide* role in (cancer) cell biology, with the melanoma-specific lncRNA *SAMSSON* as an exciting example[23]. However, several studies question the importance of lncRNAs as they might be coding for micro-peptides or as they might be by-products of transcription of nearby genes. A very recent paper by the Proudfoot team studied the transcription and RNA-processing of lincRNAs (intergenic lncRNAs) and identified several differences in comparison to pre-mRNA transcription and processing. Of particular interest, they state that lincRNAs are unstable and cleaved during transcription[24]. This is one of the papers that illustrates that lncRNA research is only a very recent scientific field and a lot remains to be explored. Certainly, not all detected ncRNAs will have a function, but their detection might for example still be interesting as a biomarker. On the other hand, there is a rapidly growing list of lncRNAs with a function in the cell and/or in development (Table 3), clearly indicating the biological role of at least a significant portion of the currently annotated lncRNAs.

The study of lncRNAs poses several challenges. First, many lncRNAs are not or poorly annotated and need to be studied in detail by other technologies, for example CAGE-seq to identify the transcription start site and 'rapid amplification of cDNA ends' (RACE) to identify the exact sequence of the lncRNA. Furthermore, there is no unique lncRNA annotation method, so different

databases make use of different identifiers for the same lncRNA locus, which further complicates lncRNA research.

A crucial next step towards the identification of the function of an lncRNA is through *in vitro* knockdown or knockout of lncRNA expression. For knockdown, the use of the RNA interference machinery with siRNAs (short interfering RNAs) or shRNAs (short hairpin RNAs) may not work as they make use of the DICER machinery of the miRNA biogenesis pathway, which is mostly functional in the cytoplasm. Therefore, the use of antisense oligonucleotides (ASOs) is a valid alternative option, as these are functional in the nucleus by using the RNaseH enzyme to degrade the RNA molecule. CRISPR-mediated knockout is another option but may be hampered in subsequent phenotypic interpretation as lncRNAs can overlap with protein coding genes (ex. antisense lncRNAs) or with regulatory regions (ex. enhancers). Alternatively, lncRNAs can be overexpressed through viral transduction, but again pitfalls may occur. First of all, the knowledge of the exact transcript sequence is necessary and secondly *cis*-regulatory functions cannot be studied. For both the knockdown and the overexpression there might however be a solution using a modified CRISPR/Cas9 system. Here, a Cas9 molecule could be used that has lost its ability to cut the DNA sequence (dCas9) but that has an extra functional domain that can either activate or repress the transcription of the targeted locus[25-27].

Yet another difficulty in lncRNA research in comparison to mRNA research is that for protein coding genes the amino acid sequence of the proteins might already hint towards a function, for example a DNA-binding domain, but for lncRNAs there is very little information that can be extracted from its sequence and RNA secondary structure prediction algorithms remain predictions and do not provide extra information on possible interaction partners. To test the interaction between an lncRNA and a protein or genomic locus, several *in vitro* tools are already available (ex. *ChIRP*, *CHART* and *RIP*)[28-30], however they remain challenging to introduce in the lab. Next to the challenges for *in vitro* research, the low species conservation of lncRNAs makes it difficult to check the effects of lncRNAs in model organisms. The introduction of human lncRNA transcripts in overexpression models or the use of primary xenografts, are two possibilities to partially overcome this. Taken together, while lncRNAs hold promise for discovery of novel roles in biology and disease, many technical hurdles currently hamper the deciphering of their function and slow down the discovery process.

Completing the T-ALL subtype specific transcriptome by microRNA profiling

The role of miRNAs in T-ALL is already more broadly described than lncRNAs, with proven *in vivo* oncogenic effects (see **review** in Chapter 1). Most of the described miRNAs with a role in T-ALL oncogenesis could be linked to several known T-ALL oncogenes or tumor suppressors either as activators of miRNA transcription or as targets for the miRNAs. No studies were however present that described the subtype specific expression pattern of the miRNAs in T-ALL, but we hypothesized that there should be at least some miRNAs with a T-ALL subtype specific role, as there are several of the miRNAs linked to subtype specific T-ALL oncogenes (ex. *TLX1*, *TLX3* and *TAL1*). One example is miR-223 that is activated by the TAL1 oncogene and can downregulate the expression of the T-ALL tumor suppressor gene FBXW7 (see Chapter 1)[31, 32]. We profiled for

the first time the miRNA expression of a large cohort of primary T-ALL patient samples by means of small RNA-sequencing instead of the RT-qPCR or micro-array platforms previously used (see **paper 4**). Within our dataset I could identify miRNAs that were not detected on the RT-qPCR platforms and I could also delineate different isomiR forms of the miRNAs, which was not possible with the previous platforms.

Several of the highest expressed novel detected miRNAs in T-ALL in general are known oncomiRs in other cancer entities, for example *hsa-miR-181b-5p* that enhances AML proliferation and *hsa-miR-92b-3p* that was identified as an oncomiR in glioblastoma and that targets SMAD3, which is known to be lost in several pediatric T-ALL cases. Next to that, I detected that in each subtype several miRNAs were specifically up- or downregulated in comparison with the other subtypes. However, this was the most obvious for the immature T-ALL and the *TAL-R* subgroup. Several of these miRNAs also appear to be differentially expressed in comparison to healthy donor subsets, making them ideal candidates for further T-ALL research.

Remarkably, for several miRNAs, it was not the canonical miRNA sequence that was the highest expressed, which is something that should be taken into account for further *in vitro* and *in vivo* analysis of these miRNAs. Isomeric forms of a miRNA can occur due to the imprecise cleavage by DROSHA or DICER, by exonuclease activity at the miRNA ends or by post-transcriptional addition of one or more bases at the 3' end of the miRNA. A miRNA interacts with its target mRNA by its seed region, which starts at the second nucleotide at the 5' end and is 6-8 nucleotides in length. This means that if there is a shift in the miRNA sequence at the 5' end (addition or deletion), the seed sequence changes, which has a major impact on the target recognition of the miRNA. However, in this dataset, I mostly detected isomiR forms that have an addition or deletion of one or more nucleotides at the 3' end, which does not immediately affect the target recognition site of the miRNA. The 3' end isomiR variants are also in other studies the most common detected modifications[33-36], but their functional mechanism is less clear than for the 5' isomiRs. It was however discovered that 3' isomiRs can have a different affinity for the AGO-complexes as the canonical miRNAs, which can have an effect on the mRNA modulation (as for example only AGO2 is able to cleave the target mRNA)[33]. Furthermore, it was already described in plant biology that 3' isomiRs have a different half-life compared to the canonical miRNA[37], which is however not yet proven in animals. The presence of isomiRs also has an effect on follow-up studies, as specific RT-qPCR primers should be designed to detect their expression level and also specific ASO inhibitors are necessary to target 5' isomirs as they must bind nucleotides 2-20 of the miRNA.

FUTURE PERSPECTIVES

With this PhD thesis, I generated a panoramic landscape of noncoding RNA expression in T-ALL, thus providing a unique resource for further exploration and data mining by other researchers. The raw data for paper 1 and paper 2 are publically available through the Gene Expression Omnibus (GEO) database (GSE62006 and GSE74272) and upon publication of the two other papers, these datasets will also be oploaded to the GEO database. Furthermore, the results of the differential expression analyses are available as supplementary materials for the papers, which makes these results available for a quick examination.

This study was the first RNA-seq based characterization of a large T-ALL patient cohort for both miRNA and lncRNA. Further, the poly(A) and total RNA-sequencing data can also be used to check specific splice variants of protein coding genes or circular RNAs, which was out of the scope of this PhD thesis. Finally, an in depth study comparing poly(A) and total RNA-sequencing in our T-ALL patient cohort will be performed as a follow-up on this project.

In addition to the above, a further challenge is the integration of the currently available mRNA, lncRNA and miRNA datasets in order to gain further insights into the biology of the disease and the interconnection of regulatory networks and integrated perturbed biological processes. A first example could be the negative correlation between a miRNA and an mRNA/lncRNA in the subtype specific profiles, which could identify potential targets of the miRNA (Figure 12 A). MicroRNA target prediction algorithms could aid in this analysis, however there are some disadvantages as different tools give different results, several special miRNA-mRNA target interactions are described (for example G-bulges) and isomiR forms can have different target preferences. AGO2-pull down or reporter assays could on the other hand be used for *in vitro* identification of miRNA-mRNA/lncRNA interaction. A second interesting focus could be the identification of transcription factor binding sites near the promoters of lncRNAs and miRNAs that could explain a possible activation or repression of the subtype specific lncRNA/miRNA by mRNAs linked to the same subtype (Figure 12 B). A third and final example I would like to point out here is the identification of competing endogenous RNAs (ceRNAs), which are lncRNAs that bind miRNAs, acting as sponges and titrating them away from their mRNA target, leading to the increased expression of the mRNA, as has been described for the PTENP1 lncRNA (Figure 12 C)[38]. This could be identified by mRNAs and lncRNAs that are upregulated in the same subtype and that can be targeted by the same miRNAs. An extra clue would be the presence of multiple binding sites for the same miRNA in the lncRNA.

Further insights into T-ALL biology might lead to novel therapeutic targets as even lncRNAs and miRNAs are nowadays already druggable with some clinical trials using antisense oligonucleotide ongoing (see Chapter 1). Despite the fact that there is no specific link between T-ALL subtypes and prognosis (as discussed in Chapter 1), these subtype specific pathways might still be of interest in clinic. For example, the relative highest amount of NOTCH1-oncogenic mutations (druggable with GSIs) is present in *TLX1* T-ALL[1] and oncogenic IL7R pathway mutations that are associated with a poor outcome are mostly detected in immature and *TLX3* T-ALL[39]. Furthermore, the comparison with healthy donor thymocyte subsets remains important as these might

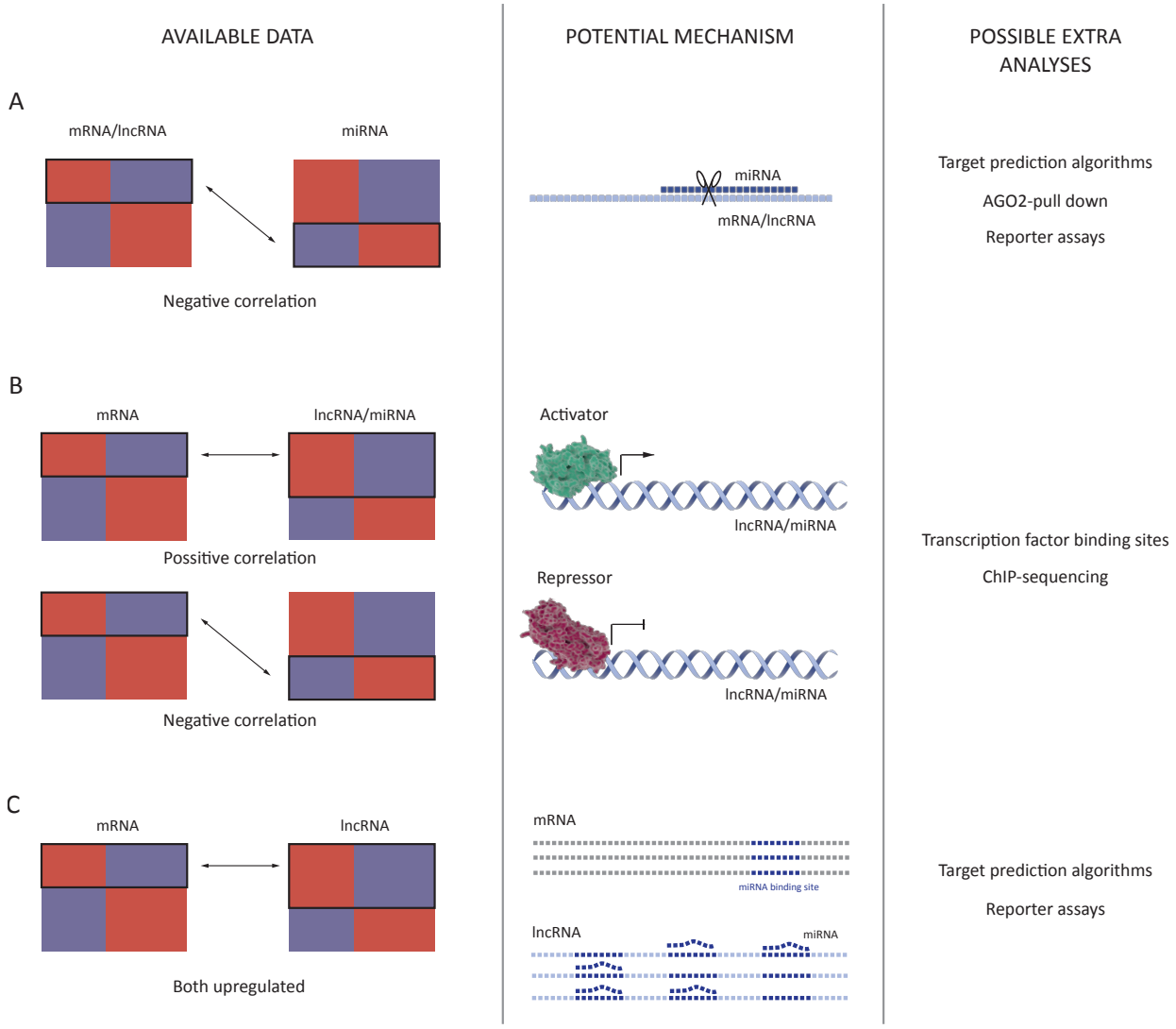


Figure 12. Future research directions using the datasets provided in this thesis. (A) Negative correlation with a miRNA might point at a miRNA target. (B) Positive or negative correlation with a transcription factor might indicate transcriptional regulation. (C) An mRNA and lncRNA that are both upregulated in a specific group and that have binding sites for the same miRNA might identify the lncRNA as a ceRNA. (Illustrations from www.somersault1824.com)

help in identifying true oncogenic pathways. Next to that, online available gene-expression datasets might aid in the identification of cancer type or tissue type specific lncRNAs. For example, 'The Cancer Genome Atlas' (TCGA) contains RNA-seq data for 33 different cancer types which could be queried for a lncRNA of interest. A T-ALL lncRNA that is not expressed in any other cancer type might indicate a specific T-cell leukemogenic function.

Nowadays, the major challenges for T-ALL in the clinic are the patients that show initial therapy resistance or the patients that relapse, as their prognosis is very dismal. It is therefore very important to quickly identify these patients and find novel druggable targets and combination therapies might aid in improving their survival.

References

1. Weng AP, Ferrando AA, Lee W, Morris JPt, Silverman LB, Sanchez-Irizarry C, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science. 2004;306(5694):269-71.
2. Taghon T, Waegemans E, Van de Walle I. Notch signaling during human T cell development. Curr Top Microbiol Immunol. 2012;360:75-97.
3. Pear WS, Aster JC, Scott ML, Hasserjian RP, Soffer B, Sklar J, et al. Exclusive development of T cell neoplasms in mice transplanted with bone marrow expressing activated Notch alleles. J Exp Med. 1996;183(5):2283-91.
4. Palomero T, Barnes KC, Real PJ, Glade Bender JL, Sulis ML, Murty VV, et al. CUTLL1, a novel human T-cell lymphoma cell line with t(7;9) rearrangement, aberrant NOTCH1 activation and high sensitivity to gamma-secretase inhibitors. Leukemia. 2006;20(7):1279-87.
5. Deangelo DJ, Stone RM, Silverman LB, Stock W, Attar EC, Fearen I, et al. A phase I clinical trial of the notch inhibitor MK-0752 in patients with T-cell acute lymphoblastic leukemia/lymphoma (T-ALL) and other leukemias. J Clin Oncol. 2006;24(18_suppl):6585.
6. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature. 2016;539(7629):452-5.
7. Paralkar VR, Taborda CC, Huang P, Yao Y, Kossenkov AV, Prasad R, et al. Unlinking an lncRNA from Its Associated cis Element. Mol Cell. 2016;62(1):104-10.
8. Trimarchi T, Bilal E, Ntziachristos P, Fabbri G, Dalla-Favera R, Tsirigos A, et al. Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. Cell. 2014;158(3):593-606.
9. Medyouf H, Gusscott S, Wang H, Tseng JC, Wai C, Nemirovsky O, et al. High-level IGF1R expression is required for leukemia-initiating cell activity in T-ALL and is supported by Notch signaling. J Exp Med. 2011;208(9):1809-22.
10. Wang Y, Wu P, Lin R, Rong L, Xue Y, Fang Y. LncRNA NALT interaction with NOTCH1 promoted cell proliferation in pediatric T cell acute lymphoblastic leukemia. Sci Rep. 2015;5:13749.
11. Gu Y, Masiero M, Banham AH. Notch signaling: its roles and therapeutic potential in hematological malignancies. Oncotarget. 2016;7(20):29804-23.
12. Rizzo P, Osipo C, Foreman K, Golde T, Osborne B, Miele L. Rational targeting of Notch signaling in cancer. Oncogene. 2008;27(38):5124-31.
13. Katsushima K, Natsume A, Ohka F, Shinjo K, Hatanaka A, Ichimura N, et al. Targeting the Notch-regulated non-coding RNA TUG1 for glioma treatment. Nat Commun. 2016;7:13616.
14. Soulier J, Clappier E, Cayuela JM, Regnault A, Garcia-Peydro M, Dombret H, et al. HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). Blood. 2005;106(1):274-86.
15. Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC, et al. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer Cell. 2002;1(1):75-87.
16. Garzon R, Volinia S, Papaioannou D, Nicolet D, Kohlschmidt J, Yan PS, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. Proc Natl Acad Sci U S A. 2014;111(52):18679-84.
17. Diaz-Beya M, Navarro A, Cordeiro A, Pratcorona M, Castellano J, Torrente MA, et al. Exploring the Expression Profile of Long Non-Coding RNA (lncRNA) in Different Acute Myeloid Leukemia (AML) Subtypes: t(8;16)(p11;p13)/MYST3-Crebbp AML Harbors a Distinctive Lncrna Signature. Blood. 2015;126(23):1397-.
18. Fernando TR, Rodriguez-Malave NI, Waters EV, Yan W, Casero D, Basso G, et al. LncRNA Expression Discriminates Karyotype and Predicts Survival in B-Lymphoblastic Leukemia. Mol Cancer Res. 2015;13(5):839-51.
19. Ronchetti D, Manzoni M, Agnelli L, Vinci C, Fabris S, Cutrona G, et al. lncRNA profiling in early-stage chronic lymphocytic leukemia identifies transcriptional fingerprints with relevance in clinical outcome. Blood Cancer J. 2016;6(9):e468.
20. Flippot R, Malouf GG, Su X, Mouawad R, Spano JP, Khayat D. Cancer subtypes classification using long non-coding RNA. Oncotarget. 2016;7(33):54082-93.
21. Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A. 2003;100(26):15776-81.
22. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10(12):1213-8.
23. Leucci E, Vendramin R, Spinazzi M, Laurette P, Fiers M, Wouters J, et al. Melanoma addiction to the long non-coding RNA SAMMSON. Nature. 2016;531(7595):518-22.

24. Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ. Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. Mol Cell. 2016.
25. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013;152(5):1173-83.
26. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods. 2013;10(10):973-6.
27. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 2013;10(10):977-9.
28. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell. 2011;44(4):667-78.
29. Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, et al. The genomic binding sites of a noncoding RNA. Proc Natl Acad Sci U S A. 2011;108(51):20497-502.
30. Gilbert C, Svejstrup JQ. RNA immunoprecipitation for determining RNA-protein associations in vivo. Curr Protoc Mol Biol. 2006;Chapter 27:Unit 27 4.
31. Mansour MR, Sanda T, Lawton LN, Li X, Kreslavsky T, Novina CD, et al. The TAL1 complex targets the FBXW7 tumor suppressor by activating miR-223 in human T cell acute lymphoblastic leukemia. J Exp Med. 2013;210(8):1545-57.
32. Correia NC, Durinck K, Leite AP, Ongenaert M, Rondou P, Speleman F, et al. Novel TAL1 targets beyond protein-coding genes: identification of TAL1-regulated microRNAs in T-cell acute lymphoblastic leukemia. Leukemia. 2013;27(7):1603-6.
33. Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Nishibu T, Ukekawa R, et al. A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. Genome Res. 2010;20(10):1398-410.
34. Wyman SK, Knouf EC, Parkin RK, Fritz BR, Lin DW, Dennis LM, et al. Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. Genome Res. 2011;21(9):1450-61.
35. Lee LW, Zhang S, Etheridge A, Ma L, Martin D, Galas D, et al. Complexity of the microRNA repertoire revealed by next-generation sequencing. RNA. 2010;16(11):2170-80.
36. Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, et al. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biol. 2011;12(12):R126.
37. Lu S, Sun YH, Chiang VL. Adenylation of plant miRNAs. Nucleic Acids Res. 2009;37(6):1878-85.
38. Yu G, Yao W, Gumireddy K, Li A, Wang J, Xiao W, et al. Pseudogene PTENP1 functions as a competing endogenous RNA to suppress clear-cell renal cell carcinoma progression. Mol Cancer Ther. 2014;13(12):3086-97.
39. Li Y, Buijs-Gladdines JG, Cante-Barrett K, Stubbs AP, Vroegindeweij EM, Smits WK, et al. IL-7 Receptor Mutations and Steroid Resistance in Pediatric T cell Acute Lymphoblastic Leukemia: A Genome Sequencing Study. PLoS Med. 2016;13(12):e1002200.

SUMMARY

Normal T-cell development is a hierarchically orchestrated process and requires strict regulation of underlying gene expression programs. Various genomic lesions can lead to either aberrant expression/activation of proto-oncogenes or repression/loss of tumor suppressors that induces a differentiation arrest at a specific developmental T-cell stage, thereby increasing the risk for transformation of arrested progenitor thymocytes to T-cell acute lymphoblastic leukemia (T-ALL).

The prognosis of T-ALL patients has increased in the last decades due to intensified therapies, but for patients with relapse and therapy resistance outcome with current treatment regimens is very poor. These hurdles, as well as the problem of long-term side effects of current harsh treatment protocols, underscore the importance of further research to discover novel, more efficient and less toxic targeted therapies. During my PhD mandate, I explored the expression patterns of two types of noncoding RNAs, lncRNAs and miRNAs, in T-ALL patient samples, sorted T-cell subsets and T-ALL *in vitro* models as their functional role in cancer biology has emerged strongly in the last decades and, particularly for lncRNAs, their role in T-ALL remained largely unexplored. This thesis provides a unique resource depicting the miRNA and lncRNA landscape of a well-annotated series of primary T-ALL samples and provides the first steps in exploring the putative role of selected candidate novel lncRNAs and miRNAs implicated in T-ALL with potential as biomarkers or as novel therapeutic targets.

Given the key role of the NOTCH1-signaling pathway in both normal and malignant T-cell development, this cascade has been exploited in the treatment of T-ALL patients through the use of the NOTCH1 inhibitor GSI. Although the use of GSIs in *in vitro* studies seemed very promising, clinical trials unfortunately failed as a consequence of therapy resistance and gastrointestinal toxicity issues. Hence, other components of the pathway should be more properly dissected and characterized to pinpoint novel targets for innovative treatment regimens. During the first part of my thesis, I therefore identified lncRNAs directly regulated by NOTCH1 in T-ALL and in normal T-cell development.

Interestingly, T-ALL cases can be classified in different genetic subtypes based on a specific oncogenic driver event and/or stage of differentiation arrest during T-cell development. Furthermore, the mRNA expression profiles of T-ALL cases cluster together according to subtype. I aimed in the second part of my PhD thesis to explore the expression of lncRNAs next to the protein coding gene expression for each of the known T-ALL subtypes by means of micro-array technology and I could show that the T-ALL subgroups are also characterized by a specific lncRNA expression profile, adding a novel layer of gene signatures that characterize these groups. Through a comparative analysis with lncRNA expression profiles of sorted healthy donor T-cell subsets, I could further delineate the subtype specific and in T-ALL ectopically expressed lncRNAs, which are of particular interest for in depth research on oncogenic lncRNAs. Next-generation sequencing techniques have dramatically boosted our ability to explore transcriptome profiling in more depth. In this respect, RNA-sequencing has now become the method of choice for gene expression profiling. Therefore, I reprofiled a large set of T-ALL patient

samples, that we previously screened by means of micro-array technology, using poly(A) and total RNA-sequencing. Comparative analysis with the array based technology showed that RNA-sequencing allows detection of differentially expressed lncRNAs with a larger fold change thus offering a broader detection range than micro-arrays. Another advantage of the RNA-sequencing technologies, specifically in the context of studying lncRNA expression, is the possibility to detect novel, previously unannotated lncRNAs, which is of particular interest as these might be lineage specific lncRNAs that could be used for targeted therapy.

In the last part of this thesis, I focused on the subtype specific expression pattern of miRNAs in T-ALL. Different research teams, including the host lab, already studied the role of individual miRNAs and mRNA-miRNA networks in T-ALL. In this thesis, a pioneering study was set up to evaluate subtype specific miRNA expression patterns of T-ALL patient samples by means of small RNA-sequencing. In contrast to previously developed RT-qPCR based and micro-array platforms, we were able to scrutinize a much wider spectrum of miRNAs and were also able for the first time to detect and evaluate isomiR expression in T-ALL. This dataset allowed me to define a subtype specific miRNA profile for T-ALL cases and to identify novel potential oncogenic miRNAs by comparing these profiles with miRNA expression profiles from healthy donor T-cell subsets.

To conclude, with this PhD project I established unique noncoding RNA expression datasets (miRNA and lncRNA) and identified potential oncogenic miRNAs and lncRNAs in T-ALL that could be used for further in depth exploration on both a functional and therapeutic level. As several therapies for noncoding RNAs (miRNAs and lncRNAs) are already in (pre-) clinical trials, the identification of a miRNA or lncRNA with a specific oncogenic role in T-ALL, NOTCH1-driven T-ALL or a specific T-ALL subtype might lead to the establishment of a novel targeted therapy in T-ALL.

SAMENVATTING

De ontwikkeling van T-cellen in de thymus is een hiërarchisch proces dat een strikte regulatie van gen expressie programma's eist. Verscheidene genetische defecten kunnen leiden tot enerzijds een overexpressie/activatie van proto-oncogenen of anderzijds een onderdrukking/verlies van tumor suppressor genen die een differentiatie stop in de normale T-cel ontwikkeling kunnen introduceren, waardoor het risico tot de ontwikkeling van T-cel acute lymfatische leukemie (T-ALL) verhoogt. Door een intensieve behandeling is de prognose voor T-ALL patiënten in de laatste decennia sterk gestegen, maar de patiënten die hervallen of resistentie tegen de behandeling vertonen hebben nog steeds een slechte overlevingskans. Dit, maar ook de nevenwerkingen op lange termijn van de intensieve behandelingsschema's tonen het belang aan van verder onderzoek naar nieuwe, meer efficiënte en minder toxische doelgerichte therapieën.

Tijdens mijn doctoraatsproject onderzocht ik het expressie patroon van twee types niet-coderende RNAs, miRNAs en lncRNAs, in T-ALL patiënten stalen, verscheidene T-cel ontwikkelingsstadia en T-ALL *in vitro* model systemen, aangezien hun rol in kankerbiologie in het laatste decennium aangetoond is en, vooral voor lncRNAs, hun rol in T-ALL nog zeer weinig beschreven is. Deze thesis brengt een unieke dataset aan die het miRNA en lncRNA landschap van primaire T-ALL patiëntenstalen in kaart brengt en een eerste stap aanbiedt richting het onderzoeken van een functionele rol voor kandidaat miRNAs en lncRNAs betrokken in T-ALL als mogelijke biomerkers of nieuwe therapeutische doelwitten.

Door de sleutelrol van NOTCH1-signalisatie in zowel normale als maligne T-cel ontwikkeling, werd deze signalisatie gebruikt in het onderzoek naar behandelingen voor T-ALL patiënten door middel van de NOTCH1-inhibitor GSI. De *in vitro* testen met GSIs waren veelbelovend, maar de klinische studies moesten helaas vroegtijdig stopgezet worden wegens resistentie en toxische reacties in het spijsverteringsstelsel. Hierdoor zouden andere componenten in de NOTCH1-signalisatie grondiger onderzocht en gekarakteriseerd moeten worden om nieuwe doelwitten voor innovatieve therapieën te kunnen aanduiden. Daarom heb ik tijdens het eerste deel van mijn doctoraatsstudie lncRNAs geïdentificeerd die direct gereguleerd worden door NOTCH1 in T-ALL en normale T-cel ontwikkeling.

T-ALL patiënten kunnen ook ingedeeld worden in verschillende genetische subtypes gebaseerd op een specifiek oncogeen defect en/of het stadium van differentiatie arrest tijdens T-cel ontwikkeling. Daarnaast is ook aangetoond dat de mRNA expressieprofielen van T-ALL patiënten samen clusteren naargelang het subtype. In het tweede deel van mijn doctoraatsstudie heb ik de lncRNA expressie in deze T-ALL subtypes in kaart gebracht door middel van micro-array technologie en kon ik aantonen dat de T-ALL subtypes ook gekarakteriseerd worden door specifieke lncRNA expressieprofielen, waardoor deze subtypes nu door een extra niveau van genexpressie gekarakteriseerd zijn. Door het vergelijken van deze data met de lncRNA expressieprofielen in gesorteerde T-cel stadia van gezonde donoren, kon ik de subtype specifieke en ectopisch geëxpresseerde lncRNA's identificeren, wat interessant is voor verder onderzoek naar oncogene lncRNA's in T-ALL. Door de nieuwe generatie van sequeneringstechnologiën is de mogelijkheid tot het grondig onderzoeken van transcriptoom profielen gegroeid, waardoor RNA-

sequenering nu de voorkeur krijgt op micro-arrays voor genexpressieprofilering. Hiervoor heb ik een groot deel van de T-ALL patiënten stalen uit vorig micro-array onderzoek opnieuw geprofileerd door middel van poly(A) en totale RNA-sequenering. Een vergelijkende analyse met de array technologie toonde aan dat RNA-sequenering de differentieel geëxpresseerde lncRNAs met een groter verschil in expressie detecteerde en dus een groter detectie gebied heeft dan de micro-arrays. Een extra voordeel van RNA-sequenering, vooral in de context van lncRNA expressie, is de mogelijkheid om nieuwe lncRNAs te identificeren die nog niet geannoteerd zijn in databanken, wat zeer interessant is voor de detectie van T-ALL specifieke lncRNAs die kunnen gebruikt worden voor doelgerichte therapieën.

In het laatste deel van deze thesis, heb ik mijn focus gelegd op het subtype specifieke expressiepatroon van miRNA's in T-ALL. Verschillende onderzoeksgroepen, waaronder het labo waar ik mijn doctoraatsstudie heb uitgevoerd, hebben al de rol van individuele miRNAs en mRNA-miRNA netwerken in T-ALL bestudeerd. In deze thesis werd een eerste studie opgezet om de subtype specifieke miRNA expressiepatronen in primaire T-ALL patiëntenstalen te evalueren door middel van een RNA-sequeneringstechnologie gericht op de detectie van korte RNA moleculen. In tegenstelling tot vroeger ontwikkelde RT-qPCR en micro-array platformen, waren we nu in staat om een veel breder spectrum van miRNAs te detecteren en konden we ook, voor de eerste keer, isomiR expressie detecteren en evalueren. Deze dataset liet me daarnaast ook toe om subtype-specifieke miRNA profielen voor T-ALL patiënten te definiëren en, door een vergelijking met miRNA profielen van gezonde T-cel donoren, mogelijke nieuwe oncogene miRNAs in T-ALL te identificeren.

Samengevat heb ik met dit doctoraatsproject unieke niet-coderende RNA expressie datasets (miRNAs en lncRNAs) gegenereerd en mogelijke oncogene miRNAs en lncRNAs in T-ALL geïdentificeerd die gebruikt kunnen worden voor verder functioneel en therapeutisch onderzoek. Aangezien er al voor verscheidene therapieën gericht tegen niet-coderende RNAs (miRNAs en lncRNAs) (pre-) klinische studies opgestart zijn, kan de identificatie van miRNAs en lncRNAs met een specifieke oncogene rol in T-ALL, NOTCH1-gedreven T-ALL of in een T-ALL subtype leiden tot de ontwikkeling van nieuwe doelgerichte therapieën in T-ALL.

DANKWOORD

Het schrijven van mijn dankwoord brengt een dubbel gevoel bij mij teweeg. Enerzijds komt er een lichte euforie naar boven, aangezien dit betekent dat ik er na iets meer dan vier jaar in geslaagd ben om een volwaardige doctoraatsthesis te schrijven. Anderzijds betekent dit ook het einde van een zeer leerrijke periode in mijn leven die ik samen met heel wat leuke collega's mocht beleven.

Mijn doctoraat zou nooit tot stand zijn gekomen zonder mijn promotor Prof. dr. Frank Speleman. Hij zorgde er in eerste instantie voor dat ik al tijdens het laatste jaar van mijn studies een FWO mandaat met succes kon binnenrijven. Zijn enthousiasme voor het kankeronderzoek werkte zeer aanstekelijk en de manier waarop hij snel een nieuw en goed project kon bedenken was zeer inspirerend. Verder zal ik ook blijven onthouden dat, alhoewel hij vaak van de ene naar de andere vergadering spurtte en zijn stoel in onze bureau zo goed als ongebruikt is, hij toch een zeer bereikbare prof was. Ook mijn co-promotoren waren van groot belang voor het slagen van dit project. In eerste instantie zorgde dr. Pieter Rondou ervoor dat ik door mijn masterthesis de microbe van het wetenschappelijk onderzoek te pakken kreeg, zorgde hij voor een goede opstart van mijn project en was hij ook altijd bereikbaar voor een goed gesprek. Vervolgens pikte ook Prof. dr. Pieter Van Vlierberghe in en was hij als 'T-ALL encyclopedie' een zeer grote meerwaarde voor mijn doctoraat. Bedankt Frank, Pieter R en Pieter VV om mijn project te begeleiden en te sturen en ervoor te zorgen dat mijn doctoraat een heel leerrijke periode was.

Verder wil ik enkele collega's specifiek bedanken. Eerst en vooral het steeds kleiner wordende FSP T-ALL team: Kaat, Karen en Siebe. Ik vind het heel spijtig dat ik nu ook dit team verlaat, maar ik ben er zeker van dat jullie de T-ALL projecten tot goed eind zullen brengen (een Nature paper bijvoorbeeld?). Ook wil ik al mijn bureaugenootjes bedanken: Kaat, Karen, Aline, Jolien, Pieter, Joni, Sofie, Béa, Filip, Gloria, Els, Suzanne, Evelien en Nadine. Bedankt voor de vele wetenschappelijke reflecties bij 'acute' problemen of toch alvast de mogelijkheid om even frustraties te uiten, vooraleer er weer met volle moed in te vliegen. Maar ook zeker en vast bedankt voor de grappige momenten en de ontspannende gesprekken tussendoor. Daarnaast wil ik ook Wouter Van Loocke bedanken voor zijn onmisbare hulp bij de bio-informatica analyses, team Jo voor de interessante discussies omtrent lncRNA problemen en Tom Taghon voor de samenwerking omtrent normale T-cel ontwikkeling; zonder jullie input was mijn werk helemaal niet zoals het nu is. Ook alle CMGG collega's wil ik bedanken voor de ontspannende lunchgesprekken, de leuke congres uitstappen en de georganiseerde activiteiten (ik denk dan bv. aan C4CC, de monthly drinks, de lab retreats etc.). Door jullie kan ik terugkijken op een zeer toffe periode en ik zal nog vaak aan deze momenten moeten denken tijdens de nieuwe uitdagingen die ik nu tegemoet ga. Een speciaal woordje van dank wil ik ook richten aan mijn paranimfen Karen, Jolien en Celine, die me in de laatste eindspurt van mijn doctoraat hebben bijgestaan om heel wat praktische zaken te regelen.

Tot slot wil ik ook een woordje van dank richten aan mijn familie en vrienden. Bedankt om interesse te tonen in iets wat voor velen onder jullie redelijk abstract is en om voor de nodige ont-

spanning te zorgen. De steeds terugkomende vragen “Wat is het nu weer juist dat je onderzoekt?” en “Is er al een datum die we moeten vastzetten voor je doctoraatsverdediging?” zijn nu hopelijk beantwoord. In het bijzonder wil ik ook mijn ouders, mijn broer en Laurens bedanken voor hun onvoorwaardelijke steun, niet alleen tijdens mijn doctoraat maar ook in de vele jaren ervoor.

CURRICULUM VITAE

Annelynn Wallaert

Personalia

Address: Joanna van de Cauterlaan 3
9820 Merelbeke
Belgium
Phone: +32 472 69 19 35
E-mail: annelynn.wallaert@ugent.be
annelynn.wallaert@gmail.com

Date of birth: January 16, 1990
Nationality: Belgian

Experience

Molecular Biologist
Period: March 2017 – present
Institute: az groeninge, Kortrijk, Belgium

Doctoral Research Assistant (4 year FWO scholarship)
Period: October 2012 – February 2017
Institute: Center for Medical Genetics, Ghent University, Ghent, Belgium
Thesis: The role of long noncoding RNAs in T-cell acute lymphoblastic leukemia
Promotor: Prof. dr. Frank Speleman
Co-promotors: Prof. dr. Pieter Van Vlierberghe and dr. Pieter Rondou

Education

Master of Science in Biochemistry and Biotechnology (Biomedical Biotechnology)
Period: 2007 – 2012
Institute: Ghent University, Ghent, Belgium
Thesis: Functional evaluation of candidate miRNAs in T-ALL
Promotors: Prof. dr. Bruce Poppe and dr. Pieter Rondou

Doctoral training program

- Introduction to Next-Generation Sequencing, EMBL/EBI course, April 13-19 2015, Hinxton, United Kingdom
- N2N Multidisciplinary Seminar Series on Bioinformatics 2015-2015 – Doctoral Schools of Life Sciences and Medicine, Ghent University
- HPC Unix command line, shell scripting and HPC basics 2015 – Doctoral Schools of Life Sciences and Medicine, Ghent University
- Personal effectiveness - Doctoral Schools of Life Sciences and Medicine, Ghent University
- Bioinformatics: Introduction and Methods, Peking University, Coursera online course
- Advanced Academic English: Conference Skills - Doctoral Schools of Life Sciences and Medicine, Ghent University
- Certificate FELASA category C, Laboratory Animal Science, Ghent University

Scientific achievements

Publications

- **The Notch driven long non-coding RNA repertoire in T-cell acute lymphoblastic leukemia.** Durinck K*, Wallaert A*, Van de Walle I, Van Loocke W, Volders PJ, Vanhauwaert S, Geerdens E, Benoit Y, Van Roy N, Poppe B, Soulier J, Cools J, Mestdagh P, Vandesomepele J, Rondou P, Van Vlierberghe P, Taghon T* and Speleman F*. Haematologica, Dec 2014. (IF 2014: 5.814, SCI Hematology: 4/70) * = Shared first/last authors

- **Novel biological insights in T-cell acute lymphoblastic leukemia.** Durinck K, Goossens S, Peirs S, Wallaert A, Van Loocke W, Matthijssens F, Pieters T, Milani G, Lammens T, Rondou P, Van Roy N, De Moerloose B, Benoit Y, Haigh J, Speleman F, Poppe B and Van Vlierberghe P. Exp Hematol, Jun 2015. (IF 2015: 2.303, SCI hematology: 46/70)

- **Characterization of genome-wide TLX1 binding profile in T-cell acute lymphoblastic leukemia.** Durinck K, Van Loocke W, Van der Meulen J, Van de Walle I, Ongenaert M, Rondou P, Wallaert A, de bock CE, Van Roy N, Poppe B, Cools J, Soulier J, Taghon T, Speleman F and Van Vlierberghe P. Leukemia, Jun 2015. (IF 2015: 12.104, SCI hematology: 1/70)

- **Long noncoding RNA signatures define oncogenic subtypes in T-cell acute lymphoblastic leukemia.** Wallaert A, Durinck K, Van Loocke W, Van de Walle I, Matthijssens F, Volders PJ, Avila Cobos F, Rombaut D, Rondou P, Mestdagh P, Vandesompele J, Poppe B, Taghon T, Soulier J, Van Vlierberghe P and Speleman F. Leukemia, 2016 Sep;30(9):1927-30. (IF 2015: 12.105, SCI hematology: 1/70)

- **Unique long non-coding RNA expression signature in ETV6/RUNX1-driven B-cell precursor acute lymphoblastic leukemia.** Ghazavi F, De Moerloose B, Van Loocke W, Wallaert A, Helsmoortel HH, Ferster A, Bakkus M, Plat G, Delabesse E, Uyttebroeck A, Van Nieuwerburgh F, Deforce D, Van Roy N, Speleman F, Benoit Y, Lammens T* and Van Vlierberghe P*. Oncotarget, Sep 2016, 2016 Nov 8;7(45):73769-73780. (IF 2014/2015: 6.359, SCI oncology: 35/322) *= Shared first/last authors

- **EV-TRACK: transparent reporting and centralizing knowledge in extracellular vesicle research.** EV-TRACK Consortium. Van Deun J, Mestdagh P, Agostinis P, Akay Ö, Anand S, Anckaert J, Martinez ZA, Baetens T, Beghein E, Bertier L, Berx G, Boere J, Boukouris S, Bremer M, Buschmann D, Byrd JB, Casert C, Cheng L, Cmoch A, Daveloose D, De Smedt E, Demirsoy S, Depoorter V, Dhondt B, Driedonks TA, Dudek A, Elsharawy A, Floris I, Foers AD, Gärtner K, Garg AD, Geeurickx E, Gettemans J, Ghazavi F, Giebel B, Kormelink TG, Hancock G, Helsmoortel H, Hill AF, Hyenne V, Kalra H, Kim D, Kowal J, Kraemer S, Leidinger P, Leonelli C, Liang Y, Lippens L, Liu S, Lo Cicero A, Martin S, Mathivanan S, Mathiyalagan P, Matusek T, Milani G, Monguió-Tortajada M, Mus LM, Muth DC, Németh A, Nolte-'t Hoen EN, O'Driscoll L, Palmulli R, Pfaffl MW, Primdal-Bengtson B, Romano E, Rousseau Q, Sahoo S, Sampaio N, Samuel M, Scicluna B, Soen B, Steels A, Swinnen JV, Takatalo M, Thaminy S, Théry C, Tulkens J, Van Audenhove I, van der Grein S, Van Goethem A, van Herwijnen MJ, Van Niel G, Van Roy N, Van Vliet AR, Vandamme N, Vanhauwaert S, Vergauwen G, Verweij F, Wallaert A, Wauben M, Witwer KW, Zonneveld MI, De Wever O, Vandesompele J and

Hendrix A. Nat Methods. 2017 Feb 28;14(3):228-232. (IF 2015: 25.328, SCI biochemical research methods: 1/77)

- **T-ALL and thymocytes: a message of noncoding RNAs.** Wallaert A, Durinck K, Taghon T, Van Vlierberghe P and Speleman F. Journal of Hematology and Oncology. 2017 10:66. (IF 2015: 6.263, SCI hematology: 6/70)

- **Long non-coding RNAs in leukemia: biology and clinical impact.** Lammens T, Durinck K, Wallaert A, Speleman F and Van Vlierberghe P. Current Opinion in Hematology. In press. (IF 2015: 3.331, SCI Hematology: 25/70)

- **Comprehensive miRNA expression profiling in human T-cell acute lymphoblastic leukemia by small RNA-sequencing.** Wallaert A, Hernandez L, Taghon T, Speleman F and Van Vlierberghe P. Submitted to Scientific Reports.

- **RNA-sequencing profiling across T-ALL and thymocyte subsets identifies candidate oncogenic lncRNAs in T-ALL.** Wallaert A, Durinck K, Van Loocke W, Hernandez L, De Decker M, Taghon T, Soulier J, Van Vlierberghe P and Speleman F. In preparation.

Awards

- Ablynx prize for best Masterthesis in Biochemistry and Biotechnology

Presentations

- **The T-ALL lncRNA transcriptome: the long and winding road for T-ALL drivers.**
  Course and Master Classes on Molecular Aspects of Hematological Disorder, Rotterdam, The Netherlands (10-12/6/2014)

- **The Notch driven long noncoding RNA repertoire in T-cell acute lymphoblastic leukemia.**
  Durinck K*, Wallaert A*, Van de Walle I, Van Loocke W, Volders PJ, Vanhauwaert S, Geerdens E, Benoit Y, Van Roy N, Poppe B, Soulier J, Cools J, Mestdagh P, Vandesomepele J, Rondou P, Van Vlierberghe P, Taghon T*, Speleman F*. * = Shared first/last authors
  Research Day, Faculty of Medicine and Health Sciences, Ghent, Belgium (5/3/2015)

- **Deciphering the NOTCH1-TLX1 regulome in T-cell acute leukemia.**
  IAP Progress Meeting 2015, Ghent, Belgium (08/10/2015)

- **The lncRNAomes of T-ALL and normal immature thymocyte subsets through poly-A and total RNA-seq**. Wallaert A, Durinck K, Van Loocke W, Van de Walle I, Rondou P, Soulier J, Van Vlierberghe P, Taghon T and Speleman F.
  First joint meeting BeSHG/NVHG, Genetics and Society, Leuven, Belgium (04-05/02/2016)

- **The lncRNAomes of T-cell acute lymphoblastic leukemia and normal immature thymocyte subsets through combined poly-A and total RNA-sequencing.** Wallaert A, Durinck K, Van Loocke W, Van de Walle I, Rondou P, Soulier J, Van Vlierberghe P, Taghon T and Speleman F.
  Research Day, Faculty of Medicine and Health Sciences, Ghent, Belgium (16/03/2016)

- **The lncRNAomes of T-cell acute lymphoblastic leukemia and normal immature thymocyte subsets through combined poly-A and total RNA-sequencing.** Wallaert A, Durinck K, Van Loocke W, Van de Walle I, Rondou P, Soulier J, Van Vlierberghe P, Taghon T and Speleman F.
  Light at the dark side of the genome, fTales, Ghent, Belgium (15-16/09/2016)

Posters

- **Unique lncRNA signatures mark the major T-ALL genetic subgroups**. Wallaert A, Durinck K, Van Loocke W, Volders PJ, Benoit Y, Poppe B, Mestdagh P, Vandesompele J, Soulier J, Cools J, Van Vlierberghe P, Rondou P, Speleman F.
  Translational Genetics: from cage over bench to bed, BeSHG, Antwerp, Belgium (7/2/2014)

- **The first lncRNA landscape of major genetic T-ALL subsets and guilt-by-association analysis for ETP-ALL specific lncRNAs.** Wallaert A, Durinck K, Van Loocke W, Volders PJ, Van de Walle I, Benoit Y, Poppe B, Mestdagh P, Vandesompele J, Taghon T, Soulier J, Cools J, Rondou P, Van Vlierberghe P, Speleman F.
  19th EHA congress, Milan, Italy (12-15/6/2014)
  Travel grant winner

- **Integrative analysis of the NOTCH1 regulatory network identifies key downstream lncRNAs in acute T-cell lymphoblastic leukemia and normal T-cell development.** Wallaert A*, Durinck K*, Rondou P, Van de Walle I, Van Loocke W, Volders PJ, Ongenaert M, Van Roy N, Poppe B, Benoit Y, Vandesompele J, Taghon T, Van Vlierberghe P, Speleman F.
  Non-Coding RNA – From Basic Mechanisms to Cancer, Heidelberg, Germany (22-25/6/2014)

- **LncRNA profiling in T-cell acute lymphoblastic leukemia predicts a functional connection between lncRNAs and the major T-ALL genetic subgroups.** Wallaert A, Durinck K, Van Loocke W, Volders PJ, Van de Walle I, Benoit Y, Poppe B, Mestdagh P, Vandesompele J, Taghon T, Soulier J, Cools J, Rondou P, Van Vlierberghe P, Speleman F.
  Non-Coding RNA – From Basic Mechanisms to Cancer, Heidelberg, Germany (22-25/6/2014)

- **The immature phenotype T-ALL long noncoding RNAome as a prelude to novel therapeutic targeting.** Wallaert A, Durinck K, Van Loocke W, Volders PJ, Van de Walle I, Benoit Y, Poppe B, Mestdagh P, Vandesompele J, Taghon T, Soulier J, Cools J, Rondou P, Van Vlierberghe P and Speleman F.
  International Conference on the Long & the Short of Non-Coding RNAs, Crete, Greece (14-19/6/2015)

- **The lncRNAomes of T-cell acute lymphoblastic leukemia and normal immature thymocyte subsets through combined poly-A and total RNA-sequencing.** Wallaert A, Durinck K, Van Loocke W, Van de Walle I, Rondou P, Soulier J, Van Vlierberghe P, Taghon T and Speleman F.
  Noncoding RNAs in Health and Disease, Santa Fe, United States of America (21-24/02/2016)

Conferences

- Next Generation Sequencing and Recent Advances in Genetics, BeSHG, Liège, Belgium (02/03/2012)
- Cancer Seminar, UZ Gasthuisberg, Leuven, Belgium (25/09/2012)
- Noncoding RNAs in Development and Cancer, Keystone Symposia, Vancouver, Canada (20-25/01/2013)
- Genetics of Human Development Exposed, BeSHG, Brussels, Belgium (15/03/2013)
- Second ESH-EHA Scientific Workshop on T-Cell Acute Lymphoblastic Leukemia, Lisbon, Portugal (22-24/03/2013)
- Genome Engineering and Synthetic Biology: Tools and Technologies, VIB, Ghent, Belgium (16-17/09/2013)
- EMBO/EMBL Symposium, The Non-Coding Genome, Heidelberg, Germany (9-12/10/2013)
- 55th ASH Annual Meeting and Exposition, New Orleans, United States of America (7-10/12/2013)
- 2nd Oncopoint meeting, UZ Ghent, Ghent, Belgium (6/2/2014)
- Translational Genetics: from cage over bench to bed, BeSHG, Antwerp, Belgium (7/2/2014)
- Course and Master Classes on Molecular Aspects of Hematological Disorders, Rotterdam, The Netherlands (10-12/6/2014)
- 19th EHA congress, Milan, Italy (12-15/6/2014)
- Non-Coding RNA – From Basic Mechanisms to Cancer, Heidelberg, Germany (22-25/6/2014)
- 3rd Oncopoint meeting, Ghent, Belgium (11/2/2015)
- Research Day, Faculty of Medicine and Health Sciences, Ghent, Belgium (05/03/2015)
- IAP Progress meeting 2015, Ghent, Belgium (08/06/2015)
- Course and Master Classes on Molecular Aspects of Hematological Disorders, Rotterdam, the Netherlands (09-10/06/2015)
- International Conference on the Long & the Short of Non-Coding RNAs, Crete, Greece (14-19/6/2015)
- Genome Engineering and Synthetic Biology: Tools and Technologies, VIB, Ghent, Belgium (28-29/01/2016)
- First joint meeting BeSHG/NVHG, Genetics and Society, Leuven, Belgium (04-05/02/2016)
- Noncoding RNAs in Health and Disease, Santa Fe, United States of America (21-24/02/2016)
- Research Day, Faculty of Medicine and Health Sciences, Ghent, Belgium (16/03/2016)
- Course and Master Classes on Molecular Aspects of Hematological Disorders, Rotterdam, The Netherlands (07-08/6/2016)
- Light at the dark side of the genome, fTales, Ghent, Belgium (15-16/09/2016)

Student Guidance

2014

- Lucas Verniers, 2nd bachelor Medicine, “Z-lijn” paper
  Importance of NOTCH1 as a target for future therapy in T-ALL.
- Karen Verboom, 1st Master Biomedical Sciences, Internship
  Study of long noncoding RNAs in T-cell acute lymphoblastic leukemia.

2015

- Nienke Heireman, Nathalie Peeters and Maaike Van Trimpont, Project Writing Coaching, course of Cancergenetics of Prof. Frank Speleman, 1st Master Biomedical Sciences
  The role of JAK/STAT signaling on downstream lncRNAs in ETP-ALL.
- Karen Verboom, 2nd Master Biomedical Sciences, Masterthesis
  Study of long noncoding RNAs in T-cell acute lymphoblastic leukemia.
- Laura Keersmaekers and Maxim De Maere, 2nd bachelor Medicine, "Z-lijn" paper
  Prognostic factors in T-ALL.

2016

- Tessa Van Steenstraete and Michiel Vercruysse, 2nd bachelor Medicine, "Z-lijn" paper
  Noncoding RNAs as biomarkers in Cancer.