UNIVERSIDADE DO ALGARVE

EPIGENETICS AND ALTERNATIVE SPLICING

Soraia Vanessa Guerreiro da Silva

Dissertação de Mestrado em

Oncobiologia – Mecanismos moleculares do Cancro

Trabalho efectuado sob a orientação de:

Maria Carmo-Fonseca, MD PhD

Noélia Custódio, PhD

Ana Teresa Maia, PhD

2015

Dissertação de Mestrado em Oncobiologia – Mecanismos Moleculares do Cancro

Epigenetics and alternative splicing

Declaro ser a autora deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

_____

AGRADECIMENTOS

O percurso que conduziu ao termo da presente investigação foi marcado por inumeráveis vivências ao nível pessoal e profissional. Representa um processo de amadurecimento nestes dois domínios, o qual não teria o mesmo valor sem a amizade, apoio e dedicação de todas as pessoas que fizeram e fazem parte da minha vida. Assim, quero expressar a minha profunda gratidão a todos aqueles que tornaram este relatório possível, em especial:

Às minhas orientadoras, a Professora Carmo, pela sua inspiração, atenção e disponibilidade em receber-me no seu laboratório; à Noélia, por ter sido tão prestável, por me ter ajudado e incentivado o meu trabalho no laboratório; e à professora Ana Teresa que apesar de não ter estado tão presente na parte prática, esteve presente na minha formação e ajudou-me a que não "caísse de pára-quedas" no laboratório.

À Rita Drago que me ensinou muitas das experiências que realizei neste trabalho e que esteve sempre presente.

À Joana Tavares que me deu uma ajuda preciosa na análise bioinformática.

Às minhas companheiras de mestrado, Sara e Vanessa, que viveram comigo esta experiência maravilhosa.

A todas as outra pessoas do laboratório, que me ajudaram sempre que precisei.

A todos os meus amigos, namorado e familiares que me acompanharam e me apoiaram nesta jornada da minha vida.

Um muito obrigado a todos!

RESUMO

Epigenética é a área da genética que se foca no estudo das alterações biológicas da célula que não envolvem alterações na sequência de nucleótidos do DNA. Um dos componentes da epigenética que tem vindo a ganhar interesse na comunidade científica são os RNAs longos não codificantes (do inglês *long noncoding RNAs* - lncRNAs) que são transcritos que contém mais de 200 nucleótidos. Estes não possuem quadros de leitura abertos (do inglês *open reading frames* – ORFs) e desempenham papéis biológicos importantes em diferenciação celular, pluripotência, regulação da transcrição, processamento e tradução de moléculas de RNAs. Têm sido também muito associados ao desenvolvimento de cancro, nomeadamente, na progressão tumoral e desenvolvimento de metáteses. Várias classes de lncRNAs têm sido descritas tendo em conta, maioritariamente, a localização destes transcritos no genoma em relação a transcritos com potencial codificante, os RNAs mensageiros (mRNAs). Uma classe de lncRNAs com interesse neste projecto é a dos transcritos anti-direccionais naturais (do inglês *natural antisense transcripts* – NAT). Estes transcritos têm a particularidade de serem codificados na cadeia anti-direccional de genes que codificam mRNAs, podendo haver sobreposição parcial com a região promotora ou com intrões. Pensa-se que poderão estar implicados na regulação da transcrição dos genes codificantes de mRNA ou na regulação da remoção dos intrões (splicing).

O presente trabalho é parte de um projecto que tem como objectivo principal investigar a biologia dos lncRNAs no contexto do desenvolvimento de leucemia. Embora já exista evidências recentes que destacam a importância de lncRNAs na regulação da expressão genética, pouco se sabe sobre o seu papel na diferenciação das células T e na transformação leucémica. O objectivo final do projecto em si, é encontrar possíveis alvos para terapias direccionadas a moléculas de RNA em células cancerígenas de Leucemia Linfoblástica Aguda de células T (do inglês *T-cell Acute Lymphoblastic Leukaemia* - T-ALL). A metodologia proposta neste projecto combina técnicas de alta resolução de epigenómica, transcriptómica e biologia molecular com abordagens para monitorizar a síntese, tempos de semi-vida e localização sub-celular de lncRNAs. A análise está focada em precursores de células T primárias purificadas a

partir de tumores do timo de ratinho e em modelos celulares de T-ALL de ratinho. O factor que diferencia e dá o carácter leucémico a estas células é a sobreexpressão do oncogene TLX3 que é considerando um dos genes mais mutados neste tipo de leucemias. No entanto, estudos anteriores mostraram que, por si só, esta sobreexpressão do oncogene não é suficiente para induzir a T-ALL, deste modo, poderão existir outros factores, tais como lncRNA, que estejam envolvidos no desenvolvimento da T-ALL. No âmbito deste estudo, foi seleccionado a partir da literatura uma lista de lncRNAs que são expressos em células T e podem ser relevantes no contexto da leucemia.

Para monitorizar o tempo de semi-vida de lncRNAs realizou-se marcação do RNA nascente nas células com um pulso de incorporação (do inglês, *pulse labelling*) do análogo do uracilo, 4-thioridine (4sU), que é incorporado em todo o RNA sintetizado na célula durante esse pulso de marcação. Segue-se a extracção de todo o RNA da célula e purificação dos RNAs nascente marcados com 4sU, e dos pré-existentes e dos não marcados. A quantificação por RT-qPCR dos lncRNAs de interesse nas diferentes populações de RNA permite o cálculo da semi-vida desses lncRNA. Os resultados obtidos neste trabalho corroboram os dados já conhecidos de outras investigações dando validade e eficácia da técnica experimental executada.

Para determinar a localização sub-celular de lncRNAs foi desenvolvido um ensaio de hibridação *in situ* de fluorescência com base em sondas de LNA e amplificação do sinal de hibridação das sondas com base em sistemas que utilizam métodos enzimáticos associados a fluorescência. Neste caso, os resultados não foram conclusivos, precisando esta técnica experimental ser melhorada e optimizada.

Os lncRNAs que serão analisados por estes ensaios, no futuro, serão fornecidos por meio de análise bioinformática do transcritoma de células T em diferentes fases de transformação leucémica. No final deste estudo iniciou-se esta análise bioinformática com dados de sequenciação de RNA (RNA-seq) obtidos de células T não transformadas e de células T em diferentes fases de transformação leucémica Nesta análise pretendeu-se ter uma ideia geral dos lncRNAs e mRNAs que se encontram diferencialmente expressos entre fases diferentes de transformação leucémica. Os resultados preliminares desta análise sugerem que existe uma percentagem maior de mRNAs diferencialmente expressos do que lncRNAs, quando se comparam células não transformadas com células em diferentes fases de transformação leucémica. O objectivo é identificar entre

os lncRNAs diferencialmente expressos aqueles que poderão ser relevantes na transformação leucémica. Estes serão alvo de estudos funcionais utilizando as técnicas optimizadas neste estudo, de modo a que se perceba o seu mecanismo de acção e se possa avaliar se têm potencial para ser alvos para terapias direccionadas.

ABSTRACT

Epigenetics is the field of genetics that studies the alterations in the transcriptional potential of a cell without interfering with the DNA sequence. One of its component is the long noncoding RNAs (lncRNAs) which are transcripts with more than 200 nucleotides and no evident open reading frames (ORFs) that play important biological roles like transcription and splicing regulation and have been associated with carcinogenesis. Several classes of lncRNAs have been described according to their genomic location in relation to protein-coding genes.

The present work is part of a project aiming at gaining novel insights into the biology of lncRNAs in the context of leukemogenesis. Although recent evidence highlights the importance of lncRNAs in regulation of gene expression, little is known about their role in T-cell differentiation and leukaemic transformation. The ultimate goal of the project is finding possible targets for RNA therapeutics in T-cell Acute Lymphoblastic Leukaemia (T-ALL). The proposed methodology combines high-throughput epigenomics, transcriptomics and systems biology approaches with techniques to monitor synthesis, lifetime and sub-cellular localization of lncRNAs. The analysis is focused on primary T-cell precursors purified from the mouse thymus and on cellular mouse models of T-ALL.

The present study aims to develop functional assays to monitor lifetime and sub-cellular localization of lncRNAs in a cellular mouse model of T-ALL. To monitor the lifetime of lncRNAs we carried out pulse labeling with the uridine analogue 4-thioridine (4sU) followed by purification of labeled nascent, pre-existing unlabeled and total cellular RNAs. RT-qPCR quantification of the RNA subsets allows the estimation of the lncRNA's half-life. To determine the sub-cellular localization of lncRNAs we developed a fluorescence *in situ* hybridization assay based on LNA probes and enzyme-based signal amplification. In this study we selected from the literature a list of lncRNAs that are expressed in T-cells and may be relevant in leukemogenisis. The candidate lncRNAs that will be analysed by these assays in the future will be provided by genome-wide transcriptomic analysis of different stages of T-cell leukemic transformation.

CONTENTS

FIGURES INDEX

TABLES INDEX

ABBREVIATIONS

4sU	Uridine analogue 4-thioridine
α-Am	α-amanitin
ABL1	Abelson Murine Leukemia Viral Oncogene Homolog 1
ActD	Actinomycin D
Airn	Antisense of IGF2R Non-Protein Coding RNA
ALL	Acute Myeloid Leukemia
AML1	Acute Myeloid Leukemia 1 Protein
AP	Alkaline Phosphatase
Bio	Biotin
BSA	Bovine Serum Albumin
CD1a	Cluster of Differentiation 1a
CD3	Cluster of Differentiation 3
CD34	Cluster of Differentiation 34
CD4	Cluster of Differentiation 4
CD5	Cluster of Differentiation 5
CD8	Cluster of Differentiation 8
CLP	Common Lymphoid Progenitor
Daxx	Death Domain-associated Protein
Dig	Digoxigenin
DN	Double Negative
DNA	Deoxyribonucleic Acid
DRB	5,6-dichloro-1–D-ribofurano-syl-benzimidazole
DP	Double Positive
EDTA	Ethylenediaminetetraacetic Acid

ELF	Enzyme-Labelled Florescence
Ens87	ENSMUST00000156387
Ens91	ENSMUST00000130391
ErbAα2	Thyroid Hormone receptor
eRNA	Enhancer RNA
ETP-ALL	Early T-cell Precursor – Acute Lymphoblastic Leukaemia
ETS1	Protein C-ets-1
Eα	Enhancer alpha
Fadd	Fas-Associated Protein with Death Domain
FBXW7	F-box/WD Repeat-containing Protein 7
FISH	Fluorescent *In Situ* Hybridization
FPKM	Fragments Per Kilobase of transcript per Million
Gas5	Growth Arrest-specific 5
H_2O_2	Hydrogen Peroxide
HCl	Hydrogen Chloride
HD	Homeodomain
Hotair	HOX Transcript Antisense RNA
HOX	Family of Homeobox proteins
HRP	Horseradish Peroxidase
Igf2r	Insulin-like Growth Factor 2 Receptor
JAK/STAT	Janus Kinase/Signal Transducer and Activator of Transcription
LEF1	Lymphoid Enhancer-binding Factor 1
LincRNA	Long Intergenic Noncoding RNA
LNA	Locked Nucleic Acid
LncRNA	Long Noncoding RNA
LMO1	Rhombotin 1
NKX	Homeodomain Transcription Factor

MALAT1	Metastasis Associated Lung Adenocarcinoma Transcript 1
MHC	Major Histocompatibility Complex
mRNA	Menssenger RNA
NaCl	Sodium chloride
NAT	Natural Antisense Transcript
ncRNA	Noncoding RNA
NEAT1	Nuclear Enriched Abundant Transcript 1
NKL	Natural Killer Lymphocyte
NOTCH1	Notch Homolog 1, Translocation-associated
NPC	Nuclear Pore Complex
Nup214	Nucleoporin 214
ORF	Open Reading Frame
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
qPCR	Quantitative Real-Time PCR
RNA	Ribonucleic Acid
RNA-seq	RNA Sequencing
RT	Room Temperature
RT-qPCR	Retro Transcription – qPCR
RUNX1	Runt-related Transcription Factor 1
Ser/Arg	Serine/Arginine Aminoacids
smRNA-FISH	Single Molecule RNA – FISH
SP	Single Positive
SSC	Buffer Solutions with Citrate
T-ALL	T-cell Acute Lymphoblastic Leukemia
TAL1	T-cell Acute Lymphocytic Leukemia Protein 1

TCRα	T-cell Receptor Alpha
TCRβ	T-cell Receptor Beta
TCRγ	T-cell Receptor Gamma
TCRδ	T-cell Receptor Delta
TF	Transcription Factor
TLX1	T-cell Leukemia Homeobox Protein 1
TLX3	T-cell Leukemia Homeobox Protein 3
TSA	Tyramine Signal Amplification
TUG1	Taurine Upregulated Gene 1
VRC	Variable Resin Chemistry
XIST	X-inactive Specific Transcript
ZEB2	Zinc Finger E-box-binding Homeobox 2
Zeb2NAT	Zeb2 – Natural Antisense Transcript

1. Introduction

1.1. The "new" RNA world – Long noncoding RNAs

The general view of the genome and gene regulation in biology has been essential centred in protein-coding genes via the central dogma enunciated by Francis Crick in 1958: RNA is transcribed from DNA and translated into protein (Crick 1970). Since then, studies have pointed to the presence of large amounts of RNA that was transcribed but did not encode proteins which represents about 80% of the whole genome (Figure 1.1, **A**). The characterisation of these noncoding RNAs (ncRNAs) was restricted to a few housekeeping genes (including ribosomal RNAs, transfer RNAs and small nucleolar RNAs) and to a small number of regulatory RNAs (Rinn & Chang 2012). Currently, the number of novel ncRNAs has increased drastically and more is known about their function, biogenesis, length, structural and sequence features. According to their size, ncRNAs are divided into two major groups: small noncoding RNAs, including microRNAs, and long noncoding RNAs (lncRNAs) (Figure 1.1, **A** and **B**), that are a large and diverse class of transcribed RNA molecules with a length of more than 200 nucleotides that resemble messenger RNA in structure and size, but do not encode proteins. Only a relatively small fraction of lncRNAs have been so far characterised and we can't start to classify different types of lncRNA according to their functions because we are still far from being able to predict the function of new lncRNAs. Nevertheless, from what has already been unravelled they are thought to carry out important regulatory functions, adding yet another layer of complexity to our understanding of genomic regulation.

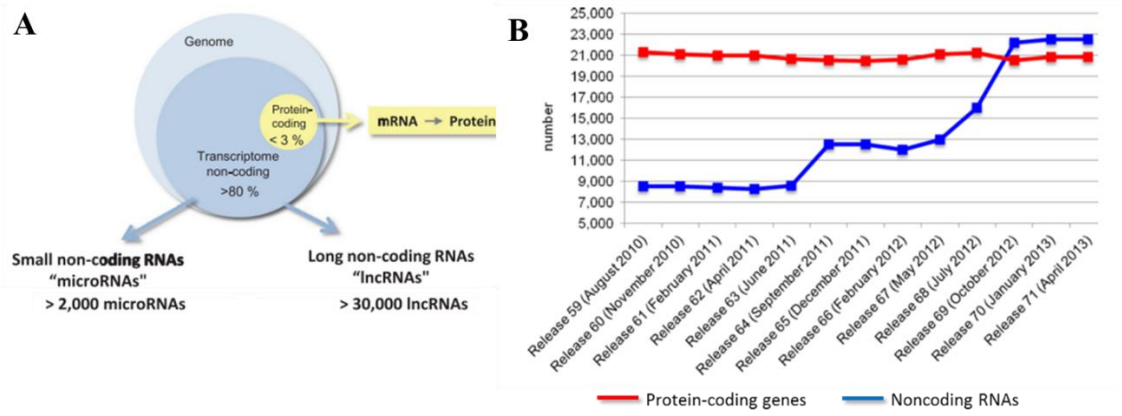
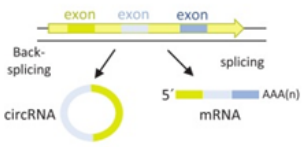


***Figure 1. 1** – Overview of the coding vs noncoding RNAs in the human genome. (**A**) Representation of the relative percentage of protein-coding and noncoding. (Adapted from Uchida & Dimmeler, 2015) (**B**) Number of noncoding (blue line) and protein coding (red line) annotated on Ensembl until 2013. The x-axis indicates the number and the date of the release (Adapted from Bussotti et al., 2013).*

1.1.1. Classification of lncRNAs

Long noncoding RNAs (lncRNAs) are generally transcribed by RNA polymerase II and undergo co-transcriptional modifications such as capping, polyadenylation and pre-RNA splicing. They harbour standard canonical splice site signals, but have less exons and are generally shorter than mRNAs (Derrien et al. 2012). The predicted open reading frames (ORFs) have a poor start codon and ORF contexts, therefore, the translation of lncRNAs is not likely to happen (Li & Chen 2013). As such, lncRNAs are defined as endogenous cellular noncoding RNA molecules longer than 200 nucleotides in length. In summary, they have small coding potential, can be spliced, capped and polyadenylated. A major feature is that these transcripts are differential expressed in tissues/cells or developmental stages (Bussotti et al. 2013). Some researchers have been classifying lncRNA based on their genomic proximity between neighbouring transcripts and established six main categories of lncRNAs: (**a**) intergenic or lincRNA; (**b**) intronic; (**c**) sense; (**d**) antisense or NAT (**e**) Enhancer or eRNA and (**f**) circular RNA (see details in Table 1.1).

***Table 1. 1** – Classes of long noncoding RNA with their representation and description. In yellow are represented protein-coding genes (Adapted from He et al., 2014; Mattick & Rinn, 2015 and Uchida & Dimmeler, 2015).*

Class	Representation	Description
(**a**) Intergenic		Do not overlap with protein-coding genes. Also known as long intergenic noncodig RNA (lincRNA).
(**b**) Intronic		Derived from an intron of a second transcript.
(**c**) Sense		Overlapping a protein-coding gene and usually sharing the same promoter.
(**d**) Antisense		Located in antisense orientation, in an opposite strand, to a protein-coding gene. Also known as Natural Antisense Transcript (NAT).
(**e**) Enhancer		Arises from the enhancer region of a protein-coding gene. Also known as enhancer RNA (eRNA).
(**f**) Circular RNA		Forms a covalently enclosed circular RNA, which usually derives from splicing of a protein-coding gene.

1.1.2. Function of lncRNAs

The first example of an lncRNA that was characterized as having a function was H19, in 1988. This lncRNA was identified as an RNA overexpressed during liver development in the mouse. The mouse H19 transcript lacked a large ORF, instead, only had small sporadic ORFs that were not evolutionarily conserved, that did not template translation in vivo and did not produce a recognizable protein product (Brannan et al. 1990). One year later, another ncRNA, termed XIST, was found to be expressed exclusively from the inactive X chromosome and then demonstrated to be required for X inactivation in mammals (Penny et al. 1996).

Nowadays, a large number of lncRNAs has been described and it's known that they can influence many cellular processes such as spatial conformation of chromosomes, chromatin and DNA modifications, RNA transcription, pre-mRNA splicing, mRNA degradation, and mRNA translation. It has been reported that they can also produce small biological active peptides (see details in Figure 1.2) (Morlando et al. 2015).

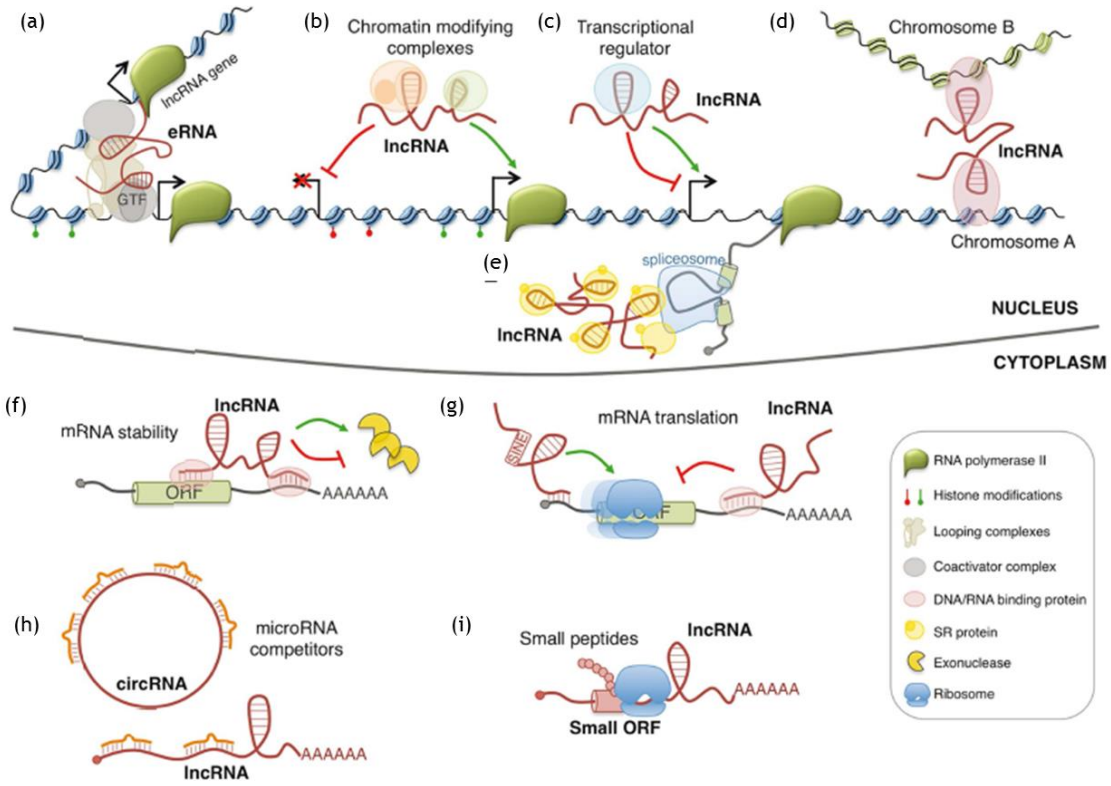


Figure 1. 2 *– Example models of lncRNA functions. Nuclear lncRNAs can regulate transcription by acting as enhancer RNA (eRNA) (**a**), by recruiting chromatin modifying complexes (**b**), or regulating transcription factors activity (**c**). They can also regulate gene expression by acting on the spatial conformation of chromosomes (**d**) or affecting pre-mRNA splicing (**e**). mRNA expression can be affected by cytoplasmic lncRNAs which regulate mRNA stability (**f**), mRNA translation (**g**) and compete with microRNA binding (**h**). A few lncRNAs contain small open reading frames (ORFs) that can be translated in biological active small peptides (**i**). (Adapted from Morlando et al., 2015).*

One of the most studied lncRNAs is the mammalian Airn (antisense of Igf2r noncoding RNA). Airn is located in antisense direction of Igf2r (insulin-like growth factor 2 receptor), using a bidirectional promoter. It can influence epigenetic events through transcription-dependent mechanisms resulting in the silencing of Igf2r transcription (Latos et al. 2012a).

Another example of lncRNA function can be described by TUG1 (taurine upregulated 1) and MALAT1 (meta- stasis associated lung adenocarcinoma transcript 1) that are linked to gene activation and repression through the organization of nuclear subdomains. These two transcripts bind Polycomb 2, but TUG1 binds methylated Polycomb 2 and MALAT1 binds the unmethylated protein. The methylation status of Polycomb 2 dictates a switch in both its lncRNA-binding specificity and nuclear sub compartment localization. This switch is accompanied by movement of Polycomb 2

target genes between active and repressive nuclear domains and ultimately influences downstream gene expression (Yang et al. 2011).

LncRNAs can also act as regulators of mRNA processing. Nascent pre-mRNAs are spliced and processed into one of potentially many isoforms. Alternative splicing and editing contribute to the increasing gene isoform diversity (Geisler & Coller 2013). LncRNA transcripts that have an antisense orientation to known protein-coding genes, also known as natural antisense transcripts (NATs), can influence splicing patterns of mRNAs. One example is the Zeb2NAT that interferes with the splicing of ZEB2 (zinc-finger E-box binding homeobox 2) pre-mRNA. Zeb2NAT expression inhibits splicing of an internal ribosome entry site (IRES)-containing intron. Translation of ZEB2 is dependent on this IRES, and therefore, expression of the NAT indirectly enables expression of ZEB2 protein (Beltran et al. 2008). ErbAα2 overlapping antisense transcript it's another NAT which its expression controls the alternative splicing of the thyroid hormone receptor ErbAα2 mRNA to form two antagonistic isoforms (Sleckman et al. 1998). In general, the mechanism by which NATs influence splicing is still unclear, but it has been hypothesized to involve splice-site masking and a subsequent block in spliceosome recruitment (Faghihi & Wahlestedt 2009).

MALAT1 (mentioned above), which is an intergenic lncRNA, also affects splicing but through a more indirect mechanism. This lncRNA associates with interchromatin granules and has been implicated in alternative splicing through the modulation of active Ser/Arg splicing factors. Ser/Arg proteins are important regulators of alternative splicing and MALAT1 interacts with them and influences the nuclear distribution and levels of phosphorylated Ser/Arg proteins. The depletion of MALAT1 changes the alternative splicing patterns of the pre-mRNAs that they target (Tripathi et al. 2010a).

LncRNAs have also been implicated in nuclear organization through the scaffolding of sub-nuclear domains. Both coding and noncoding RNAs have been associated in the nucleation of histone locus bodies, paraspeckles and nuclear stress bodies. The best-studied lncRNA of this type is NEAT1, which is important for the de novo assembly of paraspeckles (sub nuclear domains that may mediate retention of hyperedited mRNAs in the nucleus) and this lncRNA is significant for the maintenance of these paraspeckles (Mao et al. 2011).

1.2. Long noncoding RNAs in Cancer

Cancer causes 20% of deaths in Europe with more than 3 million new cases and 1.7 million deaths each year. Since the Human Genome Consortium released its final draft of the human genome, in 2001, the study of lncRNAs have been strongly associated with cancer (Gutschner & Diederichs 2012). LncRNAs are implicated in serial steps of cancer development since they can interact with DNA, RNA, protein molecules and/or their combinations. Their deregulation confers capacities for tumor initiation, growth, and metastasis (Cao 2014).

Despite the small number of well characterized lncRNAs associated with cancer, they already have several possible clinical benefits by offering great potential as novel biomarkers for diagnosis, prognosis, metastasis and predicting response to therapy (Cheng et al. 2013). For example, overexpression of MALAT1 in small cell lung cancer is an indicator of early metastasis and poor prognosis (Gutschner et al. 2013). Other studies have been developing a target approach for lncRNAs (Lee 2012). The single-nucleotide resolution of the sequencing results allows characterization of the structure, potential function and disease-associated polymorphisms of the lncRNAs (Li & Chen 2013). Recent improvement of biological drugs has broadened the types of therapeutic targets, which enables strategies targeting RNA molecules (Davis et al. 2010). These strategies show promising results for the improvement of lncRNA-based cancer therapy (Li & Chen 2013).

1.2.1. T-cell differentiation and Leukemogenisis

This work will be focused in one specific type of blood cancer, T-cell Acute Lymphoblastic Leukaemia (T-ALL). To understand how this cancer develops, referred as leukemogenisis, is important to explain basic concepts of the immune system such as T-lymphocyte differentiation.

T-lymphocyte differentiation occurs in the thymus and requires the expression of several genes in a defined temporal manner. The main characteristic that separate cells from different phases of development is the cell surface receptors that are being expressed or unexpressed during each state of differentiation. Common lymphoid progenitors (CLPs) enter into the thymus through the cortico-medullary junction and initiate commitment to the T cell lineage. At this point the cells are called double negative (DN) thymocytes because they don't express $CD4^+$ or $CD8^+$ receptors. The V, D, and J gene segments at the TCRβ, TCRγ, TCRδ or TCRα loci determine the development into either γδ or αβ T cell lineages. First, the T cell receptor (TCR) β, γ and δ loci becomes accessible and begins to be rearrange toward the γδ lineage for the expression of the pre-TCR, which helps drive cellular proliferation and leads to the $CD4^+8^+$ stage, also called double positive (DP) thymocytes (Sleckman et al. 1998). Pre-TCR surface expression, referred to as the β-selection process, is marked by arrest of TCRβ gene rearrangements and extensive cellular expansion. Afterwards, TCR*α* locus rearrangement leads to the expression of the mature TCR. Signals from the pre-TCR allow survival, cellular expansion and further differentiation of T-cells with productive rearrangement of the TCRβ (β-selection). The mature TCR, is triggered by its interaction with MHC molecules on thymic epithelial cells and allows the selection of TCR non-self-reactive thymocytes that recognize self-MHC molecules and the production of single positive (SP) mature T-cells (Figure 1.3).

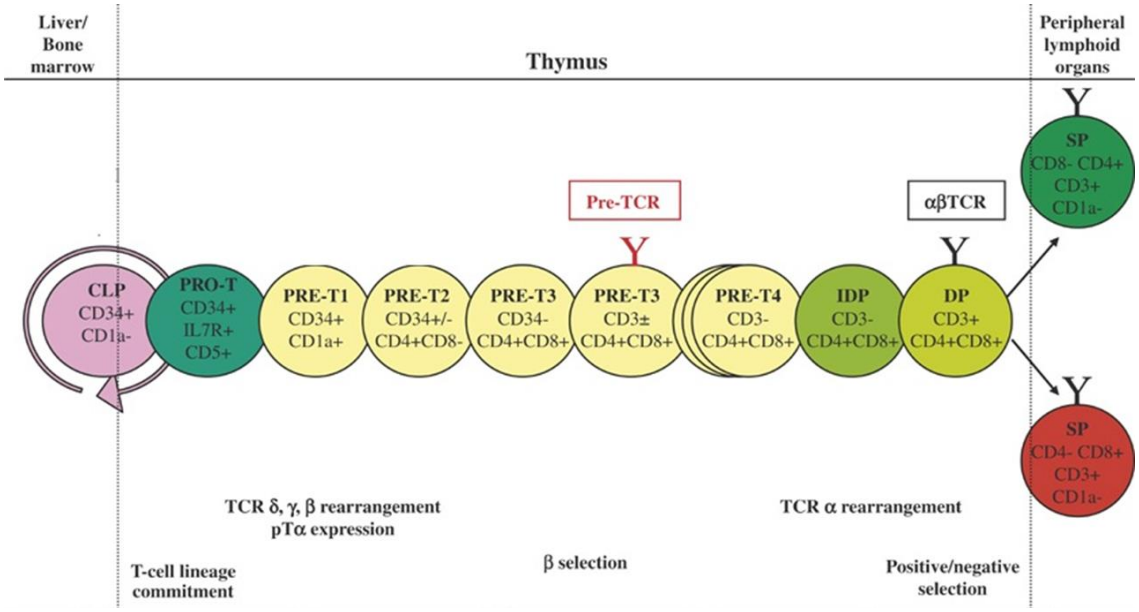


Figure 1. 3 *– T-cell differentiation process. T-cell differentiation occurs in the thymus where the ordered somatic recombination of V, D, and J gene segments at the TCRβ, TCRγ, TCRδ or TCRα loci determine the development into either γδ or αβ T cell lineages. Progressive lineage restriction and acquisition of T cell potential following migration from the bone marrow to the thymus involve successive differentiation steps defined by the acquisition of a number of surface molecules, including CD5, CD1a, CD34, CD3, CD4, and CD8(Adapted from Graux et al. 2006).*

The TCRα rearrangement is a highly regulated process, in which the TCRα enhancer (Eα) plays a primary role (Sleckman et al. 1998). The minimal Eα core contains binding sites for three transcription factors (TFs), LEF-1, RUNX1/AML1, and ETS1, which have been demonstrated to be crucial for the transcriptional and cis-chromatin opening activities of the Eα enhanceosome (Dadi et al. 2012).

Despite the differences between human and mouse thymocyte maturation, they both follow the same main developmental stages. Nevertheless, the understanding of the different processes and players in T-lymphocyte maturation comes mainly from experiments in mice (Graux et al. 2006).

1.2.2. T-cell Acute Lymphoblastic Leukemia

T-cell acute lymphoblastic leukaemia (T-ALL) is a blood cancer characterized by an uncontrolled increase in the number of immature T lymphocytes. It accounts for approximately 15% of paediatric and 25% of adult ALL cases. Even with improved

outcome, about 25% of children and 50% of adults still fail to respond to intensive chemotherapy protocols (Kalender Atak et al. 2013). The current therapies are highly toxic. The patients with primary resistance to chemotherapy and the ones who have relapsed after initial treatment have an extremely poor prognosis. In this context, the research for novel therapeutic strategies is extremely important (De Keersmaecker et al. 2014)

1.2.3. Genetic alteration related to T-ALL

Oncogenic events in T-ALL include transcriptional activation of proto-oncogenes, inactivation of tumor suppressor genes, and activation of signal transduction pathways, such as Notch1 pathway by NOTCH1 or FBXW7 mutations (Aifantis et al. 2008). TCR chromosomal translocations represent a frequent oncogenic hallmark of T-ALL (Cauwelier et al. 2006). This translocations are generally the result from mistakes on the V(D)J recombination events that lead to the ectopic activation of oncogenes owing to their relocation to the vicinity of potent cis-activating elements within the involved TCR locus.

A specific set of recurrently over-expressed transcription factors (TFs) have been documented in T-ALL, including TLX1, TLX3, TAL1, LMO1, HOXA, and NKX family members (Van Vlierberghe et al. 2008). T-ALL samples expressing each of these transcription factors show a characteristic gene expression signature and as such these transcription factors define distinct molecular subtypes in T-ALL (Vlierberghe & Ferrando 2012). Point mutations and small insertions/deletions have also been described leading to oncogenic events, like mutations that activate NOTCH1 in more than 60% of T-ALL cases (Weng et al. 2004), or mutations in cytokine receptors and tyrosine kinases such as IL7R and JAK3 (Kalender Atak et al. 2013).

Oncogenic nucleoporin fusions have been described in diverse types of hematologic malignancies, more frequently in acute myeloid leukemia (AML), but also in T-cell acute lymphoblastic leukemia (T-ALL) (Takeda & Yaseen 2014). A NUP214-ABL1 fusion was described in a series of patients with T-ALL (Graux et al. 2004). This fusion

is cytogenetically cryptic and is often located in amplified episomes. In T-ALL, NUP214-ABL1 is usually associated with rearrangement and/or overexpression of TLX1 or TLX3, although, in some patients there are more cells with abnormalities of these genes than with NUP214-ABL1, suggesting that the protein fusion is a secondary mutation (Graux et al. 2009).

Overexpression of the orphan homeobox (HOX) proteins TLX1 and TLX3 represents the most frequent oncogenic event due to chromosomal translocation in human T-ALL. TLX1 and TLX3 belong to a subtype of HOX proteins. They contain a highly conserved homeodomain (HD) that is known to be involved in DNA and protein-protein interactions (Holland et al. 2007).

Physiological expression of TLX1 and TLX3 is restricted to embryonic development and no specific function of these genes in the T cell lineage has been reported (Vlierberghe & Ferrando 2012). Transgenic expression of human TLX1 in mice induces an initial DN2 thymic block followed by development of aneuploid T-ALL, mitotic checkpoint defects, clonal TCRβ rearrangements, a mostly cortical phenotype, and a transcriptional profile similar to that observed in human TLX1+ T-ALLs (Dadi et al. 2012).

TLX3 overexpression is a result of the translocation t(5;14)(q35;q32) in approximately 25% of pediatric and 5%–10% of adult T-ALL cases (Hatano et al. 1991). Results have demonstrated that the maturation block observed in TLX+ T-ALLs is, in large part, due to ETS1-mediated TLX recruitment to the Eα core, leading to repression of Eα and blocked Vα-Jα rearrangement. Failure to express a TCRα gene, leads to the arrest development of αβ-committed thymocytes around β-selection, when a variety of cell-proliferation signals are likely to be maintained, hence contributing to oncogenesis (Figure 1.4). This blockage can be overcome by TLX1/3 abrogation or by down-stream TCRαβ expression within an appropriate cellular context. These observations have fundamental consequences both for targeted therapy in TLX+ T-ALLs and for the role of aberrant TCR expression in T lymphoid oncogenesis (Dadi et al. 2012).

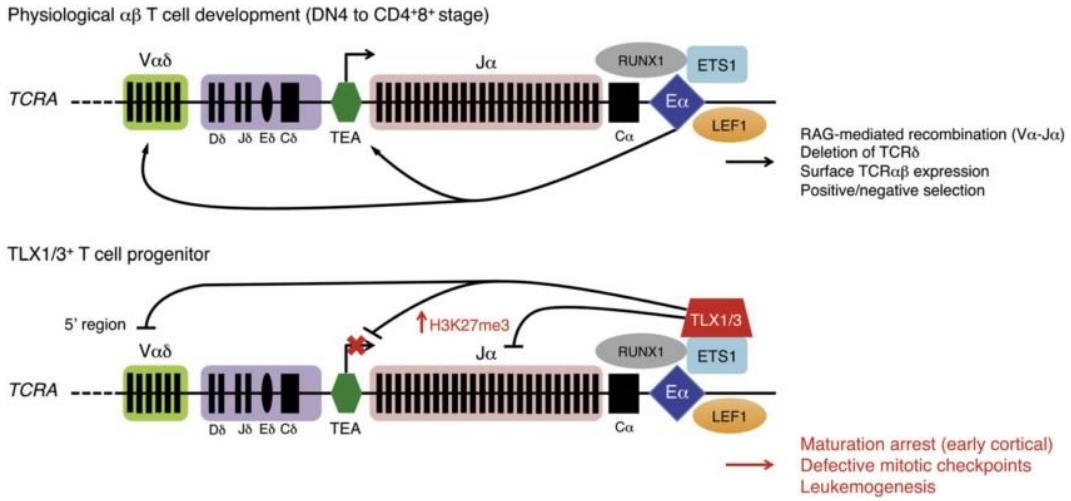


Figure 1. 4 – *TLX-Mediated Repression of the TCRa Enhancer. Transcriptional access to the TCRα locus at the late double-negative (DN) to CD4+8+ stage is regulated primarily by the function of the Ea enhancer, shown bound by its transcriptional activators ETS-1, RUNX1, and LEF1. The onset of Va-Ja recombination begins with transcription from the TEA promoter and increased histone acetylation throughout the Jα region. In a subset of T cell acute lymphoblastic leukaemia where TLX1 or TLX3 are misexpressed (bottom), ETS-1 can recruit TLX1/3 to Ea and this correlates with an enrichment of repressive histone modifications and lack of TCRA gene expression, ultimately leading to an arrest in differentiation (Adapted from (King et al. 2012).*

1.2.1. LncRNA expression during T-cell development and differentiation

The knowledge of long non-coding RNAs in immune systems is so far limited. John S. Mattick's group was the first to discover a lncRNAs expressed in $CD8^+$ T cells and suggested that many of these transcripts likely play roles in adaptive immunity (Pang et al. 2009). There is a well-established role of transcription factors as instructive signals for cell differentiation toward a given lineage, however, other features, including components of epigenetics, can regulate the maintenance of cellular states (Ranzani et al. 2015).

Several studies have identified an important contribution of lncRNAs to the development and function of adaptive immune cells. The analysis of lncRNA expression during T-cell development (RNA-seq data from 42 different T-cell types at various developmental and differentiation stages) led to the identification of 1,524 genomic regions expressing lncRNAs that are specific for lineage or developmental stage (Hu et al. 2013).

Another example is a research focused in studying the expression of lncRNAs in $CD4^+$ T cells during development and activation. In this study they analyse several lncRNAs in different stages of T cell differentiation and compare the expression levels between different stages. They found that the expression profiles of lncRNAs in different stages of CD4+ were significantly different and that many lncRNAs may exert their function through certain mRNAs that play pivotal roles in T-cell development and activation. This study suggests that the expression of lncRNAs can regulate and were correlated with the expression of neighbouring mRNAs (Xia et al. 2014).

Recently, a lncRNA profiling in human T- and B-lymphocytes at different differentiation stages identified over 500 previously unknown lncRNAs (Ranzani et al. 2015).

1.3. Genome research of long noncoding RNAs

LncRNAs are one of the emerging topics in genome research. They can be associated with gene regulatory networks, and their deregulation may be involved in a large number of complex diseases (Figure 1.5). Taking into account more than 500 publications, the database of experimentally verified lncRNA-related diseases (LncRNADisease) did a short list of 321 lncRNAs that are associated with 221 diseases where the most common is cancer (Chen et al. 2013).

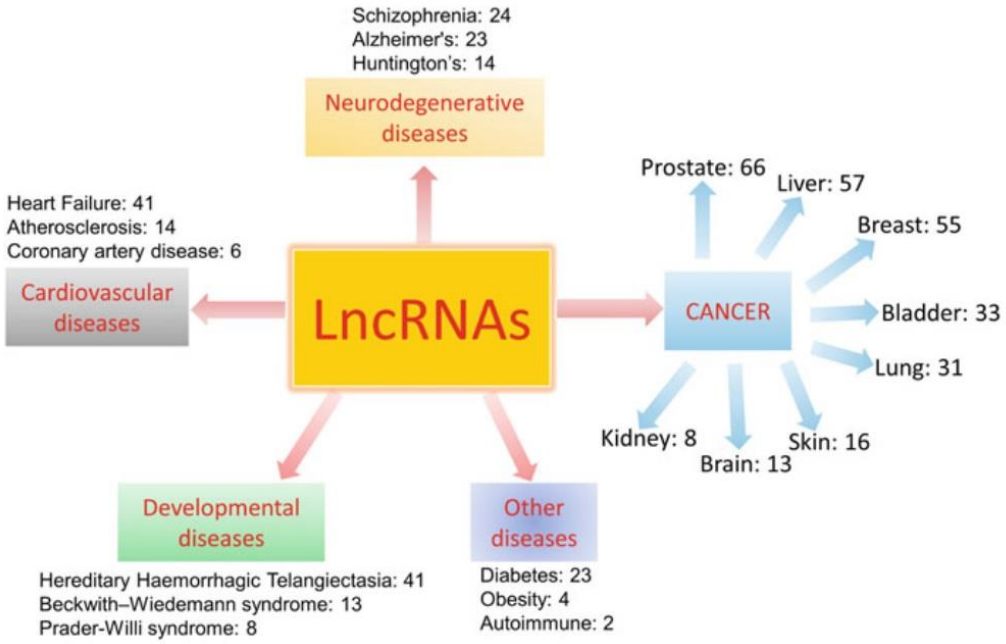


***Figure 1. 5** – Several types of diseases associated with lncRNAs. The number shown for each type of disease is the number of lncRNAs found associated with the disease by experimental evidence on interactions, epigenetics, mutation, expression, and genomic location (Adapted from Nguyen & Caninci 2015).*

The screening of lncRNAs for potential therapeutic targets is being developed and several lncRNAs have been shown as promising biomarkers in diagnosis and prognosis.

LncRNAs can be poorly conserved between species, contrasting with protein-coding genes and shorter RNAs, which leads to an additional uncertainty about whether a given lncRNA is functional. In situ hybridization, genomic, and the perturbation of their expression by overexpression and mediated knockdown are important tools to explore the roles of lncRNAs (Yan et al. 2012).

1.3.1. Expression of lncRNA

Determining under what conditions and in what cells the lncRNAs are expressed can provide significant perceptions into their function (Atkinson et al. 2012).

Studies in mouse and human demonstrated that many lncRNAs are expressed in a cell- and tissue-specific manner during development and differentiation, which suggests

that they might participate in the regulation of these biological processes (Amaral & Mattick 2008).

RNA sequencing is a good technique to quantify transcripts and enables expression levels to be easily compared across different conditions and tissues, without the need for complicated normalization methods. An RNA-seq study revealed that 78% lncRNAs are expressed as tissue-specific compared to only ~19% of the coding genes (Cabili MN et al. 2011). Another RNA-seq study has analysed the transcriptomes of 102 prostate cancer samples, defining 121 lncRNAs whose expression patterns distinguish two stages of cancer development (Prensner et al. 2011). Similar RNA-seq studies that analyse the dynamic lncRNA expression across different conditions and developmental stages in health and disease will provide comprehensive, sensitive and high-resolution data of lncRNA expression regulation (Atkinson et al. 2012).

1.3.2. Stability of lncRNA

One of the aspects of lncRNAs that is poorly understood is their post-transcriptional regulation and their metabolism in the cell. RNA levels within a cell are determined by the rates of transcription, RNA processing, and RNA decay (Windhager et al. 2012). The general expectation is that lncRNAs are less stable than protein-coding mRNAs due to their lower average level of expression and the existence of known unstable classes of lncRNAs, however, recent studies have suggested that, like has been seen for mRNAs, lncRNAs also have a wide diversity of half-lives (Dinger et al. 2009). The variation in lncRNA stability is consistent with their functional diversity and is possible that it is a reflection of their complex post-transcription regulation. Actually, post-transcriptional regulation seems to be particularly important for lncRNAs because they do not have any further translational and post-translational opportunities for regulation like protein-coding genes have (Clark et al. 2012).

Metabolic pulse labelling of nascent RNA is a powerful approach to assess the kinetics of RNA metabolism and enables the determination of their half-life. Activated uridine analogues, like 4-thiouridine (4sU), have been used to measure nascent RNA

synthesis and their metabolism at single nucleotide level with high sensitivity (Windhager et al. 2012). A genome-wide study using metabolic pulse labelling with a uridine analogue have determined similar median half-lives for mRNA and ncRNA around 3.4h in HeLa cells (Tani et al. 2012). Another genome-wide study that used treatment with actinomicin D, instead of metabolic labelling, to determine transcripts half-life reported a median half-life of 3.5h (mean 4.8h) for lncRNAs and 5.1h (mean 7.7h) for protein-coding transcripts (Clark et al. 2012).

1.3.3. Subcellular localization and single molecule quantification of lncRNA

The knowledge of the subcellular localization patterns can provide fundamental insights into the biology of lncRNAs and suggest potential molecular roles. Unlike mRNAs, which localize manly in the cytoplasm to produce their proteins, lncRNA themselves probably localize in their particular site of action, making their location within the cell important. A study in human cell lines suggests that ~30% of lncRNAs are found exclusively in the nucleus, ~15% are found exclusively in the cytoplasm, while ~50% show both nuclear and cytoplasmic localization (Kapranov et al. 2007). Sequencing studies cannot discriminate whether an lncRNA localizes in the nucleus or cytoplasm, and so there is as yet no systematic categorization of lncRNA localization patterns. This feature may be due to the fact that the expression of most lncRNAs tends to be lower than mRNA, and so their total abundance is likely far lower, which can restrict the number of sites in which a lncRNA may be active (Cabili et al. 2015).

One hypothesis is that, even with the low average abundance of lncRNAs, a small numbers of cells in the population might express high numbers of lncRNA, and this could allow an increased number of sites of action in those cells (Dinger et al. 2009). RNA fluorescence *in situ* hybridization (FISH) is a technique that can address these questions and propose potential mechanisms for lncRNA activity (Singer & Ward 1982). One of the best studied lncRNAs is XIST in which RNA FISH demonstrated that it accumulates on the inactive X-chromosome (Clemson et al. 1996). Recent examples of lncRNAs that were studied by FISH include MALAT1 and NEAT1 which were

localized to nuclear bodies (Ip & Nakagawa 2012). However, these examples represent highly abundant RNAs in the cell, though, the majority of lncRNAs are considerably less expressed, which difficult the use of conventional RNA FISH techniques that have quite low sensitivity (Cabili MN et al. 2011). One improvement of conventional RNA FISH is the application of single-transcript imaging that allows the validation of regulatory interactions. Single-molecule RNA FISH (smRNA-FISH) can provide quantitative *in situ* measurements of a variety of different transcripts and highlight the behaviour of the lncRNAs in the specific tissue or cell type that is being analysed. For example, in tumor models, single-molecule transcript imaging can enable the visualization of transcriptional heterogeneity in tumor progression and the relation between spatial context and phenotypic states of cells, represented by their expression signatures (Itzkovitz & van Oudenaarden 2011).

1.4. Aims of the Study

Although several lines of recent evidence highlight the importance of lncRNAs in regulatory gene expression networks, little is known about their role in T-cell differentiation and leukemic transformation. The present work is inserted in a project that aims at gaining a novel insight into the biology of lncRNAs in the context of leukemogenesis with the ultimate purpose of finding possible transcript targets for RNA therapeutics in T-ALL. To reach this goal the proposed methodology combines high-throughput epigenomics, transcriptomics and systems biology approaches with techniques to monitor synthesis, lifetime and sub-cellular localization of lncRNAs. The analysis is focused on primary T-cell precursors purified from the mouse thymus and on cellular mouse models of T-ALL.

The main goal of the present study is to develop functional assays to monitor synthesis, lifetime and sub-cellular localization of lncRNAs in a cellular mouse model of T-ALL. The candidate lncRNAs that will be analysed by these functional assays will be provided by the genome-wide transcriptomic analysis of different stages of T-cell leukemic transformation.

2. Methodology

2.1. Cell culture

The cultured cells used in the work were produced in our collaborators laboratory, Pierre Ferrier Lab, in France. In their laboratory, double negative (DN) thymocytes were purified from mice's thymus and transduced with the oncogene TLX3. TLX3 encodes a DNA-binding nuclear transcription factor that interacts with the transcription factor ETS1 which binds to Enhancer α and leads to maturation arrest (by inhibiting TCR rearrangements). The cells transduced with TLX3 were injected in an immune-deficient mice that developed T-ALL and the cells obtain from the tumors in lymphoid organs, named TAP cells, were stabilized in cell culture (Figure 2.1) (Dadi et al. 2012). We then received frozen TAP cells and cultured them in suspension in RPMI 1640 medium with 10% fetal bovine serum and 2mM L-glutamine at 37°C and 5% CO_2.

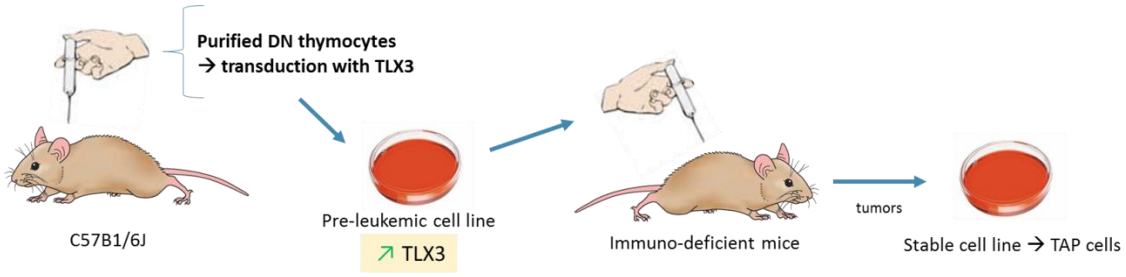


***Figure 2. 1** – Scheme of the T-ALL cell culture model developed by Dadi et al. 2012. Double negative (DN) thymocytes were purified from mice's thymus and transduced with the oncogene TLX3. These cells were then were injected in an immune-deficient mice that developed T-ALL and the cells obtain from the tumors were stabilized in a immortalized cell culture.*

2.2. RNA extraction, purification and quantification

Total RNA was extracted from cells using PureZOL (Bio-Rad) followed by DNase I treatment (Roche) according to the manufacturer's instructions. Reverse transcription was carried out with the High Fidelity cDNA synthesis Kit (Roche) according to the manufacturer's instructions, using random hexamers. Quantitative real-time PCR (qPCR) was performed using iTaqUniversal SYBR Green Supermix (Bio-Rad) in the Vii7 Real-Time PCR System (Applied Biosystems). Gene-specific primers are presented in Table 1 in the supplementary material. Each sample was run in duplicate. The $2^{-\Delta Ct}$ method was used to measure the relative changes in transcript levels using GAPDH as normalizer. At least three biological replicates were performed for all target RNAs.

2.3. DNA extraction, purification and sequencing

For total DNA extraction the cells were washed in PBS, ressuspended in Lysis Buffer (50mM Tris-HCl pH8.0; 150mM NaCl; 100mM EDTA; 1% SDS) supplemented with 0.1mg/ml of Proteinase K (Sigma) and incubated overnight at 55ºC. The genomic DNA was then recovered by phenol/cloroform extraction and ethanol precipitation. Standard PCR was performed using NZYLong DNA polymerase (NZYTech) with the followed parameters of the thermal cycler: 5 cycles of 5min at 95°C, 1min at 95°C, 1min3sec at 65°C; 35 cycles of 2min at 68°C, 1min at 68°C, 45sec at 95°C, 45sec at 65°C; 1 cycle of 1min 30sec at 68°C and 1 cycle of 10min at 68°C. The primers used to amplify and to sequence the PCR products are listed in Table 6.1 in the supplementary material. The presence of the PCR products was confirmed by agarose gel electrophoresis and the bands purified with High Pure PCR Product Purification Kit (Roche) according to the manufacturer's instructions. The purified PCR products were sequenced by STABVIDA (http://www.stabvida.net)

2.4. Labelling with 4sU-tagging

To measure transcript stability we carried out pulse labelling with the uridine analogue 4-thioridine (4sU) followed by isolation of total cellular RNAs and purification of labelled (newly synthesized) and unlabelled (pre-existing) RNA (Windhager et al. 2012) (see scheme of the protocol in Figure 2.2). Exposure of eukaryotic cells to 4sU (500µM, 60min) results in its rapid uptake, phosphorylation to 4sU-triphosphate, and incorporation into newly transcribed RNA. The protocol was performed as described by Dölken et al. 2008. Briefly, following isolation of total cellular RNA with PureZOL (Bio-Rad), the thiol-labeled RNA was biotinylated using EZ-Link Biotin-HPDP (Pierce). The cellular RNA was then quantitatively separated into labelled (newly transcribed) and unlabelled (pre-existing) RNA with high purity using µMACS streptavidin-coated magnetic beads (Miltenyi) in µMACS columns (Miltenyi) placed in an OctoMACS Seperator magnetic stand (Miltenyi). Finally, labelled RNA was recovered from the beads by simply adding a reducing agent, dithiothreitol (15,4mg/mL) that cleaves the disulphide bond and releases the newly transcribed RNA from the beads.

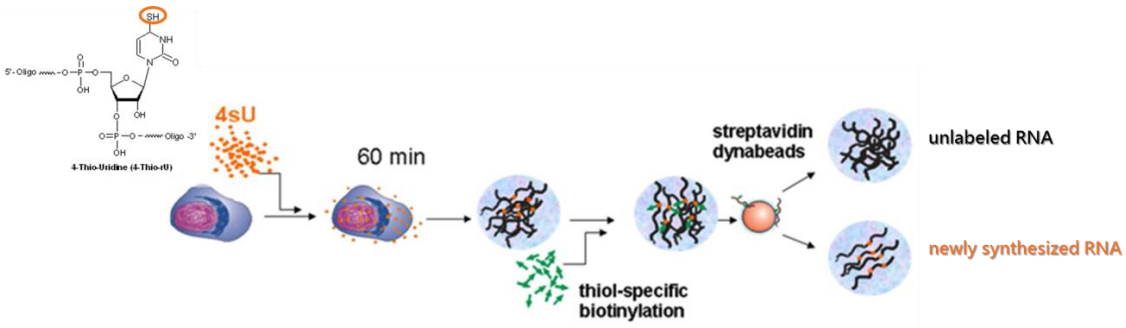


Figure 2. 2 *– Scheme of the technique of Pulse labeling with 4sU-tagging. 4-thioridine (4sU) is an uridine analogue that incorporate the RNA when it's being synthetized. The separation of the labeled from unlabeled transcripts is based on the high affinity of 4sU to biotin (Adapteded from Rädle et al. 2013).*

The ratios of newly synthesized/total RNA obtain by RT-qPCR after 60minutes of incubation with 4sU, allows non-invasive access to precise RNA half-life according to the formula given in Figure 2.3 (Rädle et al. 2013). The samples were normalized do Gapdh according to its known half-life of 7 hours (Kudla et al. 2006).

$$t_{1/2} = \frac{t_L \times \log\left(\frac{R}{norm\ factor}\right)}{\log 2} \qquad \begin{array}{l} t_L = Labeling\ time \\ R = \frac{newly\ synthesized\ RNA}{total\ RNA} \end{array}$$

Figure 2. 3 *– Mathematical equation used to determine the transcript half-life developed by Rädle et al. 2013. The key factors are the labeling time, which was 60min, and the ratio between newly synthetized and total RNA.*

2.5. LNA-RNA single-molecule FISH

Fluorescence *in situ* hybridization (FISH) is a method to detect specific nucleic acids in their cellular environment. The analysis of individual RNAs hybridized with a single FISH probe makes possible the counting of an RNA transcript with high spatial resolution (Trcek et al. 2012a). Locked nucleic acid (LNA) is a nucleic acid analogue that contains at least one nucleotide monomer with a bicyclic furanose ring locked in a conformation mimicking RNA (See Figure 2.4).

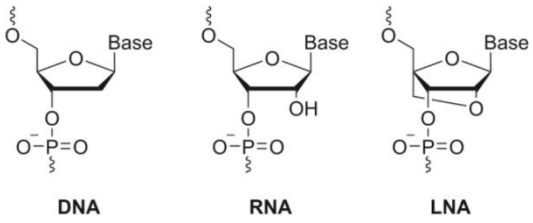


Figure 2. 4 *– Representation of the chemical structures of DNA, RNA and locked nucleic acid (LNA) nucleotide units (Adapted from Astakhova 2014).*

LNA probes provide better specificity and sensitivity and have demonstrate much higher thermal stability and higher melting temperatures when hybridized with target RNA sequences compared to unmodified counterparts (Vester & Wengel 2004). In this work we used one LNA probe for each transcript. The probes were labelled at the

5’end with either biotin, for mRNA detection, or digoxigenin for lncRNAs detection. The sequences are listed in Table 6.7 in the supplementary material.

For hybridization the cells were allowed to adhere onto poly-L-lysine coated coverslips, washed with PBS and fixed with 3.7% PFA/5% acetic acid/1xPBS for 15 min at room temperature (RT). Following fixation the cells were washed 3x5min at RT with 1xPBS/2mM VRC permeabilized in 75% ethanol and stored at -20ºC in 75% ethanol/2mM VRC. Before hybridization the cells were re-hydrated in PBS, digested with 0.1% pepsin in 0.01 M HCl with 2mM VRC for 1min at 37ºC and fixed with 3.7% PFA/1xPBS for 5 min. After 3 washes in PBS and a 5 min wash in 2xSSC/0,05% Tween20, the cells were hybridized at 55°C (30°C bellow the RNA Tm of the probes) for 1 hour with the LNA probes (Exiqon) at 40nM in hybridization mix (50% formamide/2xSSC/10% dextran sulphate/50mM sodium phosphate pH7.0. Stringent post-hybridisation washes were with 0,1xSSC/0,05% Tween 20 3x5min at 65°C. For detection of the hybridized probes the samples were subject to Tyramine Signal Amplification (TSA) system or Enzyme-Labeled Florescence (ELF) signal amplification system.

2.5.1. Tyramine Signal Amplification (TSA) system

TSA system provides enhanced sensitivity of the signal to further detection methods. The Horseradish Peroxidase (HRP) reacts with hydrogen peroxide and the phenolic part of tyramide produces a quinone-like structure with a radical on the C2 group, becoming activated. Activated tyramide then rapidly and covalently binds to all nearby tyrosine residues with proximity to the initially immobilized HRP site (Figure 2.6) (Bobrow et al. 1992).

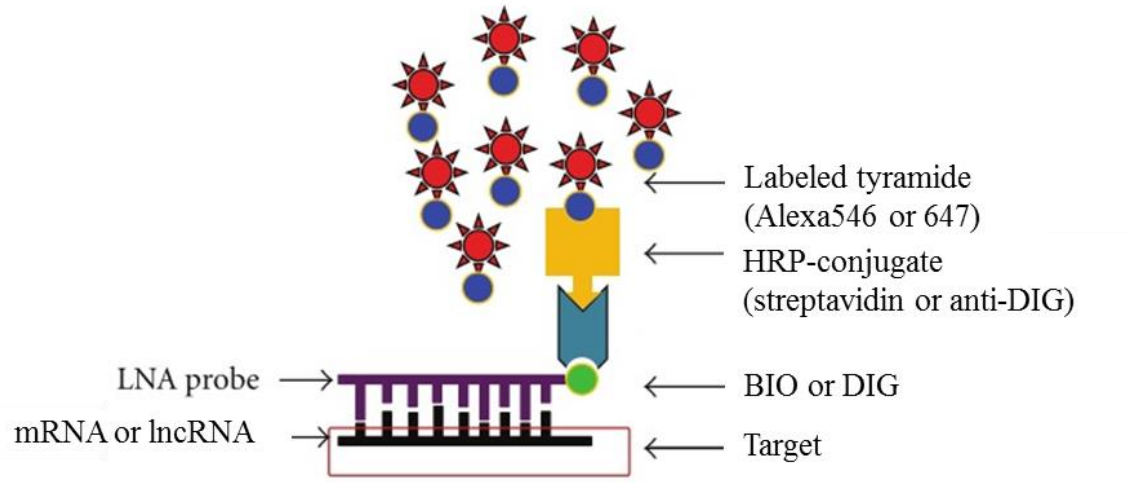


Figure 2. 5 *– Simplified scheme of TSA system for RNA detection. The LNA probes bind complimentary to the sequence of RNA. The 5' end of the probe was labeled with digoxigenin (DIG) or biotin (BIO). After hybridization HRP-conjugated streptavidin or HRP-anti-digoxigenin antibody were added to detect biotin or digoxigenin respectively, followed by the HRP substrate Alexa Fluor 546 or Alexa Fluor 647 tyramide (Adapted from Shi et al. 2012).*

In this work the Tyramide Signal Amplification Kit (Molecular Probes® by Life Technologies™) was used according to the manufacturer's instructions. Briefly, the cells were incubated in blocking buffer (1% BSA in PBS) for 30 min at room temperature. Then HRP-conjugate streptavidin (dilution 1:200 from the stock provided in the kit) or HRP-conjugate anti-digoxigenin (Abcam; dilutions tested 1:200, 1:1000 or 1:2000), both in blocking buffer, was added to the cells and incubated at room temperature for 30 min. For the multiple detections, after the first incubation with HRP-conjugate, HRP was deactivated by incubation with 1%H_20_2 in PBS for 10 min and proceeded to the next incubation with the other HRP-conjugate. After three washes in 1xPBS at 37°C, signals were amplified with labeled tyramide working solution (Alexa Fluor 546 or Alexa Fluor 647 tyramide at a 1:100 dilution in amplification buffer/0,0015% H_2O_2) for 10 min. The coverslips were then counterstained in 1mg/ ml Hoechst 33342 (Molecular Probes®) for 5 min and mounted in Prolong® Diamond Antifade Mountant (Life Technologies™).

2.6. Imaging acquisition and analysis

Samples were imaged using a Spinning Disk Confocal Microscope (Zeiss Axio Observer) equipped with an Evolve 512 EMCCD camera (Photometrics®) and an Yokogawa CSU-x1 confocal scanner (Andor). A Plan-Apochromat 100x oil- immersion objective, N.A. 1.40, was used for all imaging experiments. Briefly, after randomly selecting cells in a field, a 3D stack viewed image was taken with 0.3mm increments in the z-direction and a total of 35 sections.

2.7. RNA-seq libraries

RNA samples for Deep sequencing were prepared in Pierre Ferrier's Laboratory (Centre d'Immunologie de Marseille-Luminy, France). Deep sequencing was performed by Centro de Análisis Genómico (CNAG), Spain. The sequences obtained from the Illumina Genome Analyzer (stranded total RNA-seq with >135M reads) were aligned against the mouse genome (mmusculus.9).

2.8. Data analysis

The reads alignment to a reference genome was performed with TopHat-Bowtie2 version 2.0.9. The input arguments used were the mouse genome mmusculus.9, and the .fastq files containing the reads from both ends. In order to get only the unique alignments for a given read to the reference genome and eliminate the duplicates, the "-g" parameter was set to 1; "--num-threads" and "--mate-inner-dist" were also changed to 4 and 100, respectively, but all the remaining additional default options were left unaltered.

In order to get the fragments per kilobase of exon per million fragments mapped (FPKM) values and the differentially expressed genes, the annotation files and the

mapped reads were processed by cuffdiff tool using the 1 wild-type samples (RAGZ) against the 3 (TL3-1_1, Tumeur7_2 and TAP1B) T-ALL samples. In order to find the genes that were up- or downregulated in the samples, the downstream analysis was performed with customized scripts in R 3.1.0. The “gene_exp.DIFF” file resulted from cuffdiff analysis was filtered to include a false discovery rate (FDR) of 0.05 and a fold change of 1.5.

The differential expression of lncRNAs was obtain by running the RNA-seq data against the NONCODE v.4 database (Xie et al. 2014) and for the protein coding gene, the analysis was against the UCSC database.

3. Results

Recent findings have revealed that lncRNAs are implicated in serial steps of cancer development. LncRNAs are transcripts with more than 200 nt in length and no evident ORFs. These lncRNAs can interact with DNA, RNA and protein molecules acting as essential regulators in chromatin organization, transcriptional, splicing and post-transcriptional regulation. Their differential expression confers the cancer cell capacities for tumor initiation, growth, and metastasis (Cao 2014). Despite recent evidence highlighting the importance of lncRNAs in regulatory gene expression networks, little is known about their role in T-cell differentiation and leukemic transformation.

The main goal of the project where this work is inserted is to identify and validate lncRNAs as potential targets for RNA therapeutics in T cell acute lymphoblastic leukaemia (T-ALL). In this present work, we developed a methodology to monitor the stability of lncRNA transcripts by metabolic labelling with 4sU (Chapter 3.1) and a methodology to visualize the sub-cellular localization of lncRNA/mRNA molecules by the use of fluorescence in situ hybridization with LNA oligonucleotidic probes and confocal microscopy (Chapter 3.2). For the purpose of methodology development the lncRNA analysed in this study were selected from the literature and their expression in the T-ALL cellular model experimentally validated. The candidate lncRNAs that will be analysed by these assays in the future will be provided by genome-wide transcriptomic analysis of different stages of T-cell leukemic transformation (Chapter 3.3).

3.1. Analysis of lncRNA stability in a T-ALL cellular model

Cellular RNA levels are determined by the interplay of tightly regulated processes for RNA transcription and degradation. It has been proposed that the half-life of each mRNA is closely related to its physiological function, so it is possible that the RNA stability of lncRNAs also reflects their function. Two independent research groups

reported that non-coding RNA (ncRNA) half-lives vary over a wide range that is comparable with that of mRNAs. In this perspective, ncRNAs with short half-lives may have regulatory functions while those with long half-lives could be involved in housekeeping functions (Clark et al. 2012; Tani & Torimura 2013).

The most widely used method for genome-wide analysis of RNA stability is based on the use of transcriptional inhibitors such as actinomycin D (ActD), 5,6-dichloro-1–D-ribofurano-syl-benzimidazole (DRB) and α-amanitin (α-Am) (Tani & Akimitsu 2012). ActD inhibits transcription initiation and elongation by intercalating into DNA, while DRB and α-Am specifically inhibit RNA Polymerase II-mediated transcription. Even though transcriptional inhibitors have been widely used for determining RNA stabilities, inhibitor-mediated global transcriptional arrest has a deeply disruptive impact on cellular physiology, including splicing and polyA (Tani et al. 2012; Friedel et al. 2009).

4-thiouridine (4sU) has been used to label endogenous RNAs in mammalian cells as a non-disruptive technology for measuring RNA decay. 4sU is rapidly taken up by cells and after entering a cell, is phosphorylated by cellular uridine kinases. With the use of 4sU, additional steps, such as electroporation or lipofection, are not necessary for labeling RNA and has minimal adverse effects on gene expression (Tani & Akimitsu 2012). For this reasons we decided to follow this approach to measure the half-lives of lncRNAs in T-ALL cells.

3.1.1. Searching the literature for lncRNAs with a potential role in T-ALL

Before starting to implement the previously describes methodologies in the laboratory, we did a search in the literature and databases for lncRNAs that could have a relevant function in T-ALL and to be analysed in the study. We started by a review of the protein-coding genes that have been reported to be involved in genetic lesions that define molecular-genetic subtypes in T-ALL and other recurrent genetic alterations (Vlierberghe & Ferrando 2012). We also investigated lncRNAs that could be expressed

in this type of cells using published Chip-seq data (Hu et al. 2013) obtain from T-cell and expression levels from the NONCODE database (database for noncoding transcripts – http://www.noncode.org/). From this investigation we obtained a list of candidates that could be relevant in the context of our cellular model (see tables 3.1 and 3.2).

Table 3. 1 *– Genes involved in T-ALL and the the presence of nearby lncRNA. The reads referred in the table were obtain by the analysis of Chip-seq data from the study Hu et al. 2013.*

T-ALL genes	Reads of mRNAs in mouse's tymocytes	Presence of lncRNAs	Reads of lncRNAs in mouse's tymocytes
Jak3	~120	No lncRNA annotated nearby. At 5'end there is a slight antisense signal..	<10
Notch1	~600	Has a transcript at its 5'end that overlaps the first exon classified as antisense to the next gene.	~134
Jak1	~1700	Has a cluster nearby of intergenic lncRNAs. There's antisense signal.	~281
Tal1	<10	No lncRNA annotated nearby.	
Tlx1	<10	Has an antisense annotated in Ensembl (Tlx1os).	<10
Tlx3	<10	Has an antisense annotated in Ensembl (Gm12116).	<10
Cdkn2a/2b	<10	Has 2 noncoding transcripts (one sense and one antisense)	<10
Gata3	~600	Has an antisense that overlaps a splice junction annotated in Ensembl (Gm13256-001).	~84
Nup214	~380	Has an antisense noncoding transcript, ENSMUST00000156387.	~200-600
Il7r	~17	No lncRNA annotated nearby. Has some antisense signal in the first exon.	<10
Ezh2	~800	Has an antisense lncRNA annotated.	~70-100
Zeb2	~34	Has an antisense lncRNA, Zeb2NAT.	<10
Lyl1	~11	No lncRNA annotated nearby.	
Bcl11b	~4600	No lncRNA annotated nearby. Has antisense signal.	~697
Hoxa13	<10	Has the linRNA Hottip nearby.	<10
Nkx2-5	<10	Has a lincRNA nearby.	<10
Daxx	<10	Has an antisense annotated (Ak136742).	<10
Nrip2	<10	Has an antisense annotated (lnc1552).	<10
Fadd	~80	Has an antisense (Faddos).	~12

Table 3. 2 – LncRNA candidates. The expression levels referred in the table were taken from the NONCODE data base and are represented in FPKMs.

LncRNA	**Expression in mouse's thymus (NONCODE database)**	**Expression in human blood cells (NONCODE database)**	**Function**
Airn	**2.61896**	**0.0250699**	Imprinting at the IGF2R cluster (Latos et al. 2012b)
Hotair	**0.0390664**	**0**	Important factor in the epigenetic differentiation of skin (Gezer et al. 2014)
Tug1	**4.36793**	**0.55526**	Methylation of Polycomb 2 (Tani & Torimura 2013)
Malat1	**41.2548**	**0.011162**	Methylation of Polycomb 2 and Ser/Arg splicing factor regulation (Tripathi et al. 2010a)
Gas5	**5.14633**	**0**	Apoptosis and progression of some types of cancer (Gezer et al. 2014)
Neat1	**0.506525**	**0.0842401**	Formation and integrity of nuclear paraspeckes (Tripathi et al. 2010b)
Ens91 (ENSMUST00000130391)	**3.89755**	**Unknown**	Unknown (Xia et al. 2014)
Ens87 (ENSMUST00000156387)	**6.9745**	**Unknown**	Unknown (Xia et al. 2014)

We selected the lncRNAs Malat1, Neat1 and Airn because they have been well characterized (Cheetham et al. 2013; Fatica & Bozzoni 2014) and are potentially expressed in our cell model. We select two other lncRNAs that have not been characterized, ENSMUST00000130391 and ENSMUST00000156387 (referred now on as Ens91 and Ens87). Ens87 is a lncRNA antisense (NAT) of the protein-coding gene Nup214 and was selected from a study where the authors compared different stages of T-cell differentiation and this one was reported as being overexpressed in the DN thymocytes (Xia et al. 2014), a stage similar to the stage of the TAP cells. From the same study, we select Ens91, an intergenic lnRNA that was also upregulated in the DN thymocytes. We also choose to analyse Zeb2NAT, the antisense transcript of Zeb2 because of our interest in studying the dynamics of the mRNA and its correspondent NAT. In parallel we analysed the protein-coding genes Nup214, Zeb2 and, as controls, Fos, β-actin (β-Act) and Gapdh.

3.1.2. Determining lncRNAs/mRNA expression levels by quantitative RT-qPCR

There is evidence that lncRNAs are expressed in a cell- and tissue-specific manner during development and differentiation (Amaral & Mattick 2008), To determine which of the selected RNAs are expressed in our T-ALL cellular model (TAP cells) we performed RT-qPCR and determined their expression levels. The results obtained are represented in the graphics of Figure 3.1.

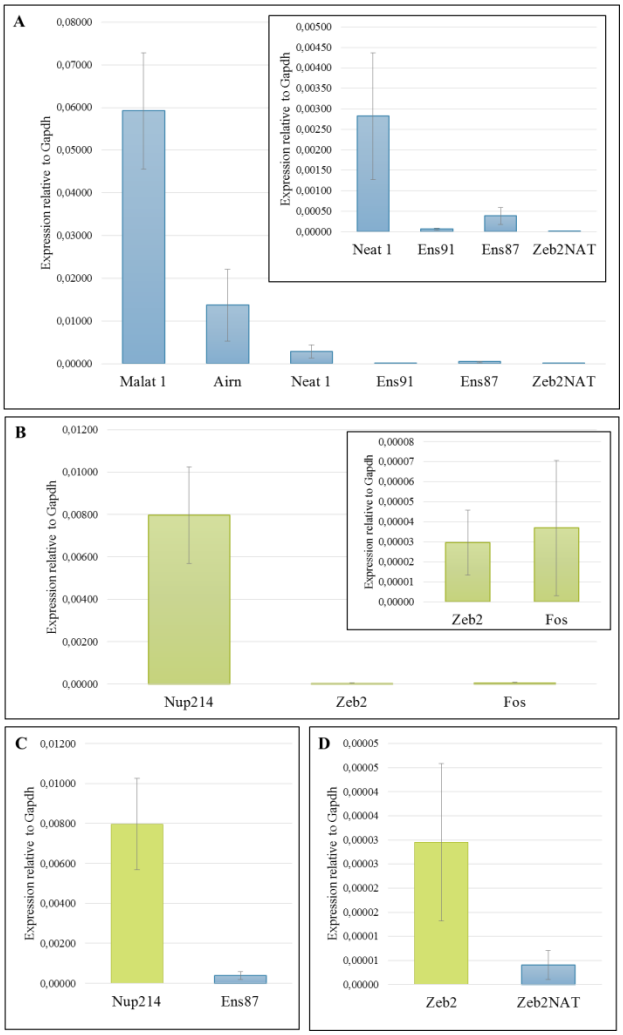


Figure 3. 1 *– Relative expression levels of selected lncRNAs and mRNAs in TAP cells. Total RNA was purified from the cells, reverse transcribed with random primers and analysed by RT-qPCR using ΔCT analysis method. (A) Comparison between the expression levels of the different lncRNAs tested. (B) Comparison between the mRNA expression levels.(C) and (D) show the differences of the transcript levels between the pairs Nup214/Ens87 and Zeb2/Zeb2NAT. The data was normalized to the mRNA Gapdh. The histograms depict mean and standard deviation of three independent experiments. The values are listed in Table 6.2 in the supplementary material.*

The results of the expression levels of the different lncRNAs analysed are compatible with the expression values found in the NONCODE database. The results show that Malat1 is the most expressed followed by Airn and Neat1 (see Table 3.2). The lncRNAs Ens91, Ens87 and Zeb2NAT are all expressed at very low levels in TAP cells. When comparing the pairs mRNA/NAT, Nup214/Ens87 and Zeb2/ Zeb2NAT, we observe that, in both cases, the mRNA has higher expression levels than its antisense transcript.

3.1.3. Measuring lncRNA/mRNA half-lives by metabolic labelling with 4-Thiouridine

After determining the expression levels of the selected transcripts in TAP cells, we proceed our research by establishing the methodology to monitor synthesis and lifetime of lncRNAs. We decided to measure the transcripts half-life using an assay based on the pulse labeling of nascent transcripts with the uridine analogue 4-thioridine (4sU) followed by purification of labeled nascent, pre-existing unlabeled and total cellular RNAs. 4sU has the great advantage of having minimal adverse effects on gene expression (Tani & Akimitsu 2012).

TAP cells were incubated with 4sU for 60min followed by purification of total and newly transcribed RNA. The separation of the labeled from unlabeled transcripts is based on the high affinity of 4sU to biotin. Following purification of the RNA subsets, cDNA was synthesized with random-primers and the specific transcripts to be analyzed were quantified by RT-qPCR. Values were normalized to Gapdh and half-lives calculated using the mathematical equation described previously in the methods section (Figure 2.3). This formula takes into account the time of labelling with 4sU, which was 60 min, and the proportion of the transcript of interest in the newly synthesised RNA fraction relative to the total cellular RNA fraction. Additionally, the calculation of the transcript's half-life takes into account the half-life of a second transcript, already known, that works as an internal normalizer (norm factor). The values obtain are therefore always relative to this other transcript. In this work, we used Gapdh as the

normaliser, considering the half-life of 7h described in the literature (Kudla et al. 2006). The results obtained are represented in Figure 3.3.

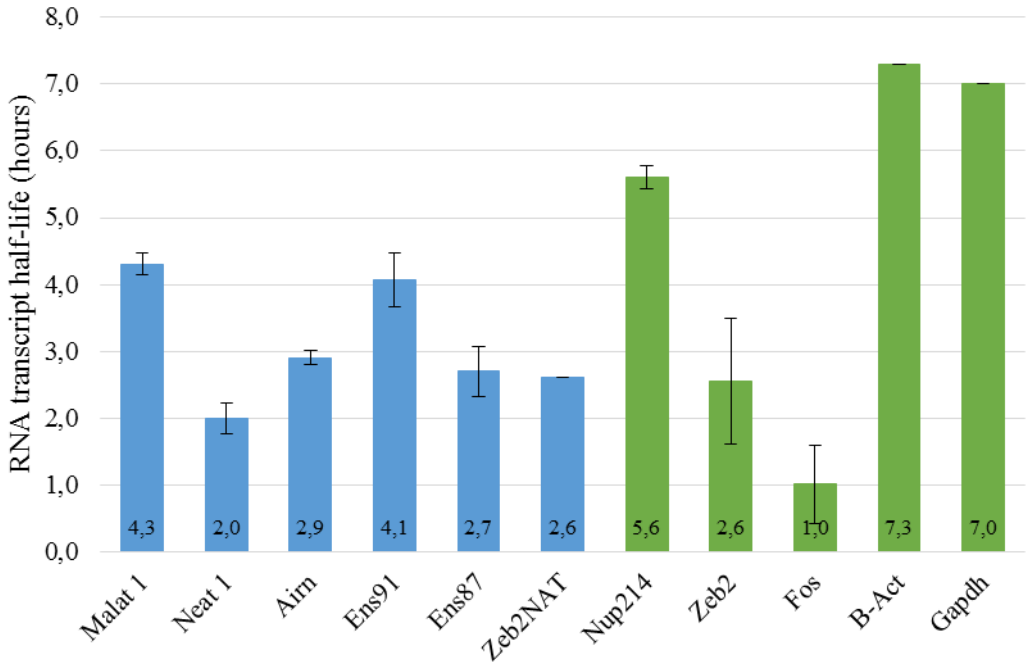


***Figure 3. 2** – LncRNA and mRNA half-lives determined based on newly transcribed RNA/total RNA ratios. TAP cells were incubated with 4sU for 60min. Total and newly transcribed RNA was purified, reverse transcribed with random primers and analysed by RT-qPCR. Values were normalized to Gapdh and half-lives calculated as described (see formula in Figure 2.3 in the methods section). LncRNAs are represented in the histogram in blue and mRNAs in green. The histograms depict mean and standard deviation of three independent experiments.*

The results indicate that Malat1 is the most stable lncRNA with a half-life of 4.3h, followed by Ens91 that has an half live of around 4h. The remaining lncRNAs analysed have half-lives of 2-3h. Nup214, which is an mRNA, has a half-life of 5,6h and Fos, known as a very short lived mRNA, has an half-life of 1h,which is in agreement with what was previously reported (Ross 1995). When comparing the pair Nup214/Ens87, mRNA and the corresponding NAT, we observe that the mRNA is more stable (5,6h) than the NAT (2,5h). However in the pair Zeb2/Zeb2NAT this difference in half-lives was not observed and both have similar and short half-lives (1,9h).

Even with the small number of targets that we have analysed, we could observe that lncRNAs are not typically unstable as a class, but rather show variation in their stability profiles in a manner similar to mRNAs, as demonstrated in other studies (Ross 1995; Tani et al. 2012). It has been suggested that the large variation in lncRNA stability is

consistent with their functional diversity and is likely a reflection of their complex post-transcription regulation. In fact, post-transcriptional regulation is particularly important for lncRNAs because, unlike protein-coding genes, they do not have any further translational and post-translational opportunities for regulation (Clark et al. 2012). LncRNA expression levels were not correlated with their stability, which could mean that lncRNAs below the expression cut-off will not be generally unstable and the same for the other way around, although, this possibilities cannot be completely discarded.

Since lncRNAs do not require translation to produce a functional gene product, it's possible that, especially those acting in the nucleus, can have a function almost immediately after transcription and so may not require a long half-life (Dinger et al. 2009). Moreover, unstable transcripts are very sensitive to changes in the level of transcription and respond quickly when transcription changes (Rabani et al. 2011).

Similarly to mRNA, highly stable lncRNAs may serve ''housekeeping'' roles and also, the fact that they are stable, may suggests that some lncRNAs have evolved to avoid degradation through diverse mechanisms, such as secondary structure and interactions with RNA-binding proteins (Clark et al. 2012).

3.2. In situ detection of individual lncRNA/mRNA molecules using LNA-modified oligonucleotides

Fluorescence microscopy is currently the best approach to study RNA localization at the single-cell level with high spatio-temporal resolution. A variety of recently developed FISH techniques make it now possible to track individual RNA molecules in single cells (reviewed by Itzkovitz & van Oudenaarden, 2011). Single-molecule RNA FISH protocols have been optimized using a set of about 50 short oligonucleotides, each one labelled with a single fluorophore (Lyubimova et al. 2013). Alternative protocols have been described based on the use of four to ten oligonucleotides about 50nt long and labels with 4 or 5 fluorescent dyes (Trcek et al. 2012b).

In our project, we are establishing a single-molecule RNA FISH (sm-RNA FISH) protocol in the cellular mouse model of T-ALL using short oligonucleotide probes with

Locked Nucleic Acid (LNA) technology. The LNA-based oligonucleotide probes have a greatly increased hybridization affinity and specificity that enable the detection of low-abundance RNAs with the ability to detect single nucleotide mismatches. This has been successfully used for *in situ* localization of microRNAs (Javelle and Timmermans, 2012) and has great potential for the detection of lncRNAs. However, the use of a single LNA probes per target will require signal amplification technology. We decided to use Tyramide Signal Amplification (TSA), which allows several probes to be hybridized and detected sequencially with different fluorophors (Shi et al. 2012).

3.2.1. Design of LNA-modified FISH probes

Our first goal with the use of sm-RNA FISH is to determine whether the lncRNAs remain restricted to the cognate gene locus or accumulate elsewhere in the nucleus. For this analysis and to establish the methodology in the laboratory we selected 2 pairs of mRNA/NAT: Nup214/Ens87 and Zeb2/Zeb2-NAT. As a positive control in this assay we chose to detect β-actin.

The criteria followed to choose the regions in the transcripts to target with the FISH probes were regions without overlap between the mRNA and the corresponding antisense lncRNA (NAT) and that were annotated as constitutive exons. To confirm the expression of the selected exons in TAP cells we designed PCR primers to detect those exons (Table 6.1 in the supplementary material) and performed RT-qPCR in total RNA samples. Some of the primer pairs were design in different exons to enable the detection of possible transcript isoforms (see scheme in Figure 3.3 A). The expression results obtained for the different regions are shown in Figure 3.3 B and C.

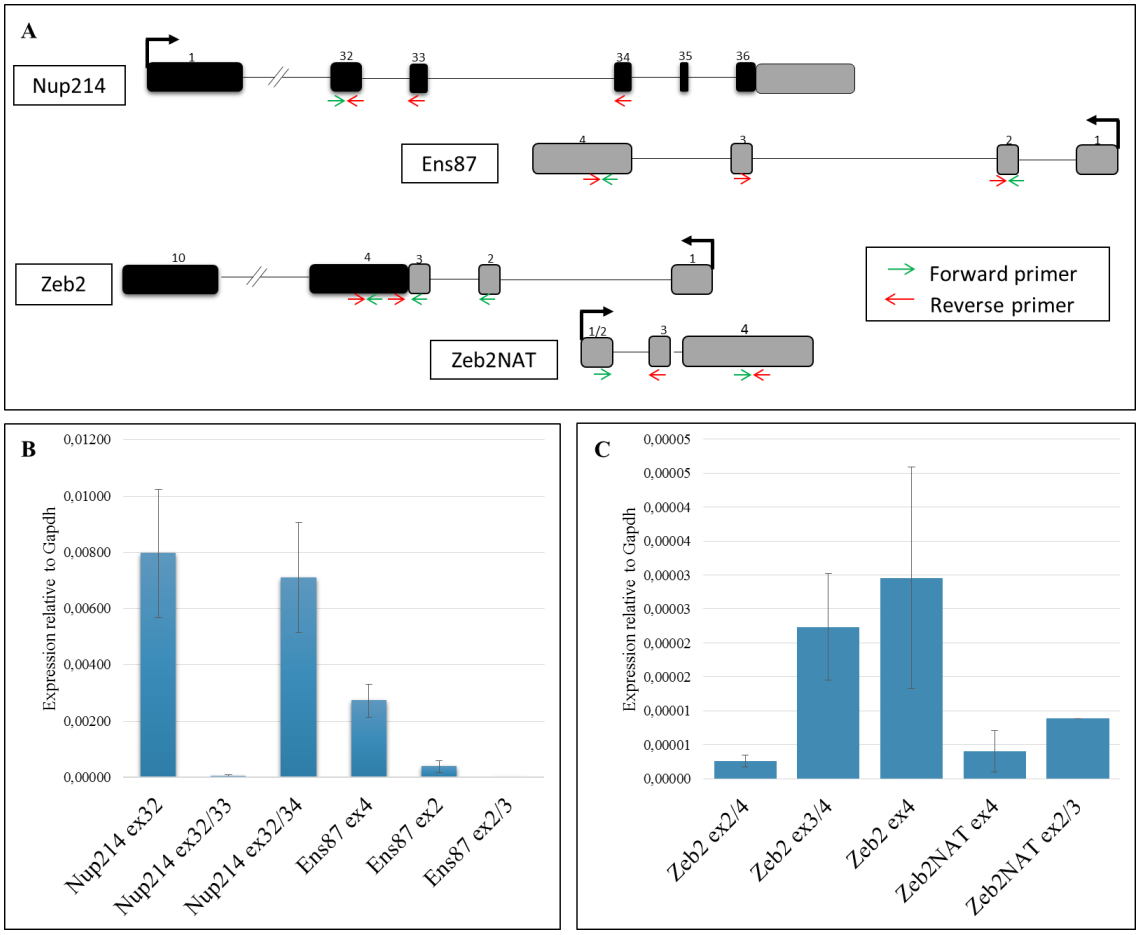


Figure 3. 3 *– Relative expression levels of selected exons present in the pairs mRNA/NAT (Nup214/Ens87 and Zeb2/Zeb2NAT) in TAP cells. Total RNA was purified from the cells, reverse transcribed with random primers and analysed by RT-qPCR using ΔCT analysis method. (A) Scheme showing the relative localization of the primers in the different transcripts. (B) Comparison between Nup214 and Ens87expression levels using primers for the indicated regions. (C) Comparison between Zeb2 and Zeb2NAT expression levels using primers for the indicated regions. The data was normalized to the Gapdh mRNA. The histograms depict mean and standard deviation of three independent experiments. The values are listed in Table 6.4 in the supplementary material.*

The analysis of the expression of Nup214 suggests that exon 33 is not constitutive, and in TAP cells is excluded, because amplification of exon 32/33 gives no expression and amplification of exon 32/34 gives expression levels similar to exon 32 alone.. This suggests that both exon 32 and 34 are present in Nup214 mRNA in TAP cells, but exon 33 is not. Analysis of the expression of Ens87 indicates that exon 3 is not constitutive and is excluded from the spliced Ens87 transcript. The analysis of the expression of Zeb2 gives lower levels when amplifying exon2/4 compared to exon 3/4 or exon4. This could be explained by lower efficiency of the amplification of the product exon 2/4 relative to the others or by the exclusion of exon 2 from Zeb2 mRNA. This data also suggests that both exons 3 and 4 are present in Zeb2 mRNA in TAP cells. Relative to

the expression analysis of Zeb2NAT we can conclude that exons 2, 3 and 4 are present in the spliced transcript.

From these results we decided design FISH probes complementary to exon 32 of Nup214 and to exon 2 of Ens87 to be sure that the regions detected do not overlap and the probes detect different transcripts. The same criteria was used for Zeb2 and Zeb2Nat where we choose to design FISH probes complementary to exon4 of Zeb2 and exon 4 of Zeb2NAT. Because LNA probes are highly specific and fail to hybridize in the presence of mismatches we decided sequence the chosen exons before designing the probes. For that purpose we amplified the selected exons from genomic DNA extracted from TAP cells using the primers listed in Table 6.5 in supplementary material. The amplified PCR products were gel purified and send for Sanger sequencing at StabVida (http://www.stabvida.com/pt/). The retrieved sequences are given in Table 6.6 of supplementary material. The identity was the sequences obtained were confirmed by a BLAST analysis in a genomic database.

We used Exiqon's Custom LNA™ mRNA/lncRNA Detection Probe design tool (www.exiqon.com) to design highly sensitive LNA™ probes specifically targeting our transcripts of interest. To feed the program we used the sequences obtained from our own Sanger sequencing of the target region. Taking into account the general probe design guidelines provided by Exiqon we chose oligonucleotide modifications at the 5'end. We decided to use Biotin modification for the probes that target mRNAs and Digoxigenin modification for the probes that target lncRNAs (see Figure 3.4).

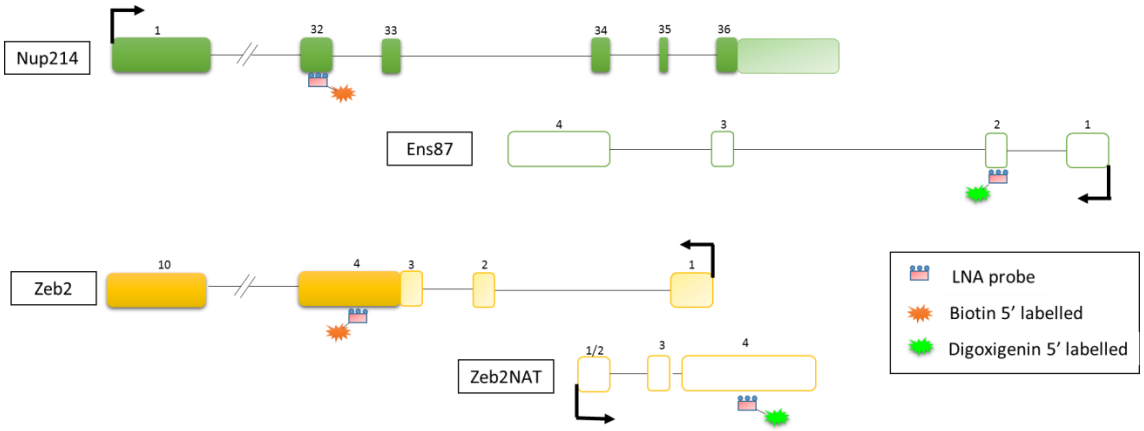


Figure 3. 4 *– Simple scheme representing the regions were the LNA probes bind to the mRNA of Nup214 or Zeb2 and its antisense lncRNAs Ens87 or Zeb2NAT, respectively.*

3.2.2. Detection of lncRNA/mRNA transcripts using single molecule LNA-FISH

LNA probes have been successfully used for in situ localization of microRNAs (Javelle and Timmermans, 2012) and they also could have great potential for the detection of lncRNAs. However, the use of a single LNA probe per target will require signal amplification technology. We decided to use Tyramide Signal Amplification (TSA) because it is highly sensitive and allows simultaneous detection of multiple probes by using sequentially different fluorophores for detection.

By implementing a single-molecule LNA-RNA FISH approach we can, not only determine the sub-cellular localisation of the target transcripts but also determine the absolute number of transcripts per cell. Fluorescent FISH signals were detected in a spinning disk confocal microscope. Confocal imaging via spinning disk involves scanning a field with laser light from a number of pinholes arranged in a pattern on a modified Nipkow disk. In the spinning disk confocal the detector is a CCD camera which can register the signal from a quarter million or million pixels simultaneously with a quantum efficiency of upwards of 90%. The high speed of image acquisition combined with the superior sensitivity of high end CCDs makes the spinning disk confocal a good system to analyse single-molecule FISH signals.

First, we evaluated the success of the experiment with a negative and a positive control. In the negative control we only add the hybridization buffer with no probes and in the positive control we hybridized the cell with β-Actin LNA probe labelled with digoxigenin at 5'end, a positive control tested and recommended by the company Exiqon. We also tried different concentrations of HRP-conjugate anti-digoxigenin to test the technic background (Figure 3.5).

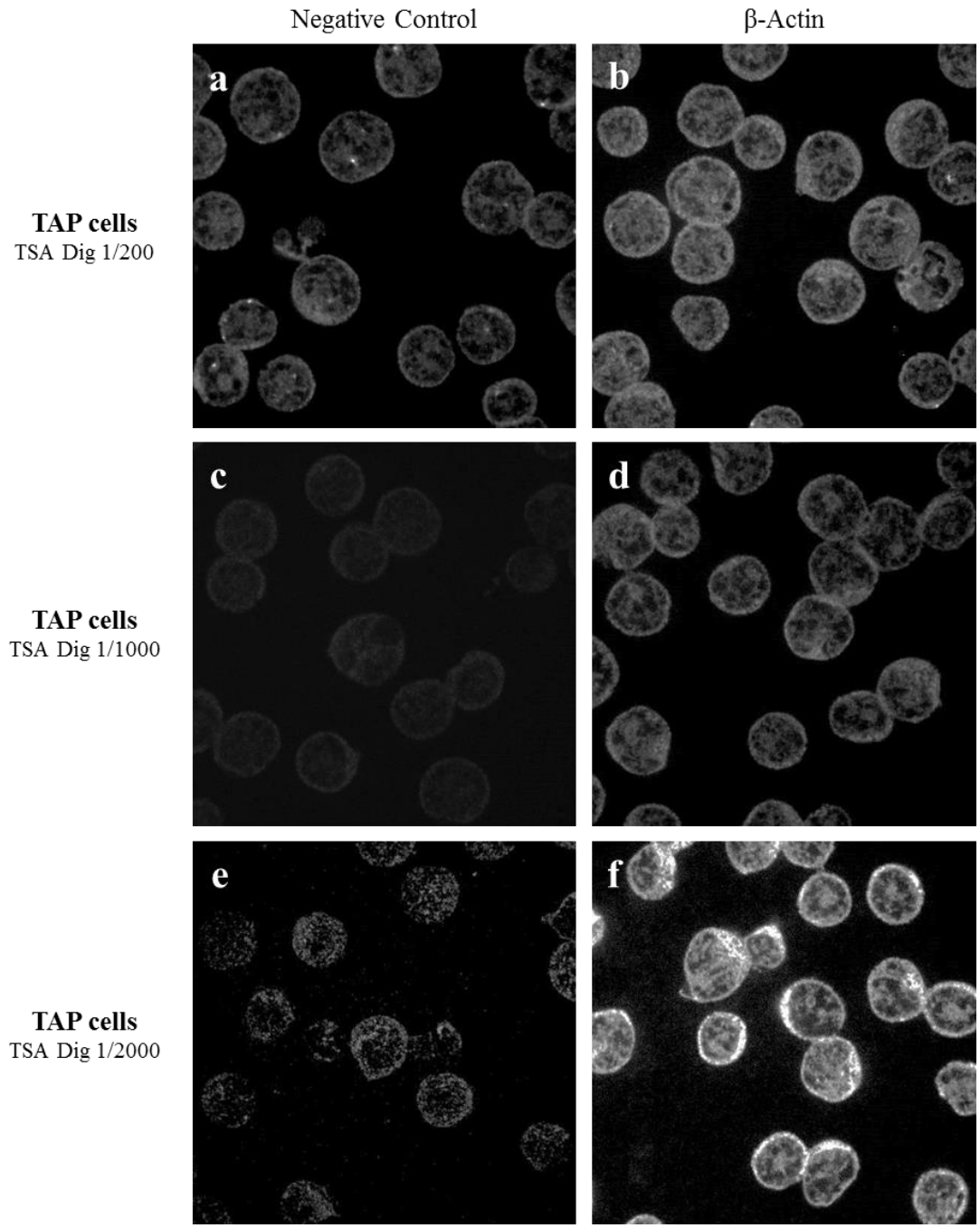


Figure 3. 5 – *LNA-RNA smFISH in TAP cells, using TSA amplification: Hybridization with β-Actin LNA probe. Was performed LNA-RNA smFISH using (a), (c) and (e) hybridization buffer with no probe and (b), (d) and (f) β-actin probe 5'digoxigenin labelled. The detection was made with TSA signal amplification using HRP-conjugate anti-digoxigenin at (a) and (b) 1:200 dilution; (c) and (d) 1.1000 dilution and; (e) and (f) 1:2000 dilution. The HRP-conjugate anti-digoxigenin was detect with Alexa Fluor 546 and imaged with spinning disk confocal microscope, objective x100 –oil, with C561 laser.*

In all the concentrations used for HRP-conjugate anti-digoxigenin we detect differences between the negative and the β-Actin positive control, with the signal of the

positive control always higher. We detected cytoplasmic staining for the β-Actin mRNA as expected with all the HRP-conjugated anti-digoxigenin dilutions tested. . Comparing the different dilutions there is differences in the ratio background/signal and though the positive signal of the 1:2000 is higher, the ratio tends to be lower. The β-Actin signal in the 1:1000 dilution is similar to 1:200 dilution, but the negative control has a lower background at 1:1000. It has been described that excess HRP- reagent can result in the formation of tyramide dimers that may be deposited in the cells and result in increased background. We think that this dimerization of tyramide is the cause of the signals observed in the negative control at 1:200 dilution. From the comparison of the signal to noise ratio of the different dilutions we concluded that 1:1000 to 1:2000 are appropriate and should be used with our custom designed LNA probes.

After this initial optimisation steps and confident that the positive control with the β-Actin probe is working we performed the LNA-RNA sm-FISH for the other probes. We started by the probes that target the lncRNAs Ens87 and Zeb2NAT and that are labelled with Digoxigenin at the 5'end. . We decided to use the HRP-conjugate anti-digoxigenin at 1:1000 dilution. In parallel with the probes we included the negative control, hybridization without probe. The results obtained are shown in Figure 3.6.

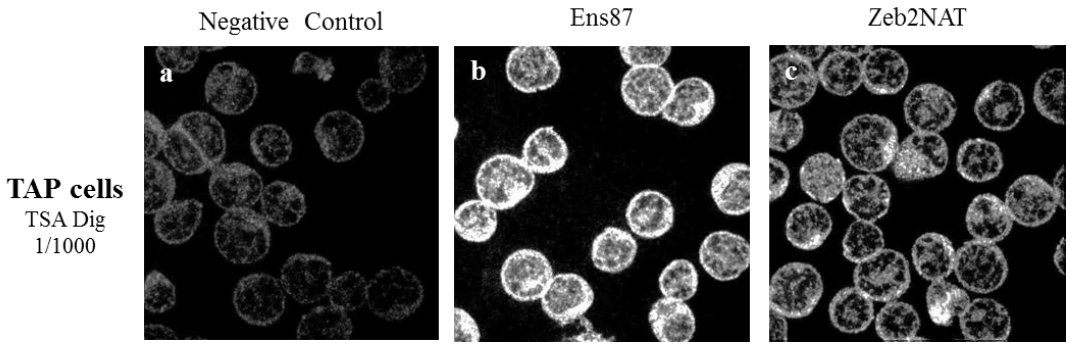


***Figure 3. 6** – LNA-RNA smFISH in TAP cells, using TSA amplification: Hybridization with Ens87 and Zeb2NAT LNA probes. Was performed LNA-RNA smFISH using (a) hybridization buffer with no probe; (b) Ens87 probe 5'digoxigenin labelled and; (c) Zeb2NAT probe 5'digoxigenin labelled. The detection was made with TSA signal amplification using HRP-conjugate anti-digoxigenin at (a), (b) and (c) 1:1000 dilution. The HRP-conjugate anti-digoxigenin was detect with Alexa Fluor 546 and imaged with spinning disk confocal microscope, objective x100 – oil, with C561 laser.*

Overall the hybridisation with the Ens87 probe tends to give a higher signal than the Zeb2NAT probe, which is in agreement with the RT-qPCR expression analysis (see Fig 3.1 A).

It is important to refer that, although these results are encouraging they are still preliminary. More experiments with different concentrations of probe and HRP-conjugate anti-digoxigenin should be done to try to improve the signal to noise ratios. Additional control to validate the specificity of the signals includes treatment with RNase A before the hybridization, which degrade the RNA molecules and should eliminate the signal completely. Another control to the specificity of the hybridization would be to repeat the hybridisation in different cell types that may not expressed these lncRNAs or express them at different levels.

Finally, we tested the LNA probes that target the mRNAs Nup214 and Zeb2, and are labelled with biotin at the 5'end. Also in this case we included the negative control, hybridization without probe. The results obtained with this panel of probes are shown in Figure 3.7.

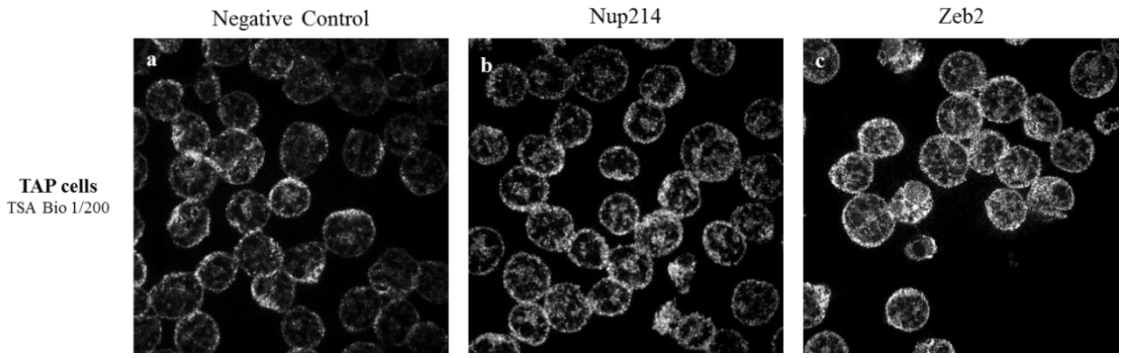


***Figure 3. 7** – LNA-RNA smFISH in TAP cells, using TSA amplification: Hybridization with Nup214 and Zeb2 LNA probes. Was performed LNA-RNA smFISH using (a) hybridization buffer with no probe; (b) Nup214 probe 5'biotin labelled and; (c) Zeb2 probe 5'biotin labelled. The detection was made with TSA signal amplification using HRP-conjugate streptavidin at 1:200 dilution. The HRP-conjugate streptavidin was detect with Alexa Fluor 546 and imaged with spinning disk confocal microscope, objective x100 –oil, with C561 laser.*

In this case our first observation was a much higher background in the negative hybridization. A strong cytoplasmic staining was detected which could correspond to the detection of endogenous biotin. It is known that some cells contain high levels of endogenous biotin, and this seems to be the case of TAP cells. It is therefore important to quench the activity of endogenous biotin which can be achieved with a commercially

available Biotin-blocking kit. Only after this optimization step the signal present in the Nup214 and Zeb2 hybridizations can be appreciated.

Although the results from the previous experiments still need optimisation we decided to try a double hybridisation experiment. In this case, we used the probes for the lncRNAs (Ens87 or Zeb2NAT) labelled at 5'end with digoxigenin and the probes for the mRNAs (Nup214 or Zeb2) labelled at 5'end with biotin. We hybridized the LNA probe pairs mRNA/lncRNA (Nup214/Ens87 or Zeb2/Zeb2NAT) together and then detect them consecutively. Firstly the mRNA probe labelled with biotin was detected with HRP-conjugate streptavidin and Alexa Fluor 546 tyramide (green signal) and afterwords the HRP was deactivated by incubation with $1\%H_20_2$ in PBS. Then the lncRNA labelled with digoxigenin was detected with HRP-conjugated anti digoxigenin and Alexa Fluor 647 tyramide (red signal). In parallel a negative control without probe was subjected to the same scheme of detection See results in Figure 3.8

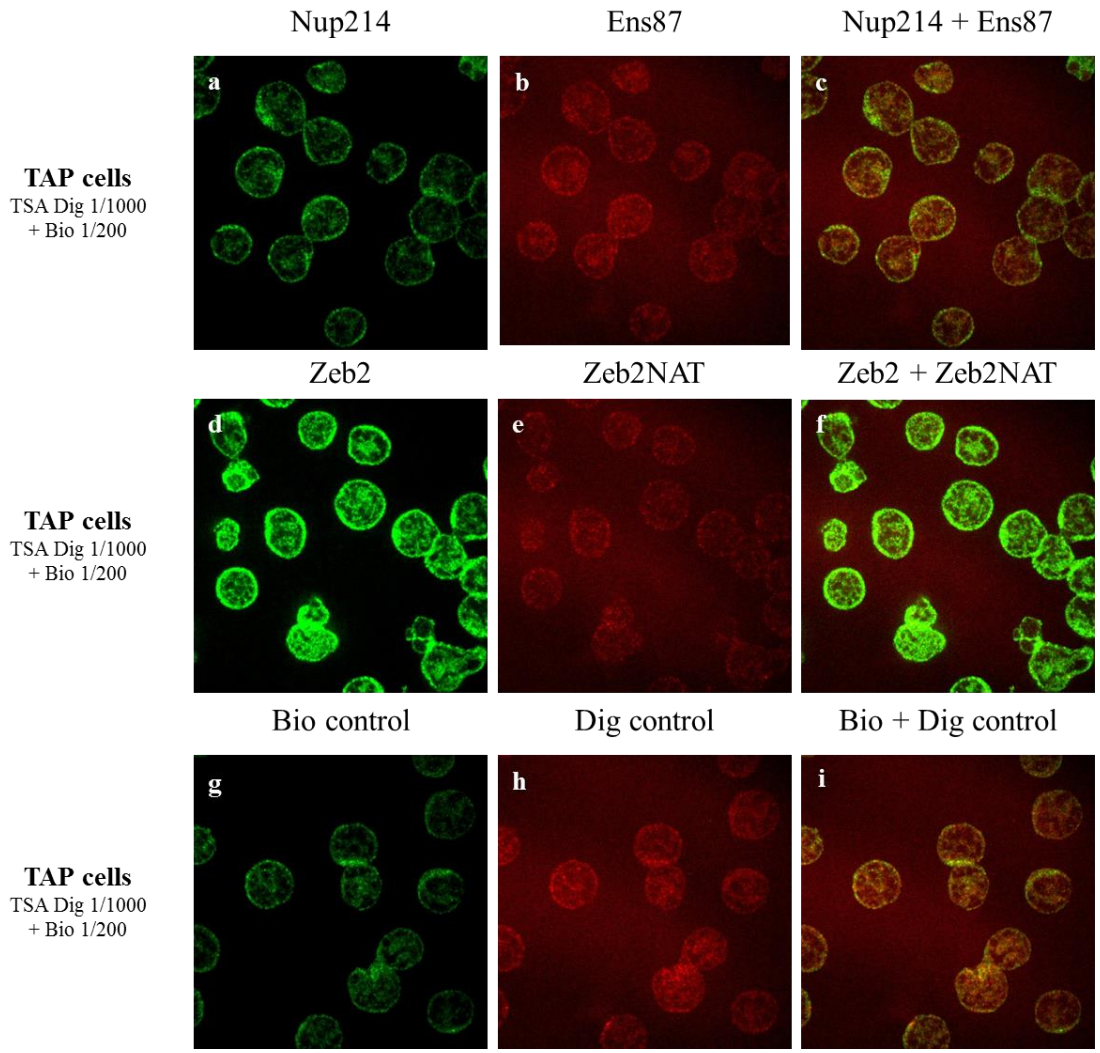


Figure 3. 8 *– LNA-RNA smFISH in TAP cells, using TSA amplification: Hybridization with Nup214/Ens87 and Zeb2/Zeb2NAT LNA probes LNA-RNA smFISH was performed using (g), (h) and (i) hybridization buffer with no probe; (a), (b) and (c) Ens87 probe 5'digoxigenin labelled and Nup214 probe 5'biotin labelled and; (d), (e) and (f) Zeb2NAT probe 5'digoxigenin labelled and Zeb2 probe 5'biotin labelled. The detection was performed with TSA signal amplification using HRP-conjugate anti-digoxigenin at 1:1000 dilution and HRP-conjugate streptavidin at 1:200 dilution. The HRP-conjugate anti-digoxigenin was detect with Alexa Fluor 546 and imaged with spinning disk confocal microscope, objective x100 –oil, with C561s laser. The HRP-conjugate streptavidin was detect with Alexa Fluor 647 and imaged with spinning disk confocal microscope, objective x100 –oil, with C640s laser.*

The signal of the digoxigenin was lost in both cases and the staining with Ens87 and Zeb2NAT was similar to the negative control. The staining with Nup214 was also similar to the negative control. Since the RT-qPCR indicate that the mRNA are more expressed than the lncRNAs, it's good to try the detection first with HRP-conjugate streptavidin and then with HRP-conjugate anti-digoxigenin to avoid losing the signal.

Several studies indicate that the majority of lncRNAs are located in nucleus, which is consistent with the major function of lncRNA in epigenetic regulation of gene expression. There is still a population of lncRNAs that are enriched in cytoplasm and have important functions including translational regulation (Derrien et al. 2012).

Finding a lncRNA primarily in the nucleus near its site of transcription may suggest that it regulates transcription of a proximal gene (Cabili et al. 2015). Since genomic locations situated in different chromosomes can be in close spatial proximity within the nucleus, this proximity would explain the observations of both cis- and trans-mediated regulatory effects of lncRNAs. This model could explain how lncRNAs, which are generally of lower abundance relative to mRNAs, can reliably identify their target genes by searching in spatial proximity near their transcription locus (Rinn & Guttman 2014). On the other hand, the lncRNAs that have been identified in the cytoplasm often show sequence complementarity with transcripts that originate from either the same chromosomal locus or independent loci. It could suggest that upon recognition of the target by base pairing, they can modulate translational control (Fatica & Bozzoni 2014). It has been suggested that spliced lncRNAs, compared with such un-spliced as single exon transcripts, intergenic and cis-antisense RNAs are more stable than those derived from introns and with the sub-cellular localization analysis there is indication that the location of lncRNAs is widespread in cell, where the nuclear-localized lncRNAs are more likely to be unstable (Cao 2014).

In our work we couldn't define weather the lncRNAs or the mRNAs tested localized in the nucleus or cytoplasm but we expect that with further optimizations of the protocols we can achieve that outcome. The correlation of Nup214 with its antisense transcript, Ens87, has not yet been described and, in this present study, we investigate a possible connection between the mRNA and the NAT. The relation between Zeb2 and its NAT (Zeb2NAT) has been highlighted and there is evidence that Zeb2NAT interferes with splicing of ZEB2 and that the expression of the NAT indirectly enables expression of ZEB2 protein (Beltran et al. 2008). In this work we didn't observe a strong relation with T-ALL but, in another study, performed with a different subtype of human T-ALL, where the cells are in a more immature stage, the authors, showed an evidence for an oncogenic driver role for ZEB2 through gain-of-function mechanisms

where they hypothesized that is possible that the NAT could be involved in the overexpression of ZEB2 (Goossens et al. 2015).

3.3. Genome-wide analysis of lncRNAs and protein-coding genes in T-cell leukemic transformation

High-throughput RNA sequencing (RNA-Seq) has provided useful information for lncRNA identification. Nevertheless, the short reads of current technologies still limit the ability to accurately delineate full-length transcriptional units, particularly those of lncRNAs, which typically are expressed at low levels (Guttman et al. 2009).

To study genome-wide the differences between diverse stages of T-ALL, strand specific RNA-seq from total RNA samples was performed by our collaborators. In our analysis, we compare 4 different samples: control sample that correspond to RAG-/- thymocytes; TLX3.1 sample that correspond to the cell culture of RAG-/- thymocytes transduced with the oncogene TLX3; sample from the tumors formed in the spleen after injection in the mouse with TLX3.1 cells and; TAP sample that correspond to the stable and immortalized cell culture derived from the T-ALL tumors.

RNA seq reads were align to the reference mouse genome (mmusculus.9).using TopHat and subsequent analysis of differential expression of protein coding and non-coding genes between the four samples described above was performed by using Cuffdiff. To map the aligned reads to annotated coding and non-coding genes were used the UCSC and NONODE databases, respectively.

The results obtain from this analysis are represented in Figure 3.9.

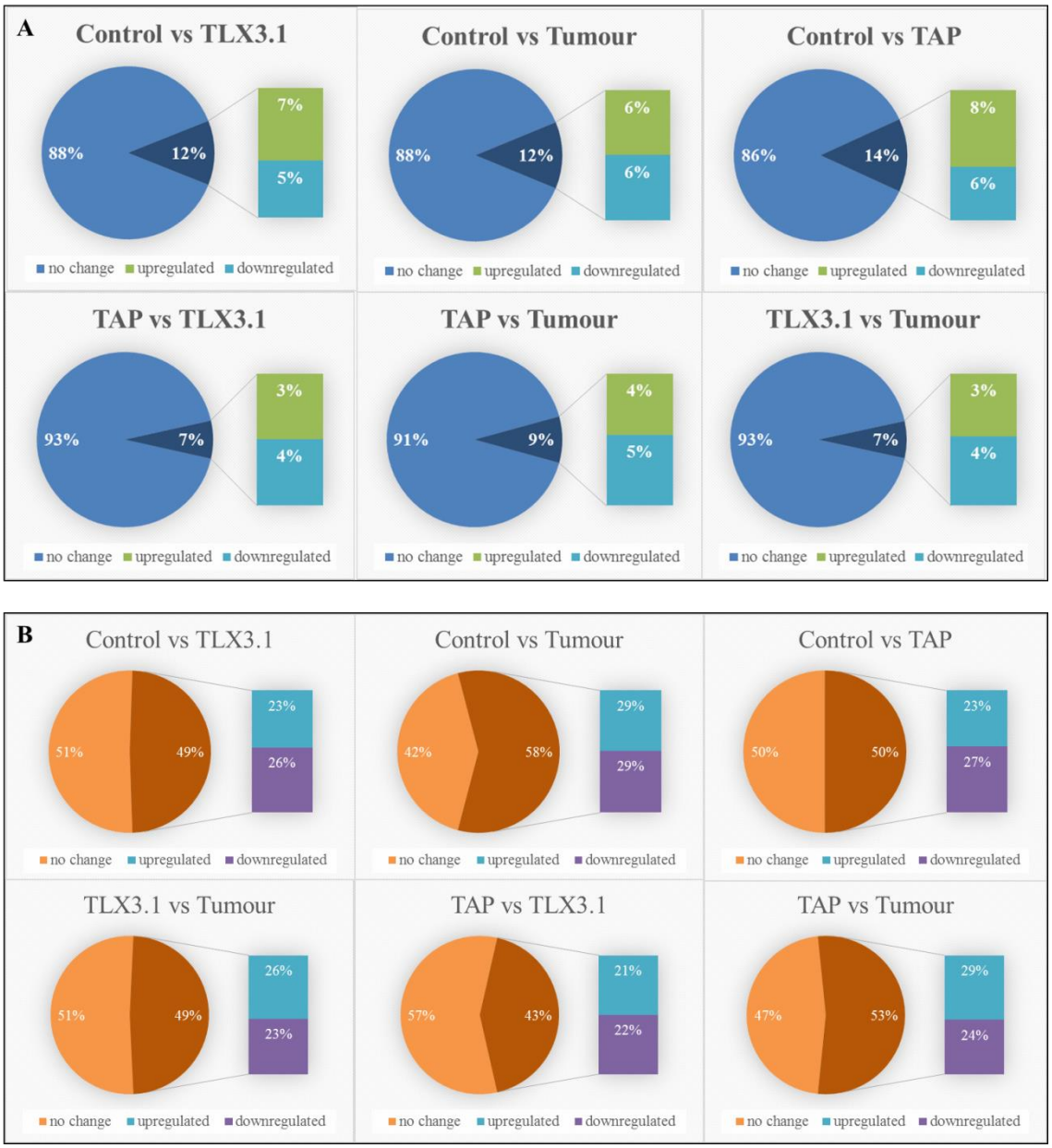


Figure 3. 9 *– Pie charts of the differential expression of lncRNAs and protein-coding genes. (A) Differential expression of lncRNAs between different samples. (B) Differential expression of protein-coding genes between different samples.In both cases the reads were align to the genome of reference mmusculus.9 with the TopHat tool and the differential expression analysis was obtain by Cuffdiff tool using the NONCODE data base for lncRNAs and the UCSC database for the protin-coding genes. The transcript was considered differential expressed if p-value<0.05.*

We could observe that there are more transcripts differentially expressed in the analysis of the proteins-coding genes corroborating the evidences that the lncRNAs, with less differential expressed transcripts, are more cell- tissue-specific (Amaral & Mattick 2008). We also observe a bigger difference of differential expressed transcripts

when comparing the samples against the control rather the comparison between each sample indicating a close relation that the samples have.

We also did a close analysis to the lncRNAs and mRNAs tested in the previously experiments with the results in Figure3.11.

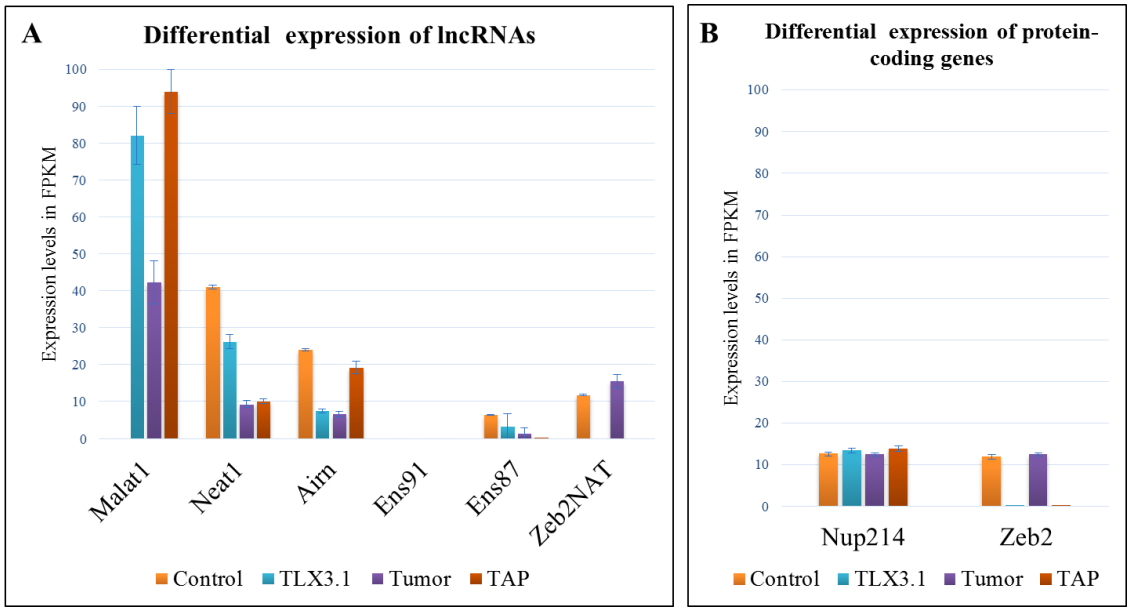


Figure 3. 10 *– Differential expression of selected lncRNAs and protein-coding genes. (**A**) Differential expression of lncRNAs between different samples. (**B**) Differential expression of protein-coding genes between different samples.In both cases the reads were align to the genome of reference mmusculus.9 with the TopHat tool and the differential expression analysis was obtain by Cuffdiff tool using the NONCODE data base for lncRNAs and the UCSC database for the protein-coding genes. The transcript was considered differential expressed if p-value<0.05.*

Based on this analysis Malat1 is not expressed in the control sample and has high values in TAP sample which is in agreement with the results from expression obtained by RT-qPCR. Neat1 and Airn have decreased expression in relation to control and they also corroborate the RT-qPCR results, with Airn more expressed in TAP sample than Neat1. Ens91 expression is non-existent and the RT-qPCR also showed a low expression. Ens87 has a decreased expression along the sample having a very low expression in TAP sample. Zeb2NAT seems to be expressed in the control and the tumor sample and, has expected by the RT-qPCR, is very low expressed in TAP samples. When looking to the mRNAs, Nup214 doesn't change significantly between the control and the T-ALL samples. Zeb2 has the same behaviour as Zeb2NAT corroborating their relationship (Beltran et al. 2008).

We also check two new pairs of mRNA/NAT: Fadd/Faddos and Daxx/BC051242. Fadd and Daxx are mRNAs that were part of the research done initially when we were selecting the lncRNAs to test experimentally (see Table 3.1). Fadd is an adaptor protein that bridges members of the tumor necrosis factor receptor superfamily and it's involved in apoptosis (Kim et al. 1996). Daxx is a protein that has been implicated in many nuclear processes including transcription and cell cycle regulation (Salomoni and Khelifi 2006). The results are represented in Figure 3.11.

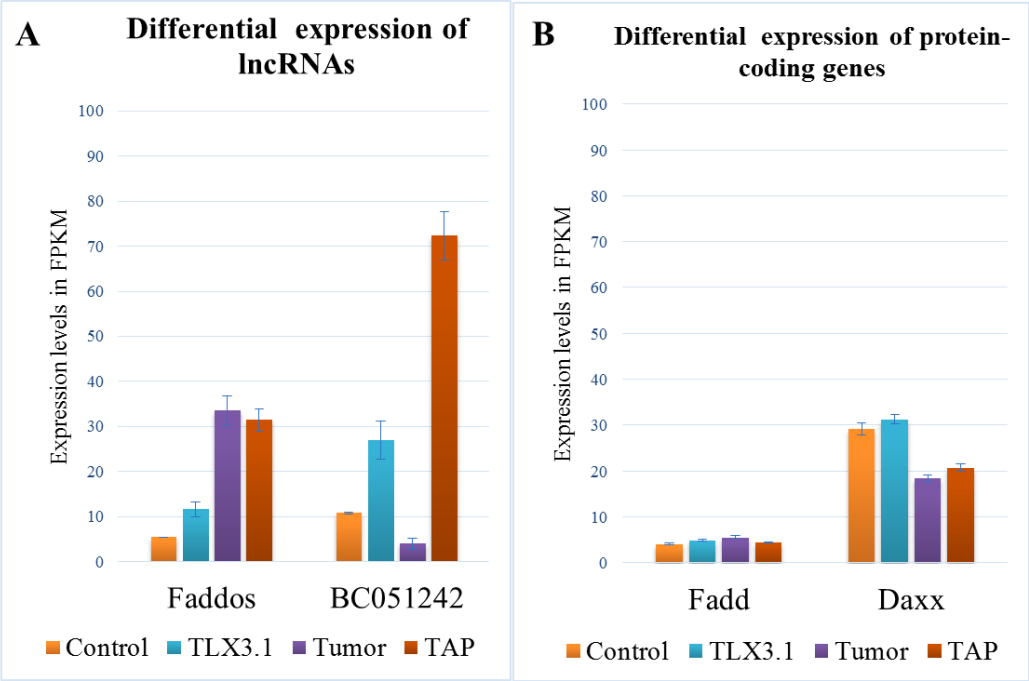


Figure 3. 11 *– Differential expression of new selected pairsmRNA/NAT. (**A**) Differential expression of lncRNAs between different samples. (**B**) Differential expression of protein-coding genes between different samples.In both cases the reads were align to the genome of reference mmusculus.9 with the TopHat tool and the differential expression analysis was obtain by Cuffdiff tool using the NONCODE data base for lncRNAs and the UCSC database for the protein-coding genes. The transcript was considered differential expressed if p-value<0.05.*

Faddos and BC051242 seem both to be upregulated in TAP cells compared to normal control. Fadd mRNA, like Nup214, also does not change and Daxx has a similar behaviour than its antisense but with less evident changes. The pair Daxx/BC051242 could be a similar example of mRNA regulation like has been suggested for Zeb2/Zeb2NAT.

A large amount of sequence data have been generated in the last few years and this already allowed the identification of thousands of lncRNAs expressed in different stages of hematopoietic development. Nevertheless, most studies merely identify lncRNAs without explaining how these molecules act within the cell. This has further generated the need to develop new experimental tools to identify and analyse the mechanisms of action of lncRNA. Targeted deletion studies in primary cells, or animal models when applicable, will be necessary to assess the specific functions of lncRNAs involved in normal and malignant hematopoiesis (Morlando et al. 2015). We expect that with further analysis of the described datasets we will be able to identify a subset of lncRNAs with a role in leukemogenesis. The tool developed in this work will be extremely useful to determine the mechanism of action of these lncRNAs and to help elucidating their role in the malignant transformation of T-cells.

4. Conclusions and Future Work

There is a long road ahead in the research of lncRNAs that will require more comprehensive transcriptome analyses and transcript assemblies, additional data on levels of expression in a specific cell- and tissue-type, and more data on cellular localization and interacting partners. We suggest that the annotation of genomes should be more transcript centric rather than gene centric, because it has become clear that the majority of the transcripts in cells are more than mere intermediates between the hereditary information encoded in DNA. In fact, numerous transcripts that may not be translated at all and are involved in critical biological functions (Bussotti et al. 2013; Mattick & Rinn 2015).

Nowadays, cancer therapy has many difficulties, for example, specific targeting of cancer cells without interfering with normal tissue function, or specific delivery of antitumor drugs may be challenging to the treatment of cancer. LncRNAs could offer a number of advantages, both as diagnostic and prognostic markers but also as novel specific therapeutic targets. But first, it is necessary to have a detailed knowledge about the tumor-specific lncRNA function and its requirement for essential cancer cell properties. LncRNAs are interesting targets in cancer therapy and especially their cancer- and tissue-specific expression could be a major advantage over other therapeutic options (Gutschner & Diederichs 2012).

In this particular work, the main objectives were the implementation of efficient experimental techniques that allow functional analysis of the candidate targets for therapy that will be provided after the bioinformatics data analysis of different subsets of T-ALL. The knowledge of the stability and geographical information of lncRNAs can be essential for its further efficient targeting in the RNA therapeutic context.

The main goal of this present study is the development and implementation of laboratory techniques to measure transcripts half-life by using metabolic pulse labelling with 4sU and a novel approach for sub-cellular localization of single RNA molecules using a FISH protocol with a single LNA probe per RNA molecule.

In the future work, our major objective is to do a more detailed bioinformatics analysis of the data that we have with the aim of finding potentially relevant lncRNA in the process of T-cell malignant transformation that could be used as candidates for therapeutic targets for T-ALL. With a list of relevant lncRNAs we will precede for their characterization using the experimental techniques developed in this study along other in vitro and in vivo assays.

5. References

Aifantis, I., Raetz, E. & Buonamici, S., 2008. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nature reviews. Immunology*, 8(5), pp.380–390.

Amaral, P.P. & Mattick, J.S., 2008. Noncoding RNA in development. *Mammalian Genome*, 19(7-8), pp.454–492.

Astakhova, K., 2014. Toward Non-Enzymatic Ultrasensitive Identification of Single Nucleotide Polymorphisms by Optical Methods. *Chemosensors*, 2(3), pp.193–206. Available at: http://www.mdpi.com/2227-9040/2/3/193/.

Atkinson, S.R., Marguerat, S. & Bähler, J., 2012. Exploring long non-coding RNAs through sequencing. *Seminars in Cell and Developmental Biology*, 23(2), pp.200–205. Available at: http://dx.doi.org/10.1016/j.semcdb.2011.12.003.

Beltran, M. et al., 2008. A natural antisense transcript regulates Zeb2 / Sip1 gene expression during Snail1-induced epithelial – mesenchymal transition. , pp.756–769.

Bobrow, M.N. et al., 1992. The use of catalyzed reporter deposition as a means of signal amplification in a variety of formats. *Journal of Immunological Methods*, 150(1-2), pp.145–149.

Brannan, C.I. et al., 1990. The product of the H19 gene may function as an RNA. *Molecular and cellular biology*, 10(1), pp.28–36.

Bussotti, G., Notredame, C. & Enright, A.J., 2013. Detecting and comparing non-coding RNAs in the high-throughput era. *International Journal of Molecular Sciences*, 14(8), pp.15423–15458.

Cabili, M.N. et al., 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biology*, 16(1), pp.1–16. Available at: http://genomebiology.com/2015/16/1/20.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al., 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes and Development*, 25, pp.1915–27.

Cao, J., 2014. The functional role of long non-coding RNAs and epigenetics. *Biol Proced Online*, 16(1), p.11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25276098.

Cauwelier, B. et al., 2006. Molecular cytogenetic study of 126 unselected T-ALL cases reveals high incidence of TCRbeta locus rearrangements and putative new T-cell oncogenes. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 20(7), pp.1238–1244.

Cheetham, S.W. et al., 2013. Long noncoding RNAs and the genetics of cancer. *British journal of cancer*, 108(12), pp.2419–25. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3694235&tool=pmcentrez&rendertype=abstract.

Chen, G. et al., 2013. LncRNADisease: A database for long-non-coding RNA-associated diseases. *Nucleic Acids Research*, 41(D1), pp.983–986.

Cheng, W., Zhang, Z. & Wang, J., 2013. Long noncoding RNAs: New players in prostate cancer. *Cancer Letters*, 339(1), pp.8–14. Available at: http://dx.doi.org/10.1016/j.canlet.2013.07.008.

Clark, M.B. et al., 2012. Genome-wide analysis of long noncoding RNA stability. *Genome Research*, 22, pp.885–898.

Clemson, C.M. et al., 1996. XIST RNA paints the inactive X chromosome at interphase: Evidence for a novel RNA involved in nuclear/chromosome structure. *Journal of Cell Biology*, 132(3), pp.259–275.

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

Dadi, S. et al., 2012. TLX Homeodomain Oncogenes Mediate T Cell Maturation Arrest in T-ALL via Interaction with ETS1 and Suppression of TCRα Gene Expression. *Cancer Cell*, 21, pp.563–576.

Davis, M.E. et al., 2010. Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature*, 464(7291), pp.1067–1070. Available at: http://dx.doi.org/10.1038/nature08956.

Derrien, T. et al., 2012. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*, 22, pp.1775–1789.

Dinger, M.E. et al., 2009. Pervasive transcription of the eukaryotic genome: Functional indices and conceptual implications. *Briefings in Functional Genomics and Proteomics*, 8(6), pp.407–423.

Dölken, L. et al., 2008. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. , pp.1959–1972.

Faghihi, M.A. & Wahlestedt, C., 2009. Regulatory roles of natural antisense transcripts. *Nature reviews. Molecular cell biology*, 10(9), pp.637–643. Available at: http://dx.doi.org/10.1038/nrm2738.

Fatica, A. & Bozzoni, I., 2014. Long non-coding RNAs: new players in cell differentiation and development. *Nature reviews. Genetics*, 15(1), pp.7–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24296535.

Friedel, C.C. et al., 2009. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Research*, 37(17).

Geisler, S. & Coller, J., 2013. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nature reviews. Molecular cell biology*, 14(11), pp.699–712. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24105322.

Gezer, U. et al., 2014. Long non-coding RNAs with low expression levels in cells are enriched in secreted exosomes. *Cell biology international*, 9999, pp.1–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24798520.

Goossens, S. et al., 2015. ZEB2 drives immature T-cell lymphoblastic leukaemia development via enhanced tumour-initiating potential and IL-7 receptor signalling. *Nature communications*, 6, p.5794. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25565005.

Graux, C. et al., 2006. Cytogenetics and molecular genetics of T-cell acute lymphoblastic leukemia: from thymocyte to lymphoblast. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 20(May), pp.1496–1510.

Graux, C. et al., 2004. Fusion of NUP214 to ABL1 on amplified episomes in T-cell acute lymphoblastic leukemia. *Nature genetics*, 36(10), pp.1084–1089.

Graux, C. et al., 2009. Heterogeneous patterns of amplification of the NUP214-ABL1 fusion gene in T-cell acute lymphoblastic leukemia. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 23(1), pp.125–133.

Gutschner, T. et al., 2013. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Research*, 73(3), pp.1180–1189.

Gutschner, T. & Diederichs, S., 2012. The Hallmarks of Cancer: A long non-coding RNA point of view. *RNA Biology*, 9(6), pp.703–719.

Guttman, M. et al., 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, 458(7235), pp.223–227. Available at: http://dx.doi.org/10.1038/nature07672.

Hatano, M. et al., 1991. Deregulation of a homeobox gene, HOX11, by the t(10;14) in T cell leukemia. *Science (New York, N.Y.)*, 253(5015), pp.79–82.

He, Y. et al., 2014. Long noncoding RNAs: Novel insights into hepatocelluar carcinoma. *Cancer Letters*, 344(1), pp.20–27. Available at: http://dx.doi.org/10.1016/j.canlet.2013.10.021.

Holland, P.W.H., Booth, H.A.F. & Bruford, E. a, 2007. Classification and nomenclature of all human homeobox genes. *BMC biology*, 5, p.47.

Hu, G. et al., 2013. Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nature immunology*, 14(11), pp.1190–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3805781&tool=pmcentrez&rendertype=abstract.

Ip, J.Y. & Nakagawa, S., 2012. Long non-coding RNAs in nuclear bodies. *Development, growth & differentiation*, 54(1), pp.44–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22070123.

Itzkovitz, S. & van Oudenaarden, A., 2011. Validating transcripts with probes and imaging technology. *Nature methods*, 8(4 Suppl), pp.S12–S19.

Kalender Atak, Z. et al., 2013. Comprehensive Analysis of Transcriptome Variation Uncovers Known and Novel Driver Events in T-Cell Acute Lymphoblastic Leukemia. *PLoS Genetics*, 9(12).

Kapranov, P. et al., 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science (New York, N.Y.)*, 316(5830), pp.1484–1488.

De Keersmaecker, K. et al., 2014. NUP214-ABL1-mediated cell proliferation in T-cell acute lymphoblastic leukemia is dependent on the LCK kinase and various interacting proteins. *Haematologica*, 99(1), pp.85–93.

King, B., Ntziachristos, P. & Aifantis, I., 2012. Hijacking T Cell Differentiation: New Insights in TLX Function in T-ALL. *Cancer Cell*, 21(4), pp.453–455. Available at: http://dx.doi.org/10.1016/j.ccr.2012.03.026.

Kudla, G. et al., 2006. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biology*, 4(6), pp.0933–0942.

Latos, P. a et al., 2012a. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science (New York, N.Y.)*, 338(6113), pp.1469–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23239737.

Latos, P. a et al., 2012b. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science (New York, N.Y.)*, 338(6113), pp.1469–72.

Lee, J., 2012. Epigenetic Regulation by Long Noncoding RNAs. *Science*, 338(6113), pp.1435–1439. Available at: citeulike-article-id:11843007\nhttp://dx.doi.org/10.1126/science.1231776.

Li, C.H. & Chen, Y., 2013. Targeting long non-coding RNAs in cancers: Progress and prospects. *International Journal of Biochemistry and Cell Biology*, 45(8), pp.1895–1910. Available at: http://dx.doi.org/10.1016/j.biocel.2013.05.030.

Lyubimova, A. et al., 2013. Single-molecule mRNA detection and counting in mammalian tissue. *Nature protocols*, 8(9), pp.1743–58. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23949380.

Mao, Y.S. et al., 2011. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nature cell biology*, 13(1), pp.95–101. Available at: http://dx.doi.org/10.1038/ncb2140.

Mattick, J.S. & Rinn, J.L., 2015. Discovery and annotation of long noncoding RNAs. *Nature Publishing Group*, 22(1), pp.5–7. Available at: http://dx.doi.org/10.1038/nsmb.2942.

Morlando, M., Ballarino, M. & Fatica, A., 2015. Long Non-Coding RNAs: New Players in Hematopoiesis and Leukemia. *Frontiers in Medicine*, 2(April), pp.1–5. Available at: http://journal.frontiersin.org/article/10.3389/fmed.2015.00023/abstract.

Nguyen, Q. & Caninci, P., 2015. Expression Specificity of Disease-Associated lncRNAs: Toward Personalized Medicine. *Springer International Publishing Switzerland*, 351, pp.139–157.

Pang, K.C. et al., 2009. Genome-wide identification of long noncoding RNAs in CD8+ T cells. *Journal of immunology (Baltimore, Md. : 1950)*, 182(12), pp.7738–7748.

Penny, G.D. et al., 1996. Requirement for Xist in X chromosome inactivation. *Nature*, 379(6561), pp.131–137.

Prensner, J.R. et al., 2011. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nature biotechnology*, 29(8), pp.742–749. Available at: http://dx.doi.org/10.1038/nbt.1914.

Rabani, M. et al., 2011. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature biotechnology*, 29(5), pp.436–442. Available at: http://dx.doi.org/10.1038/nbt.1861.

Rädle, B. et al., 2013. Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. *Journal of visualized experiments : JoVE*, (August), pp.1–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23963265.

Ranzani, V. et al., 2015. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nature Immunology*, 16(3). Available at: http://www.nature.com/doifinder/10.1038/ni.3093.

Rinn, J. & Guttman, M., 2014. RNA and dynamic nuclear organization. *Science*, 1240, pp.10–12.

Rinn, J.L. & Chang, H.Y., 2012. Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry*, 81, pp.145–166.

Ross, J., 1995. mRNA stability in mammalian cells. *Microbiological reviews*, 59(3), pp.423–450.

Shi, Z., Johnson, J.J. & Stack, M.S., 2012. Fluorescence in situ hybridization for microRNA detection in archived oral cancer tissues. *Journal of Oncology*, 2012.

Singer, R.H. & Ward, D.C., 1982. Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog. *Proceedings of the National Academy of Sciences of the United States of America*, 79(23), pp.7331–7335.

Sleckman, B.P. et al., 1998. Accessibility control of variable region gene assembly during T-cell development. *Immunological reviews*, 165(I), pp.121–130.

Takeda, A. & Yaseen, N.R., 2014. Nucleoporins and nucleocytoplasmic transport in hematologic malignancies. *Seminars in Cancer Biology*, 27, pp.3–10. Available at: http://dx.doi.org/10.1016/j.semcancer.2014.02.009.

Tani, H. et al., 2012. Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Research*, 22, pp.947–956.

Tani, H. & Akimitsu, N., 2012. Genome-wide technology for determining RNA stability in mammalian cells: Historical perspective and recent advantages based on modified nucleotide labeling. *RNA Biology*, 9(10), pp.37–41.

Tani, H. & Torimura, M., 2013. Identification of short-lived long non-coding RNAs as surrogate indicators for chemical stress response. *Biochemical and Biophysical Research Communications*, 439(4), pp.547–551. Available at: http://dx.doi.org/10.1016/j.bbrc.2013.09.006.

Trcek, T. et al., 2012a. Single-mRNA counting using fluorescent in situ hybridization in budding yeast. *Nature protocols*, 7(2), pp.408–19. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4112553&tool=pmcentrez&rendertype=abstract.

Trcek, T. et al., 2012b. Single-mRNA counting using fluorescent in situ hybridization in budding yeast. *Nature protocols*, 7(2), pp.408–19.

Tripathi, V. et al., 2010a. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular Cell*, 39(6), pp.925–938. Available at: http://dx.doi.org/10.1016/j.molcel.2010.08.011.

Tripathi, V. et al., 2010b. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular Cell*, 39(6), pp.925–938.

Uchida, S. & Dimmeler, S., 2015. Long Noncoding RNAs in Cardiovascular Diseases. *Circulation Research*, 116(4), pp.737–750. Available at: http://circres.ahajournals.org/cgi/doi/10.1161/CIRCRESAHA.116.302521.

Vester, B. & Wengel, J., 2004. Current Topics LNA (Locked Nucleic Acid): High-Affinity Targeting of Complementary. *October*, 43(42), pp.13233–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15491130.

Van Vlierberghe, P. et al., 2008. Cooperative genetic defects in TLX3 rearranged pediatric T-ALL. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 22(4), pp.762–770.

Vlierberghe, P. Van & Ferrando, A., 2012. The molecular basis of T cell acute lymphoblastic leukemia. *The Journal of clinical …*, 122(10). Available at: http://www.jci.org/articles/view/61269?elq=6b34dfd3997d4382b7927d9b459e8f66.

Weng, A.P. et al., 2004. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science (New York, N.Y.)*, 306(5694), pp.269–271.

Windhager, L. et al., 2012. Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Research*, 22, pp.2031–2042.

Xia, F. et al., 2014. Dynamic transcription of long non-coding RNA genes during CD4+ T cell development and activation. *PLoS ONE*, 9(7), pp.1–11.

Xie, C. et al., 2014. NONCODEv4: Exploring the world of long non-coding RNA genes. *Nucleic Acids Research*, 42(November 2013), pp.98–103.

Yan, B., Wang, Z.H. & Guo, J.T., 2012. The research strategies for probing the function of long noncoding RNAs. *Genomics*, 99(2), pp.76–80. Available at: http://dx.doi.org/10.1016/j.ygeno.2011.12.002.

Yang, L. et al., 2011. NcRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell*, 147(4), pp.773–788. Available at: http://dx.doi.org/10.1016/j.cell.2011.08.054.

6. Supplementary material

Table 6. 1 *– Primers used in qRT-PCR. Primer3 was used to design the primers and Oligo Calc (Oligonucleotide Properties Calculator) to validate them in terms of self-complementarity. BLAST was used to validate them in terms of specificity. The primers were synthesised by Invitrogen™.*

Transcript	Primers	Product length
- Airn	Fw:5'GTGGGAGGGATCTTAGCACA'3 Rv:5'AGAATGGGCTGTGTGTTTCC'3	114
- Malat1	Fw:5'TAAGCGCTTGCCTCTGTCTT'3 Rv:5'CACCTGCATTCTGTGTGGTC'3	148
- Neat1	Fw:5'TGGGGATTATTGAGCTGAGG'3 Rv:5'CTCCTGAGTACCGGGATGAA'3	137
- Ens91	Fw:5'TGGGTCTGAAATTCTTGCTTG'3 Rv:5'GCGATCAGTCATTCCTGTCA'3	157
- Ens87 ex2	Fw:5'GCATGTTCTGCGCTGTGA'3 Rv:5'CCAGGTCCAGGGTGAAGATA'3	153
- Ens87 ex3	Fw:5'GCATGTTCTGCGCTGTGA'3 Rv:5'TCAGCATTCCCTCAAGTTCC'3	214
- Zeb2NAT ex4	Fw:5'GGCAGGAGGAAGAAGGAGAA'3 Rv:5'CCAATCCCTTCAGAGCAAAG'3	162
- Zeb2NAT ex2/ex3	Fw:5'AGAAGGACTTGCTCCCGAAT'3 Rv:5'TCCTGCCTCCAGACACTCTT'3	149
- Nup214 ex32	Fw:5'TGGAAACAGTGGAGCTAAGACA'3 Rv:5'CAAAACCTTGGGATGCTACG'3	113
- Nup214 e 32/e 33	Fw:5'TGGAAACAGTGGAGCTAAGACA'3 Rv:5'ACCCCTCCAAACCCAGGA'3	207
- Nup214 ex32/ex34	Fw:5'TGGAAACAGTGGAGCTAAGACA'3 Rv:5'ACTGGCGAGGGTACCGAAG'3	197
- Zeb2 ex2/e 4	Fw:5'GAGGCGTAACACGTCAGTCC'3 Rv:5'GAACTTGGAGAGGGGGTCTT'3	178
- Zeb2 ex2/e 3	Fw:5'TTCCCCTTCTCCAAGTTTTG'3 Rv:5'GAACTTGGAGAGGGGGTCTT'3	179
- Zeb2 ex4	Fw:5'GTCGCTGTGTTTGGTTGCTA'3 Rv:5'GTTTTTCCTCCTGGGATTGG'3	179
- Fos	Fw:5'CTTTATCCCCACGGTGACAG'3 Rv:5'TGACACGGTCTTCACCATTC'3	172
- β-Actin	Fw:5' CCCTGAAGTACCCCATTGAA'3 Rv:5' CTTTTCACGGTTGGCCTTAG'3	158
- Gapdh	Fw:5'AACTTTGGCATTGTGGAAGG'3 Rv:5'ACACATTGGGGGTAGGAACA'3	357

Table 6. 2 – *Relative expression levels of selected lncRNAs and mRNAs in TAP cells. Total RNA was purified from the cells, reverse transcribed with random primers and analysed by RT-qPCR using the ΔCT analysis method.*

		#1	#2	#3	#Average	Std. Dev.
LncRNA	**Malat 1**	0,06370	0,061299	0,07956	**0,068186**	0,0080999
	Neat 1	0,00310	0,003429	0,00454	**0,003690**	0,0006149
	Airn	0,01875	0,025572	0,01931	**0,021210**	0,0030930
	Ens91	0,00007	0,00006	0,00008	**0,000069**	0,0000110
	Ens87 (ex2)	0,000233	0,000262	0,000349	**0,000281**	0,0000490
	Zeb2NAT (ex4)	0,000003	0,000002	0,000002	**0,000002**	0,0000003
mRNA	**Nup214 (ex32)**	0,010344	0,010996	0,010419	**0,010587**	0,0002913
	Zeb2 (ex4)	0,000016	0,000015	0,000019	**0,000017**	0,0000019
	Fos	0,000026	0,000031	0,000034	**0,000030**	0,0000029
	B-Act	0,860472	0,506874	0,803238	**0,723528**	0,1549690
Control	**Gapdh**	1,00000	1,000000	1,00000	**1,000000**	0,0000000

Table 6. 3 – *LncRNA half-lives determined based on newly transcribed RNA/total RNA ratios. TAP cells were incubated with 4sU for 60min. Total and newly transcribed RNA was purified, reverse transcribed with random primers and analysed by RT-qPCR. Values were normalized to Gapdh and the half-lifes calculated as described (see formula in Figure 2.3, in methods section).*

	#1	#2	#3	#Average	# Std. Dev.
Malat 1	4,5	4,1	4,3	4,3	0,163
Neat 1	2,0	1,9	2,4	2,0	0,232
Airn	2,9	3,1	2,9	2,9	0,108
Ens91	3,3	4,1	4,2	4,1	0,397
Ens87	2,5	2,7	2,4	2,5	0,122
Zeb2NAT	2,6	1,1	*	1,9	0,749
Nup214	5,5	5,6	5,9	5,6	0,170
Zeb2	1,6	3,5	1,9	1,9	0,831
Fos	0,9	2,2	1,0	1,0	0,589
B-Act	7,3	*	*	7,3	0,000
Gapdh	7,0	7,0	7,0	7,0	0,000

Table 6. 4 – *Relative expression levels of the pair mRNA/NAT in TAP cells. Total RNA was purified from the cells, reverse transcribed with random primers and analysed by RT-qPCR using the ΔCT analysis method.*

	#1	#2	#3	#Average	Std. Dev.
Ens87 ex2	0,000233	0,000262	0,000349	**0,000281**	0,0000490
Ens87 ex2/3	0,000005	0,000005	0,000006	**0,000005**	0,0000007
Zeb2NAT ex4	0,000003	0,000002	0,000002	**0,000002**	0,0000003
Nup214 ex32	0,010344	0,010996	0,010419	**0,010587**	0,0002913
Nup214 ex32/33	0,000026	0,000029	0,000027	**0,000027**	0,0000011
Nup214 ex32/34	0,007164	0,008712	0,008650	**0,008175**	0,0007156
Zeb2 ex3/4	0,000013	0,000032	0,000023	**0,000022**	0,0000079
Zeb2 ex4	0,000016	0,000015	0,000019	**0,000017**	0,0000019
Gapdh	1,00000	1,000000	1,00000	**1,000000**	0,0000000

Table 6. 5 – *Primers used in standard PCR and DNA sequencing. We used the tools Primer3 to design the primers and Oligo Calc (Oligonucleotide Properties Calculator) to validate them. The primers were made by Invitrogen™.*

Transcript	Primers	Product length
- Ens87 ex2	Fw:5'CCGTGACTTTGCCAGTTTCT'3 Rv:5'GGCCAGTGTTGTAGCATTCA'3	892
- Zeb2NAT ex4	Fw:5'AAGAGTGTCTGGAGGCAGGA'3 Rv:5'GAAGTGGGCTCTGAATTGGT'3	1266
- Zeb2NAT ex2/ex3	Fw:5'TGGTAAACGACTCAAGTTCCAA'3 Rv:5'GGGTCTCTCTTCTCTCCCCTAC'3	1008
- Nup214 ex32	Fw:5' TCCTTGGGTCAGAGAGCAGT'3 Rv:5' GCACGTCTTATGTATTAGAAATGCAG'3	713
- Nup214 ex33	Fw:5' ATGGGGCTAGAAAGCAACCT'3 Rv:5'GCAGTCTTGAGTGGGGGTTA'3	899
- Nup214 ex34	Fw:5'AGGAGTATTTGAGGCCAGCA'3 Rv:5'TGAGAGCAGAAAGGCAGACA'3	913
- Zeb2 ex4	Fw:5'GGACCTGCTCCCCACTCT'3 Rv:5'GCTGGATCGCTTGTTTGTTT'3	881

***.Table 6. 6** - DNA Sequences for sm-RNA-LNA FSH probe design. Fragments were obtained by standard PCR and sent to StabVida Company for sequencing.*

>Nup214_exon32_154nt

GTGGATGGACCAAGACCTCTGGGAGACCTGCACCTGAGTGTCGGTGATTTCAGACCCAGCATCTGCAGCT CTGCTGGGCTGTGCCATTGCTCCTGTCCAGAGACTCGGCTGATCTGACTCCTTTCTGTCTTCTGAGCTTCTCCT ATTCAAGCTCCCTCACACCTGTTCTTGTTCCCTTTAGGATACCTCTAATCTGTTTGGAAACAGTGGAGCTAAGA CATTTGGAGGGTTTGGCAGCTCATCCTTCGGGGAACAGAAGCCTGCCGGCACATTCAGCTCTGGCGGTGGGA GCGTAGCATCCCAAGGTTTTGGATTTTCCACTCCAAATAAAACAGGTAACCCGCGAGTGAGCCATGCTCATCT GTGTGCGTGTGTGAATGACTTGTGAGCTTATAGAAGGTGCCCCTTTGGAGGCAACATC

>Ens87_exon2_167nt

TACTGGAAACTTCTGTCCCTTGAGACGGAGTTTCACACTGTACCCCAGATTGGCCTGGGACTCGGTATGG CTCACAGACTGGCCTTGAATTTACTGTGACAACTGTCTCCACAAGGAGCAGGCATGTTCTGCGCTGTGACCAG AGAAATTTCTGATCACTCTGAATCCTGCCAAAAGCCAGTCCATTTGCTGCTGTCTGGGCTCAGCCTTTTCCACT CCTCTAGACATGACGGCTGGGAATGCAGAGACAGGATTATCTTCACCCTGGACCTGGCTTTGGTTGGTATTTT ATCAGGTGACATTTTAGACCAGATTGGTGAAGGGTGGCTCTCTGGAGTGAGCATCTAGATGTCCAGGACTATT CCTGGCTGCCACTTGTCTTGCTGAGTGAGGTCTAGTGAGCTGACCTGACTAGTCCGGTCTCCTCAGGCAGGAG ACGCACATACCTTGTCTGCAGACGGGCAGGCTTCGTGATAAACAGGGCCCTTGCTCAGAGCATCTGTCCTGGA AGCCAAAGATTTTTATCTAGAGACCAAGGATGCTGTGTTTGTCTGGGATGGGGCTATTTCTGTGTACAGACAG GCAGGAAGATAAGAGGAGGGAAAACATCTTTCTGGGTACCATGGTGACAGGGACTTACTGTACCTTTGAAGTC TTTTATGAGCCTTTGGGGCAGGTGTTTACAGCTCTAAACTTGGGTGTCCCTAATTCTTTTAGTCCAGATATAGC TCATCATCAAGATGTGGTAAGGAAATGGGGACCCTGCAGGGCAGGGGGCTCCCTGTGAA

>Zeb2_exon3_isoform_201_573nt

GGCTCGGAGACAGATGAAGAGGACAAGCTTCACATTGCTGAAGATGATAGCCTTGCAAACCCTCTGGACC AGGACACCAGCCCAGCTAGCATGCCCAACCATGAGTCCTCCCCACACATGAGCCAAGGGCTGCTACCAAGAG AGGAAGAAGAGGAGGAGCTAAGGGAGAGTGTTGTGGAGCACAGCTGGCACAGTGGCGAGATTCTGCAAGCCT CTGTAGCCGGTCCAGGTAAGTGTCTGTCTGTCCGCCACTGTGTGGCCCCTTCCTGGGCCACCTGCATGTCACC TGTGCTTCCCAGCAGAACACCTGCAGCCACATCATTCAAATAGTTCTTTCCATCGCCCTTCTGGAGTAGAGAG CAGAGCTGTTGGCAGATGATGAAGTGATAGCAATGTGGTCACATCACGTACAGTACTCGGATTTTAGGAAATG ATTGTGTATTGGTTTCAACCCAAGTTGGATATCAACTCTAGTGTAACAAACACAAAACATGAAGTACTTGTACA TTTTCAATGTGAAACATTACAGAACGCAATCAAGATCTCTACTCTCATGACAGAAAGTATAAAGAA

>Zeb2NAT_exon4_368nt

TTGGATCCCTATATTAAAAGTTTGGCATTTTAATAAAGGGTCTCTAGGTTAACATAGAGCTGTACTATGCT GGCCAATCTCTGGACCCCTCTACACCTCATGCAGACCACATTACAAAGTGAGGGGACAGGGGCAGGAGGAAG AAGGAGAAAGGTCAGAGGAACTTAGGACGAAAGAAAGAAGAGGGTGACACTAAACTGTGTGGGGTCTAAGTT TTGAAGTGCATATTTTTGATATTGCAATTAAACTAAAATTAAAAAGAAAAGACTCCACCTTTGCTCTGAAGGGA TTGGTTATGCAAATATGGAAGGGATTTCCTGGAGAATACAGCTTTCTACTGTATGGTTAATTAAATAATCACTC AAAAAGCTTCCAACGTGACGCTTCTGTCGTAATCCAATCAGGTTACATAGGTCCTAAACAAGAAAGATATTTTC CACATCTGGAAGTCAGCAATTTAGCAAGTACTGCACATTATTAACCAATTCAGAGCCCACTTCCCAGGGGATTT TTTTTTGAAGTTTAAGTGTTCTTAACCAATGCTCTGCTGTTTTGTTAATCAAAAAGCTGGTTTACACTGCACATA ATTGGAA

***Table 6. 7** – LNA probes for sm-FISH. The LNA™-enhanced oligonucleotide was produced according to custom design by the Exiqon probe design software. The synthesized oligonucleotides were analyzed by High-Performance Liquid Chromatography (HPLC) and the identity of the compound was confirmed by Mass Spectrophotometry (MS).*

Target RNA	Probe sequence	RNA Tm (°C)
- β-Actin (positive control)	/5DigN/CTCATTGTAGAAGGTGTGGTGCCA	92
- Ens87 exon 2	/5DigN/AGGTCCAGGGTGAAGATAAT	83
- Zeb2NAT exon 4	/5DigN/TTCGTCCTAAGTTCCTCTGA	84
- Nup214 exon 32	/5Biosg/TCCAAACAGATTAGAGGTAT	83
- Zeb2 exon 3	/5Biosg/ACAGTGGCGGACAGACAGACA	83

***Table 6. 8** – Differential expression of lncRNAs between different samples. The reads were aligned to the genome of reference mmusculus.9 with the TopHat tool and the differential expression analysis was obtain by Cuffdiff tool using the NONCODE data base for lncRNAs. The transcript was considered differential expressed if p-value<0.05.*

Samples	no change	Diff. Exp.	upregulated	downregulated	Total
Control vs TLX3.1	100217	13824	7736	6087	114041
Control vs Tumour	99797	14244	7107	7143	114041
Control vs TAP	98403	15638	9200	6440	114043
TLX3.1 vs Tumour	106288	7753	3677	4080	114041
TAP vs TLX3.1	106093	7948	3946	4001	114041
TAP vs Tumour	104295	9746	4513	5232	114041

Samples	no change	Diff. Exp.	upregulated	downregulated	Total
Control vs TLX3.1	88%	12%	7%	5%	100%
Control vs Tumour	88%	12%	6%	6%	100%
Control vs TAP cells	86%	14%	8%	6%	100%
TLX3.1 vs Tumour	93%	7%	3%	4%	100%
TAP vs TLX3.1	93%	7%	3%	4%	100%
TAP vs Tumour	91%	9%	4%	5%	100%

Table 6. 9 – *Differential expression of protein-coding genes between different samples. The reads were aligned to the genome of reference mmusculus.9 with the TopHat tool and the differential expression analysis was obtain by Cuffdiff tool using the UCSC database for the protein-coding genes. The transcript was considered differential expressed if p-value<0.05.*

Samples	no change	Diff. Exp.	upregulated	downregulated	Total
Control vs TLX3.1	11515	11044	5172	5872	22559
Control vs Tumour	9432	13127	6518	6609	22559
Control vs TAP	11981	2037	5512	6469	22559
TLX3.1 vs Tumour	11607	10952	5841	5111	22559
TAP vs TLX3.1	12877	9682	4838	4844	22559
TAP vs Tumour	10554	12005	6499	5506	22559

Samples	no change	Diff. Exp.	upregulated	downregulated	Total
Control vs TLX3.1	51%	49%	23%	26%	100%
Control vs Tumour	42%	58%	29%	29%	100%
Control vs TAP cells	53%	9%	24%	29%	100%
TLX3.1 vs Tumour	51%	49%	26%	23%	100%
TAP vs TLX3.1	57%	43%	21%	21%	100%
TAP vs Tumour	47%	53%	29%	24%	100%

Table 6. 10 – *Differential expression of selected lncRNAs between different samples. The reads were aligned to the genome of reference mmusculus.9 with the TopHat tool and the differential expression analysis was obtain by Cuffdiff tool using the NONCODE data base for lncRNAs. The transcript was considered differential expressed if p-value<0.05.*

Average	**Control**	**TLX3.1**	**Tumor**	**TAP**
Malat1	0	82,0259333	42,3272	93,9326
Neat1	41,0316333	26,1613167	9,27949667	10,0858
Airn	23,9768	7,41208167	6,63468167	19,208
Ens91	0	0	0	0
Ens87	6,39245667	3,31799536	1,39993056	0,00164
Zeb2NAT	11,7327	0	15,4999167	0
Faddos	5,40129	11,6305033	33,4839333	31,4905
BC051242	10,7931333	27,01235	4,04176	72,2588

Stand.Dev.	**Control**	**TLX3.1**	**Tumor**	**TAP**
Malat1	0	7,88610068	5,81430885	5,99472
Neat1	0,56617798	1,89776271	0,98333134	0,62321
Airn	0,33070038	0,55077791	0,79130909	1,72405
Ens91	0	0	0	0
Ens87	0,08817048	3,33346803	1,40779165	0,00015
Zeb2NAT	0,16181875	0	1,75480227	0
Faddos	0,07450164	1,64682478	3,29510136	2,49049
BC051242	0,14897589	4,22687759	1,23839727	5,39096

Table 6. 11 – *Differential expression of selected protein-coding genes between different samples. The reads were aligned to the genome of reference mmusculus.9 with the TopHat tool and the differential expression analysis was obtain by Cuffdiff tool using the UCSC database for the protein-coding genes. The transcript was considered differential expressed if p-value<0.05.*

Average	**Control**	**TLX3.1**	**Tumor**	**TAP**
Nup214	12,6077667	13,4363333	12,4848667	13,9157
Zeb2	11,8948	0,03936613	12,5695167	0,00539
Fadd	4,13448667	4,87612	5,58560167	4,40808
Daxx	29,1951667	31,3134167	18,4181833	20,7008
Stand.Dev.	**Control**	**TLX3.1**	**Tumor**	**TAP**
Nup214	0,55160373	0,49483026	0,42133921	0,58727
Zeb2	0,52041171	0,00339548	0,40491831	0,00087
Fadd	0,1808927	0,20610972	0,25336436	0,19814
Daxx	1,27735342	1,07188539	0,69725356	0,87271