eman ta zabal zazu

Universidad del País Vasco | Euskal Herriko Unibertsitatea

Departamento de Enfermería

Facultad de Medicina y Enfermería

Investigating the role of miRNA binding in B-cell lymphoma

Tesis presentada por

ERIKA LARREA DE ORTE

San Sebastián, 20 de diciembre de 2018

Investigating the role of miRNA binding in B-cell lymphoma

Tesis presentada por

Erika Larrea de Orte

Para la obtención del título de Doctor en
Investigación Biomédica **por la** *Universidad del País Vasco/Euskal Herriko Unibertsitatea*

Tesis dirigida por

Dr. Charles Lawrie

A mi familia

ACKNOWLEDGEMENTS

En primer lugar me gustaría agradecer a mi director de tesis, el Dr. Charles Lawrie por darme la oportunidad de realizar este trabajo en su grupo de investigación. Por apostar por mi y por su tiempo y dedicación. También a todas las entidades financiadoras, en especial al programa de becas PFIS del Instituto de Salud Carlos III por posibilitar el desarrollo de esta tesis doctoral. Al doctor Jude Fitzgibbon por darme la oportunidad de realizar una estancia en su grupo en el centro de Hemato-Oncologia del Barts Cancer Institute en Londres y a todos los miembros de su grupo que me acogieron como a una más. Fue sin lugar a dudas una gran experiencia y muy enriquecedora.

Quisiera agradecer también a todos los miembros del grupo de Oncología Molecular de Biodonostia por todo lo que me han aportado durante estos cinco años. En especial a Marta Fernández, con la que trabajé codo con codo y de la que tuve el enorme placer de aprender. Le agradezco muchísimo cada minuto que dedicó en formarme desinteresadamente. También a Lorea Manterola por guiarnos, a mí y a todos los predocs en nuestros respectivos proyectos y también por los momentos que hemos compartido. A Ibai, mi co-PhD student, porque para mí ha sido una persona muy importante durante este proceso. Y por supuesto a Maitena, nuestra bioinformática súper crack. Por su entusiasmo y su infinita ayuda en el procesamiento de los datos PAR-CLIP : sin ti no hubiese sido posible! Y a todo el resto del grupo, porque todos me habéis aportado y espero haberos aportado algo yo a vosotros también.

Me gustaría darle las gracias a Giovanni por su gran ayuda tanto en la maquetación e impresión de este manuscrito como en el cumplimentado de la documentación para la universidad. Y también por todas nuestras charlas y por su amistad. Del mismo modo, me gustaría agradecerle a mi amigo Peio por ayudarme con los trámites de la Universidad. Porque he acudido a él con preguntas y pidiéndole favores en muchas ocasiones y siempre se ha ofrecido a ayudarme de buen grado. Agradezco el tiempo que pasaste en nuestro grupo y haber tenido la oportunidad de conocer a la gran persona que eres.

En realidad, quisiera agradecer a todas y cada una de las personas que forman y han formado parte de Biodonostia porque todas en su conjunto han hecho que me siente como en casa cada vez que entro por las puertas del edificio. Puedo decir con orgullo que saldré de Biodonostia sintiéndome una versión mejorada de mi misma y esa

mejora se debe en parte a todos vosotros porque me habéis formado como investigadora y como persona. En especial, me gustaría agradecer a personas como Angela, Giovanni y Estefanía por estar ahí en mis días más oscuros. Pero también a todas las personas con las que he compartido risas y buenos momentos sobre todo durante los últimos dos años. Tanto a los que siguen estando en mi vida como a los que han decidido marcharse.

También a personas que me han inspirado como Maria Caffarell en el ámbito científico, a mis amigas Samanta y Margui por ser un ejemplo de que no hay que seguir las pautas que marca la sociedad para ser feliz, a mis amigas Maddi y Lidia por ser un ejemplo de mujeres fuertes, valientes, profesionales y bellas bachateras-salseras. Por supuesto a Mikel y a Asier por Vietnam. Fue increíble y justo lo que necesitaba. A mi amigo Pedro, por invitarme a un fin de semana de 10 en Madrid. Me siento muy agradecida por nuestra amistad que se ha mantenido a pesar de los años y la distancia. Y a Ricardo, por traer una luz tan bonita a mi vida y apoyarme tanto durante la escritura de la tesis.

Pero por encima de todo quiero agradecerle a mi familia todo su apoyo, comprensión y amor. En especial a mi madre por haber estado ahí durante toda mi vida; para levantarme si caigo y para celebrar mis alegrías y mis éxitos llena de orgullo. Lamento mucho si en algún momento de mi vida me he distanciado. Pero ya estoy aquí de nuevo y esta vez para siempre. Y a mi hermana, porque a pesar de ser diferentes en muchas cosas, me encanta ver lo parecidas que somos en otras. Y porque me encanta la persona en la que se está convirtiendo.

En resumen, mil gracias a todos los que habéis hecho posible que llegue a este momento de mi vida con más fuerza, valentía, seguridad y siendo más yo que nunca.

GRACIAS A TODOS

INDEX

Index

ABBREVIATIONS

4-SU	4-thiouridine
6-SG	6-thioguanosine
Ab	Antibody
ABC-DLBCL	Activated B-cell-like diffuse large B-cell lymphoma
Ago	Argonaute
AID	Activation-induced cytidine deaminase
ALL	Acute Lymphoblastic Leukemia
AML	Acute Myeloid Leukemia
AMPK	AMP-activated protein kinase
ARMC10	Armadillo Repeat Containing 10
ASCT	Autologous stem cell transplantation
ATP	Adenosine triphosphate
B2M	beta-2-microglobulin
BCL10	B-cell lymphoma 10
BCL2	B-cell lymphoma 2
BCL2L11	Bcl-2-like protein 11
BCL3	B-cell lymphoma 3
BCL6	B-cell lymphoma 6
BCR	B-cell receptor
BL	Burkitt lymphoma
BLIMP1	B lymphocyte-induced maturation protein-1
bp	Base pairs
BR	Bendamustine-rituximab
BSA	Bovine Serum Albumin
CARD11	Caspase recruitment domain-containing protein 11
CBM complex	CARD11–BCL10–MALT1 complex
CCND1	Cyclin D1
CD10	Cluster of differenciation 10
CD19	Cluster of differenciation 19
CD20	Cluster of differenciation 20

CD21	Cluster of differenciation 21
CD22	Cluster of differenciation 22
CD23	Cluster of differenciation 23
CD3	Cluster of differenciation 3
CD30	Cluster of differenciation 30
CD40	Cluster of differenciation 40
CD40L	Cluster of differenciation 40 ligand
CD5	Cluster of differenciation 5
CD58	Cluster of differenciation 58
CD79A	Cluster of differenciation 79A
CD79B	Cluster of differenciation 79B
CD93	Cluster of differenciation 93
CDKN	Cyclin Dependent Kinase Inhibitor
cDNA	Complementary DNA
CDS	Coding sequence
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
CHOP	Cyclophosphamide, doxorubicin, vincristine and prednisone
Chr	Chromosome
CIITA	Class II Major Histocompatibility Complex Transactivator
CIP	Calf Intestinal alkaline Phosphatase
CLIP	Crosslinking and immunoprecipitation
CLL	Chronic lymphocytic leukemia
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CNS	Central nervous system
CREBBP	CREB Binding Protein
CSR	Class-switch recombination
CVP	Cyclophosphamide, vincristine and prednisone
CXCL	C-X-C Motif Chemokine Ligand
D	Diversity gene segment from Immunoglobulin variable region

DEL	Double expressor lymphomas
DGCR8	DiGeorge Syndrome Critical Region 8
DHL	Double hit lymphomas
DICER	Endoribonuclease Dicer
DLBCL	Diffuse large B-cell lymphoma
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTPs	Deoxyribose nucleoside triphosphate
Drosha	Drosha Drosha ribonuclease III
Drosha	Drosha ribonuclease III
E2F1	E2F Transcription Factor 1
E2F2	E2F transcription factor 2
EBF1	Early B-cell factor 1
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EP300	E1A Binding Protein P300
EPHA7	EPH Receptor A7
EZH2	Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit
FBS	Fetal Bovine Serum
FDC	follicular dendritic cells
FFPE	Formalin-fixed paraffin-embedded
FISH	Fluorescent in situ hybridization
FL	Follicular lymphoma
FLAG-tag	fusion peptide tag or epitope (sequence: DYKDDDK)
FLIPI	Follicular Lymphoma International Prognostic Index
FOXO	Forkhead box O transcription factors
FOXP1	Forkhead Box P1
Fwd	Forward

Abbreviations

G2E3	G2/M-Phase Specific E3 Ubiquitin Protein Ligase
GC	Germinal center
GCB-DLBCL	Germinal center B-cell-like diffuse large B-cell lymphoma
GEO	Gene Expression Omnibus
GEP	Gene expression profiling
GFP	Green fluorescent protein
GI	Gastrointestinal
GTP	Guanosine triphosphate
H3K27	histone H3 at lysine 27
H3K27me3	Trimethylation of histone H3 at lysine 27
H3K4	histone H3 at lysine 4
HAT	Histone acetyltransferase
HA-tag	peptide derived from the human influenza hemagglutinin (HA) molecule (amino acids 98-106)
HC	Heavy chain
hg19/GRCh37	Human genome assembly *GRCh37* (*hg19*) from Genome Reference Consortium
HIST1H2BO	Histone Cluster 1 H2B Family Member O
HITS-CLIP	High-throughput sequencing of RNA isolated by CLIP
HRAS	HRas Proto-Oncogene, GTPase
HRP	Horseradish peroxidase
HSC	Hematopoietic stem cells
IFRT	Involved-field radiotherapy
IFU	Infection forming unit
Ig	Immunoglobulin
IGH	Immunoglobulin heavy chain gene
IgV	Immunoglobulin variable region
IGV	Integrative Genomics Viewer
IHC	Immunohistochemistry
IKZF1	IKAROS Family Zinc Finger 1
IL	Interleukin

INPP5D	Inositol Polyphosphate-5-Phosphatase D
IP	Immunoprecipitation
IPI	International prognostic index
IRF4	Interferon Regulatory Factor 4
J	Joining gene segment from Immunoglobulin variable region
JAK	Janus kinase
Kb	Kilobase
kDa	Kilodalton
Ki67	Proliferation marker. Encode by MKI67 gene
KLRK1	Killer Cell Lectin Like Receptor K1
KMT2D	Lysine Methyltransferase 2D
KRAS	KRAS Proto-Oncogene, GTPase
LB	Lysogeny broth
LC	Light chain
let-7	lethal-7
LHD	Lactate dehydrogenase
LIN28A	Lin-28 Homolog A
lncRNA	Long non-coding RNA
LY6D	lymphocyte antigen 6 family member D
MBP	Major breakpoint region
MCL	Mantle cell lymphoma
MCOLN2	Mucolipin 2
MEF2B	Myocyte Enhancer Factor 2B
METTL15	Methyltransferase Like 15
MHC	Major histocompatibility complex
MICB	Inhibition of class I polypeptide-related sequence B
miRNA	MicroRNA
MOI	Multiplicity of infection
MPP	Multipotent progenitor cell
mRNA	Messenger RNA

mTOR	Mechanistic Target Of Rapamycin Kinase
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MUM1	Melanoma Associated Antigen (Mutated) 1
MYB	MYB Proto-Oncogene, Transcription Factor
MYC	MYC Proto-Oncogene, BHLH Transcription Factor
MYD88	Myeloid Differentiation Primary Response 88
MZ	Marginal zone
MZL	Marginal zone lymphoma
NCCN	National Cancer Centre Network
NF-Kβ	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next generation sequncing
NHL	Non-Hodgkin lymphoma
nt	nucleotide
NTC	Non-template control
ntFL	Non-transformed follicular lymphoma
o/n	overnight
OS	Overall survival
OSTN	Osteocrin
PACT	Protein activator of protein kinase R
PAR-CLIP	Photoactivable ribonucleoside-enhanced CLIP
PAX5	Paired box 5
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFS	Progression-free survival
PGM	Ion Torrent Personal Genome Machine
PIM1	Pim-1 Proto-Oncogene, Serine/Threonine Kinase
PRC2	Polycomb Repressive Complex 2
PRDM1	PR/SET Domain 1
pre-B	Precursor B-cell
pre-miRNA	precursor miRNA

pri-miRNA	primary miRNAs
pro-B	Progenitor B-cell
PTEN	Phosphatase and tensin homolog
QC	Quality control
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
RBP	RNA binding protein
R-CHOP	Rituximab + CHOP
R-EPOCH	Rituximab, etoposide, prednisone, vincristine, cyclophosphamide, and doxorubicin
Rev	Reverse
RIP	RNA immonoprecipitation
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNase	Ribonuclease
RNP	Ribonucleoprotein complexe
RPMI	Roswell Park Memorial Institute Medium
RRAGC	Ras Related GTP Binding C
RT	Reverse transcription
S1PR1	Sphingosine-1-Phosphate Receptor 1
SCR	Scramble
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SENP1	SUMO Specific Peptidase 1
SERPINA9	Serpin Family A Member 9
SET	**S**u(var)3-9, **E**nhancer-of-zeste and **T**rithorax
SF1	Splicing factor 1
SHM	Somatic hypermutation
SILAC	Stable Isotope Labeling by/with Amino acids in Cell culture
SMAD	SMAD family member
SNORA51	Small Nucleolar RNA, H/ACA Box 51
snoRNA	Small nucleolar RNA

Abbreviations

SNP	Single nucleotide polymorphism
SNV	Single nucleotide variants
SOCS	Suppressor Of Cytokine Signaling
SOX4	SRY-Box 4
SPI1	Hematopoietic Transcription Factor PU.1
SPOCK3	SPARC (Osteonectin), Cwcv And Kazal Like Domains Proteoglycan 3
STAT	Signal Transducer and Activator of Transcription
TBE	Tris-borate-EDTA
TBP	TATA binding protein
TBS	Tris-buffered saline
TCF3	Transcription factor 3
TCR	T-cell receptor
TdT	Terminal deoxynucleotidyl transferase
tFL	transformed follicular lymphoma
TGFβ	Transforming growth factor beta
THL	Triple hit lymphoma
TLR	Toll like receptor
TNFAIP3	TNF Alpha Induced Protein 3
TNFRSF14	TNF Receptor Superfamily Member 14
TP53	Tumour protein p53
TR	Targeted re-sequencing
TRBP	HIV-1 TAR RNA binding protein
tRNase Z	Endoribonuclease that generates the mature 3'-end of tRNA molecules
UBB	Ubiquitin B
UTR	Untranslated region
UV	Ultraviolet
V	Variable gene segment from Immunoglobulin variable region
VAF	Variant allele frequence
VEGF	Vascular Endothelial Growth Factor
WGS	Whole genome sequencing

WHO	World Health Organization
WT	Wild type
XPO5	Exportin 5
ZNF	Zinc-finger nuclease

1. INTRODUCTION

1.1 Introduction to Hematologic Malignancies

Haematological malignancies are a diverse group of neoplasms affecting hematopoietic and lymphoid tissues. The World Health Organization (WHO) recognises over 160 different types of hematologic malignancies in the updated 2016 classification scheme [1]. They display very different aetiologies, clinical presentation and progression, incidence and prognosis. The incidence of hematological malignancies in Western Europe is 24.5 per 100.000 habitants, and 20.4 per 100.000 habitants in Spain. Mortality rates for the same period were 8.3 and 7.2 per 100.000 habitants in Western Europe and Spain respectively [2]. The classification of different types of hematological malignancy has been the subject of great controversy over the years involving many different classification schemes (e.g. Keal, BNLI, REAL etc.), that have culminated in the general consensus with the WHO classification system [1]. In general terms hematological malignancies can be divided according to their respective cell lineage as being either myeloid or lymphoid, and according to the degree of maturity of the respective cell-of-origin. The WHO classification scheme also utilizes morphology, immunophenotype, and genetic and clinical criteria to further subdivide each category [3, 4] (Table 1.1).

WHO Classification of tumours of hematopoietic and lymphoid tissues

Myeloid neoplasms	
Myeloproliferative neoplasms	Mature neoplasms with effective hematopoiesis
Myelodisplasic síndromesº	Mature neoplasms with ineffective hematopoiesis
Myelodisplasic/Myeloproliferative neoplasms	Mature neoplasms
Acute myeloid leukemia (AML) and related precursor neoplasm	Precursor neoplasms
Lymphoid neoplasms	
Precursor lymphoid neoplasms	Precursor neoplasms
Mature B-cell neoplasms	Mature neoplasms
Mature T-cell and NK-cell neoplasms	Mature neoplasms
Hodgkin lymphoma	Mature neoplasms
Other neoplams	
Acute leukemias of ambiguous lineage	Precursor neoplasms
Myeloid and lymphoid neoplasms with eosinophilia and abnormalities of *PDGFRA, PDGFRB* or *FGFR1*	-
Histiocytic and dendritic cell neoplasms	Mature neoplasms
Post-transplant lymphoproliferative disorders (PTLD)	-

Table 1.1: WHO Classification of tumours of hematopoietic and lymphoid tissues

1. Introduction

Mature B-cell neoplasms are a group of tumours affecting lymphoid tissues, where the tumour cells retain a resemblance to mature B-lymphocytes, their probable cell of origin. This large group includes leukemias (e.g. chronic lymphocytic leukemia, B-cell prolymphocytic leukemia and hairy cell leukemia), lymphomas (e.g. Hodgkin and non-Hodgkin lymphomas) and plasma cell neoplasms (e.g. plasma cell myeloma, plasmacytoma, immunoglobulin deposition disease, osteosclerotic myeloma).

2.2 B-cell Non-Hodgkin lymphomas

B-cell Non-Hodgkin lymphomas (NHLs) comprise a heterogeneous group of diseases caused by neoplastic proliferation of B-lymphocytes. The incidence of these lymphomas is 16.2 new cases per 100.000 habitants in Westen countries [5]. Follicular lymphoma (FL) and Diffuse Large B cell lymphoma (DLBCL) are the most common forms of B cell lymphoma, cumatively accounting for 60 % of all B-cell NHLs [4]. According to the World Health Organization International Agency for Research on Cancer (IARC), the rate of non-Hodgkin B-cell lymphomas is rising worldwide, with an increase of 4.2% per annum [6], and is higher in western countries than developing countries [2]. There are over 30 distinct types and subtypes of non-Hodgkin B-cell lymphomas recognised by the WHO classification scheme that represent a wide spectrum of clinical presentations from small localized lesions to bulky lesions and disseminated tumour growth [1]. The aggressiveness of these subtypes is also very variable, ranging from indolent or low-grade lymphomas (ie. FL and marginal zone lymphoma (MZL)) to aggressive or high-grade lymphomas (ie. DLBCL and Burkitt lymphoma (BL)). Consequently, different lymphoma types require different treatments, and patient outcome varies markedly [7] (Table 1.2). For the majority of B-cell lymphomas, tumour cells recapitulate the normal stages of B-cell differentiation so they can be classified according to their normal counterpart [8]. Therefore, it is important to know the B cell differentiation process to better understand the biology of B-cell lymphomas.

Major types of non-Hodgkin B-cell lymphomas	Incidence (per 100.000)	5-year OS (% patients)	Cell of origin
Follicular lymphoma	3.23	75.6	Germinal centre B-cells
Mantle cell lymphoma	0.86	25	Naïve pre-germinal center B cells of the mantle zone
Diffuse large B-cell lymhoma	8.31	46.3	Germinal centre B-cells (GCB molecular subtype)
			Post-germinal centre B-cells (ABC molecular subtype)
Burkitt lymphoma	0.36	51	Germinal centre B-cells
Marginal zone lymphoma	3.44	61.2	Marginal zone B-cells
Plasmablastic lymphoma	0.11	16.3	Post-germinal center B-cell
Lymphoplasmacytic lymphoma	0.3	62	Post-germinal center B-cell

Table 1.2: Main types of non-Hodgkin B-cell lymphomas. *The table shows the incidence of these lymphomas, the 5-year overall survival (OS) and the cell of origin of each lymphoma. Epidemiologic data obtained from Smith et al 2015 [5].*

1.2.1 B-cell differentiation

1.2.1.1 Early hematopoiesis

B-lymphocytes originate from hematopoietic stem cells (HSC) in the bone marrow. During hematopoiesis, HSCs differentiate to multipotent progenitor cells (MPPs) and subsequently these cells differentiate into common myeloid progenitors (CMPs) and common lymphoid progenitors (CLPs) giving rise to myeloid and lymphoid lineages respectively. The commitment of CLPs to the lymphoid differentiation pathway is highly regulated via a number of specialised transcription factors including *SPI1, IKZF1, TCF3, EBF1 and PAX5* [9, 10]. *EBF1* and *PAX5* expression are highly dependent on IL-7 signalling, which is also a critical factor in this process. In addition, it has been described that the expression of LY6D and CD93 mark the earliest B-linage committed progenitors [11, 12].

Cells differentiate from CLPs to progenitor B cells (pro-B). These pro-B cells are particularly dependent on IL-7 for survival and proliferation [13]. In pro-B cells heavy chain (HC) immunoglobulin (Ig) rearrangement begins. Once this process has been

successfully initiated cells become precursor B cells (pre-B). During the pre-B cell stage HC Ig gene rearrangement is completed while light chain (LC) gene rearrangement takes place to generate a fully functional B cell receptor (BCR). At this stage the cell becomes an immature B cell [14]. A schematic representation of hematopoiesis is shown in Figure 1.1.
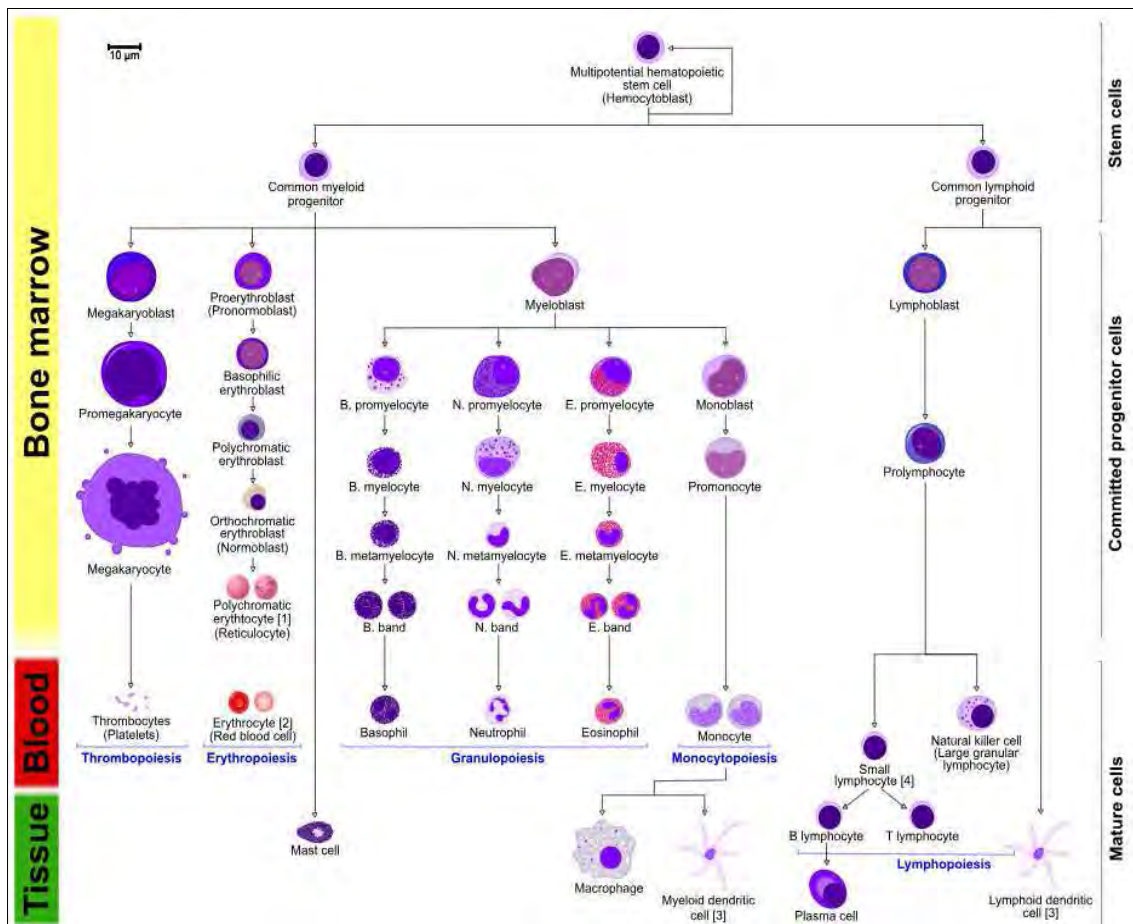
Figure 1.1: Schematic diagram of hematopoiesis depicting the various lineages and their gross morphological characteristics.

1.2.1.2 Ig gene rearrangement

The rearrangement of Ig genes is a somatic recombination process that only occurs in developing lymphocytes during early stages of T and B cell maturation. The immunoglobulin heavy chain gene (IGH) (located in chromosome 14) contains multiple copies of three different types of gene segments that constitute the variable region of the antibody protein (Variable (V), Diversity (D) and Joining (J) gene segments) as well

as two constant gene segments. This recombination process starts with the recombination of a D segment and a J segments in a stochastic way. Subsequently, a V gene segment will recombine with the previously produced D-J fragment. Variability is introduced in this junction due to TdT enzyme which is a template-independent DNA polymerase [15]. A primary transcript will be then generated and sliced to remove the sequence between VDJ segment and one of the constant gene segments. Translation of this mRNA leads to the production of the IgM heavy chain protein.

A similar process occurs for the synthesis of Ig light chain in pre-B cells with the exception that light chain genes do not have D segments so only V-J gene segments will be rearranged. This light chain protein will be assembled from one of the two existing gene loci: Ig-κ locus on chromosome 2 or Ig-λ locus on chromosome 22. Upon successful assembly, two light chains are combined with two heavy chains resulting in the formation of IgM molecule expressed on the surface of the immature B cell (Figure 1.2). At this point the immature B cell will be exposed to self-antigens in the bone marrow to exclude cells those showing auto-reactivity. Those cells with the capacity to bind self-molecules are eliminated by apoptosis while the non-reactive cells receive signals for survival and maturation [16-19].
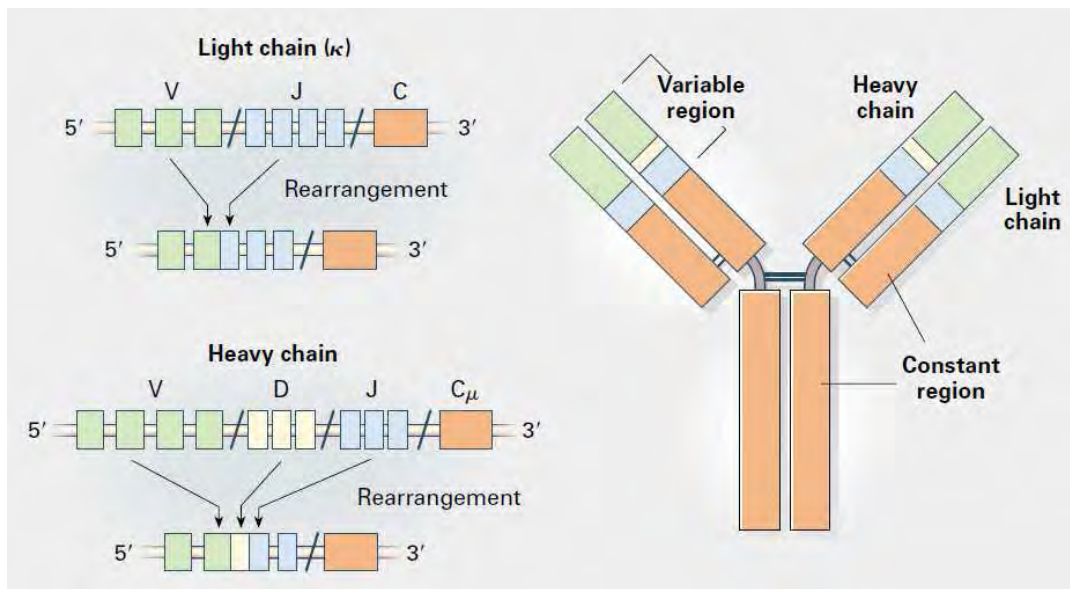
Figure 1.2: V(D)J Recombination in B-Cell Development. Antibodies consist of two identical heavy chains and two identical light chains. Each of these chains is subdivided into constant (C) and variable (V) regions. The variable-region genes are assembled by somatic recombination of

V, diversity (D), and joining (J) segments for the heavy chains, and V and J segments for the light chains. In this recombination process, the DNA located between the rearranging gene elements is deleted from the chromosome and variability is introduced between the segments. The germ-line DNA contains many different V, D, and J segments, so many different antibodies can be generated. Reproduced from Kuppers, et al. 1999 [20].

1.2.1.3 B-cell maturation

Once immature B cells leave the bone marrow and enter the blood circulation, they reach the spleen where they proceed through their final developmental stages to become mature B cells. Immature B cells can either become marginal zone (MZ) B cells, resident B cells of the splenic marginal zone, which are specialized to respond quickly to blood-borne pathogens, or follicular B cells, which are circulatory cells that constitute the main B cell compartment capable of mounting T-dependent immune responses. Immature B cells enter the spleen and populate an area known as the red pulp where they are exposed to more self-cells and circulating proteins. Cells showing autoreactivity undergo apoptosis. After this last checkpoint, cells diffferentiat into naive follicular B cells, which are fully mature B cells that can participate in immune responses. These mature B cells will circulate through all lymphoid tissues [14].

In the resting state, mature B cells are compartmentalized in secondary lymphoid organs as lymph nodes (Figure 1.3). Once antigen enters lymph nodes, it is captured by antigen presenting cells including macrophages, dendritic cells and also B cells. In B cells the antigen bound to the BCR is internalized, processed so antigen peptides can be exposed on major histocompatibility complex (MHC) class II molecules on the surface of the B-cell. This activation promotes chemokine receptor expression changes allowing these cells to migrate toward T cell zone. At the T-cell B-cell interface the MHC class II molecule with the presented peptide in the membrane of the activated B cells interacts with the T-cell receptor (TCR) of T helper lymphocytes that recognize the presented antigen. Apart from this recognition, interaction between co-stimulatory protein CD40 in B cells and CD40L in T helper cells is needed. These interactions promote B-cell proliferation and the migration from T-cell zone to primary follicle. Some B-cells will move to the interfollicular zone and differentiate into short-living

antibody secretor plasma cells, providing a rapid antibody response to the infection, although it will be a low affinity response. In contrast, some B cells remain in the follicles and form a germinal centre (GC) where cells experience an affinity maturation process. In the germinal centre reaction, iterative cycles of selection, proliferation and mutation occur resulting in a selection of multiple B cell clones with increased specificity and affinity towards the antigen [21, 22].
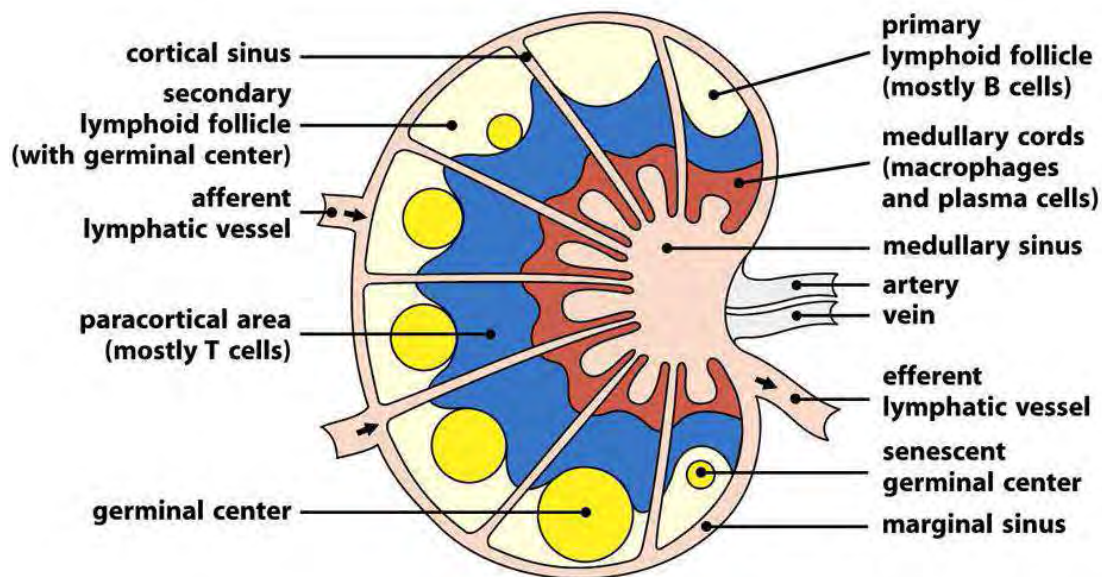
Figure 1.3: Lymph node structure. Naïve B-cells enter the lymph node through the high endothelial venules in the medullary cords. B-cells that bind antigen become activated at the border between the T-cell (blue) and the B-cell (yellow) areas. Activated B-cells then migrate into a primary follicle, proliferate and form a germinal center. Reproduced from Immunobiology: The Immune System in Health and Disease, 2005 (6th Edition) [23].

Mature GCs form dark and light zones within the lymph node. Dark zones contain intensively packed proliferating cells known as centroblasts and in this area, B cells undergo somatic hypermutation (SHM) of their Ig V genes, involving the expression of BCL6 [24, 25]. This protein acts as a transcriptional repressor of several genes of DNA damage response, including TP53 [26]. The key enzyme in this process is activation-induced cytidine deaminase (AID) that deaminates cytosine residues to uracil resulting in random mutations in the variable region of Ig genes due to errors in the DNA repairing machinery creating more diversity in the antigen recognition capacity of antibodies [27, 28]. A balance of sensitivity to CXCL13 expressed by follicular dendritic

1. Introduction

cells (FDC) in the light zone and CXCL12 expressed by stromal cells at the GC-T interface will promote the migration of B-cells between the dark zone and the light zone migration [29, 30] (Figure 1.4).
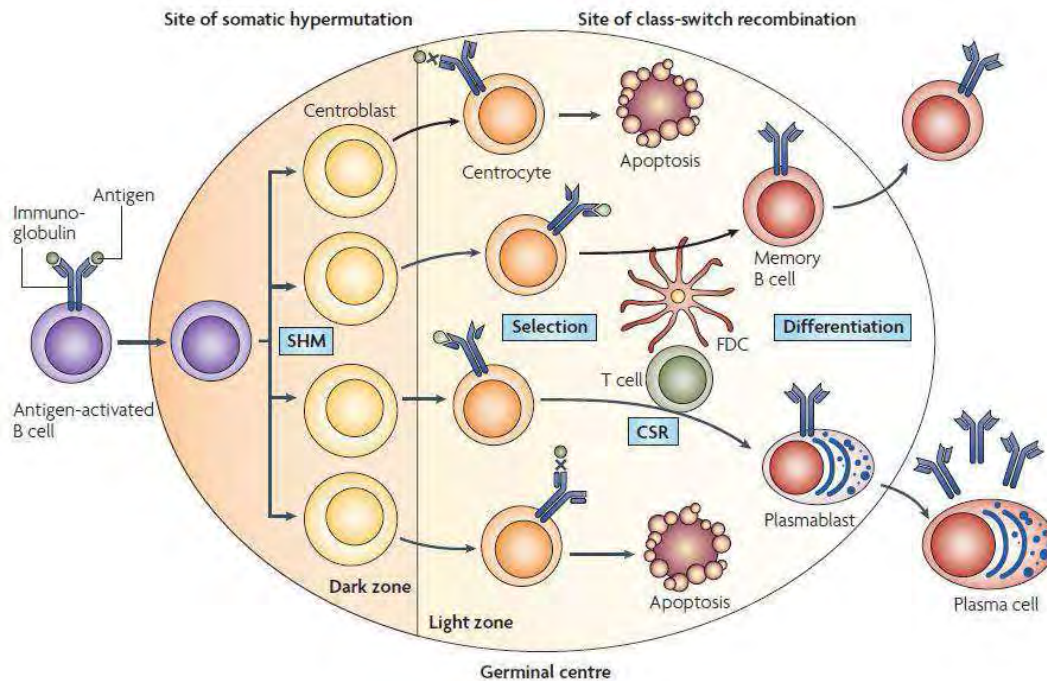
Figure 1.4: The germinal centre microenvironment. Antigen-activated B cells differentiate into centroblasts that undergo clonal expansion in the dark zone of the germinal centre. During proliferation, the process of somatic hypermutation (SHM) introduces base-pair changes into the V(D)J region of the rearranged genes encoding the immunoglobulin variable region (IgV) of the heavy chain and light chain. Centroblasts then differentiate into centrocytes and move to the light zone, where their BCRs are selected if improved binding to the antigen. Otherwise undergo apoptosis and are eliminated. A subset of centrocytes undergoes immunoglobulin class-switch recombination (CSR). Cycling of centroblasts and centrocytes between dark and light zones seems to be mediated by a chemokine gradient, presumably established by stromal cells in the respective zones. Antigen-selected centrocytes eventually differentiate into memory B cells or plasma cells. Reproduced from Klein, et al. 2008 [31].

The light zone is characterised by the presence of centrocytes which are B-cells wit lower proliferative activity and smaller size than centroblasts. This area also includes FDC, macrophages and T-helper lymphocytes. After the SMH process in the dark zone, cells expressing BCRs with high affinity for antigens are positively selected as a result of interactions with FDC in the light zone. In the FDC the antigen is held in native form and

B cells recognizing antigens with a high affinity will receive survival signals. One of these signals is the expression of BCL2 anti-apoptotic protein in positively selected B-cells [32]. In contrast, B-cells showing a low affinity are eliminated by apoptosis and phagocytised by the follicular macrophages. T-helper cells in the GC express cytokines such as IL-4 and IL-10 that help regulate B-cell differentiation [33-35]. Also, cytokines produced by antigen specific T-helper cells stimulate Ig class-switch in GCs in order to produce a more specific and effective immune response according to the pathogen (Figure 1.5). In addition, some hypothesis hold that interaction of B cells with antigen held on FDC is then followed by antigen uptake and presentation to local T helper cells in the light zone [36, 37].
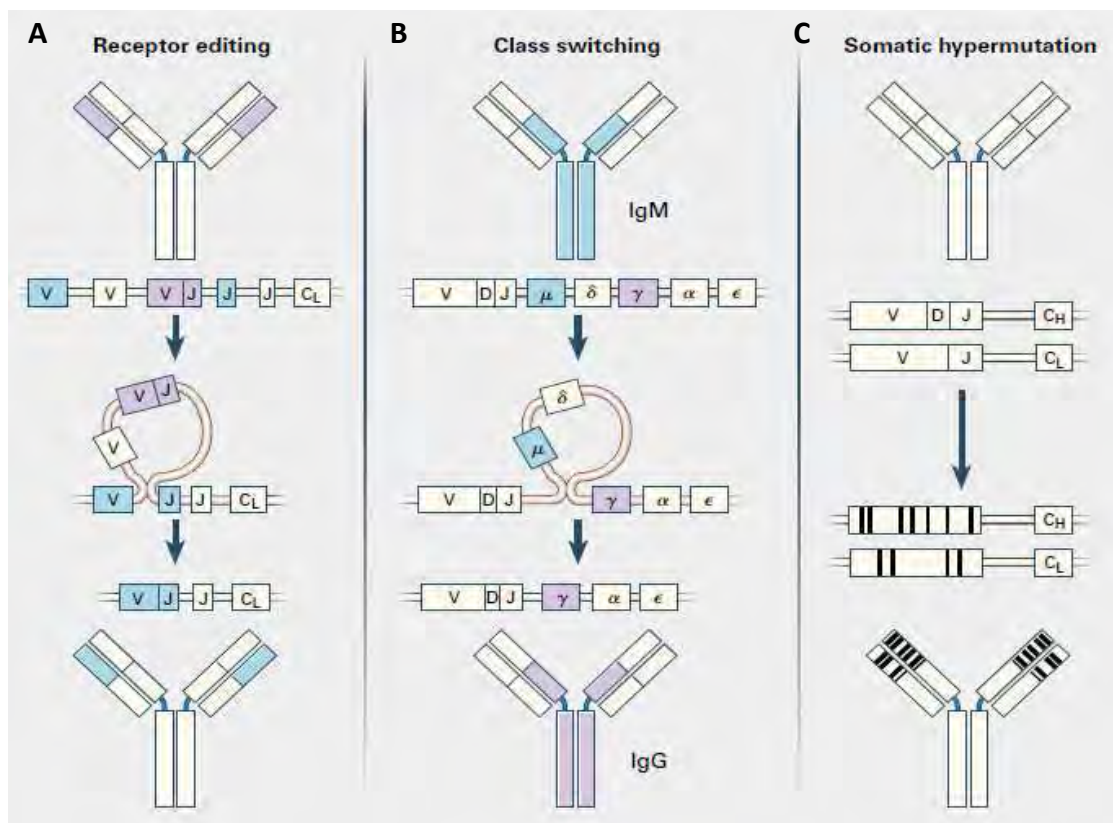
Figure 1.5: Molecular Processes Modifying the Genes Encoding Antibody Molecules. A) Receptor editing is a process by which an originally expressed antibody polypeptide chain, usually the light chain, is replaced by another one. B) Class switching process results in the replacement of an expressed heavy-chain constant-region gene (IgM or IgD) by a downstream constant-region gene, allowing expression of IgG, IgA, or IgE and therefore changing the effector functions of the antibody. C) The DNA sequences of the variable-region genes can be modified by SHM. This process introduces point mutations and occasional deletions and

duplications at a very high rate, specifically into the DNA of heavy- and light-chain variable-region genes. Reproduced from Kuppers, et al. 1999 [20].

Finally, antigen specific B-cells differentiate into plasma cells or memory cells that leave the germinal centre. The function of plasma cells is to secrete antibodies to the bloodstream. Some of these plasma cells will become long-lived plasma cells in the bone marrow niches and will be able to live up to decades secreting antibodies to the bloodstream [38, 39]. On the other hand, long-lived memory cells express BCR with somatically hypermutated BCR and a specific heavy chain for the antigen (IgG, IgA, IgE etc.). Upon activation of these cells, a rapid and specific immune response will be generated. These cells tend to home to the site where they were formed [40, 41].

1.2.2 Cell of origin in B-cell lymphomas

As mentioned above, non-Hodgkin B-cell lymphomas are a heterogeneous group of lymphomas affecting B cells. These lymphomas are classified according to different features, including immunophenotype and cytogenetic aberrations, but also according to their cell of origin [4]. It has been shown that most B-cell lymphoma subtypes are derived from either GC or immediate post-GC stage of B-cell differentiation. Actually, among lymphomas derived from mature B cells, unmutated variable regions of the Ig genes have been only found in mantle-zone lymphomas and some cases of B-cell chronic lymphocytic leukemia [42, 43]. The somatic mutated variable region in the BCR in a lymphoma indicates that it has been originated from GC or post-GC. Also, lymphomas can be distinguished between these two possibilities. In germinal centre B cell tumours, ongoing somatic mutation of the variable region of the BCR within the tumour clone can be observed, as in the case of follicular lymphoma. In contrast, if intraclonal variable region Ig gene diversity is absent, this may indicate that the cell of origin is a post-germinal-centre B cell (Figure 1.6). The fact that most B-cell lymphomas have their origin in cells from the germinal centre or post-GC suggests that the microenvironment provided by germinal centres represents a risk for malignant transformation. Indeed, both proliferative expansion processes along with somatic hypermutation processes and Ig class switching are taking place simultaneously in GCs. Failures in the control of these processes could be critical to B-cell lymphomagenesis as

they can promote the appearance of the characteristic chromosomal translocations and point mutations of many of these lymphomas [44]. However, some chromosomal translocations can occur during Ig genes rearrangements in the bone marrow, perhaps reflecting an early event in lymphomagenesis [20].
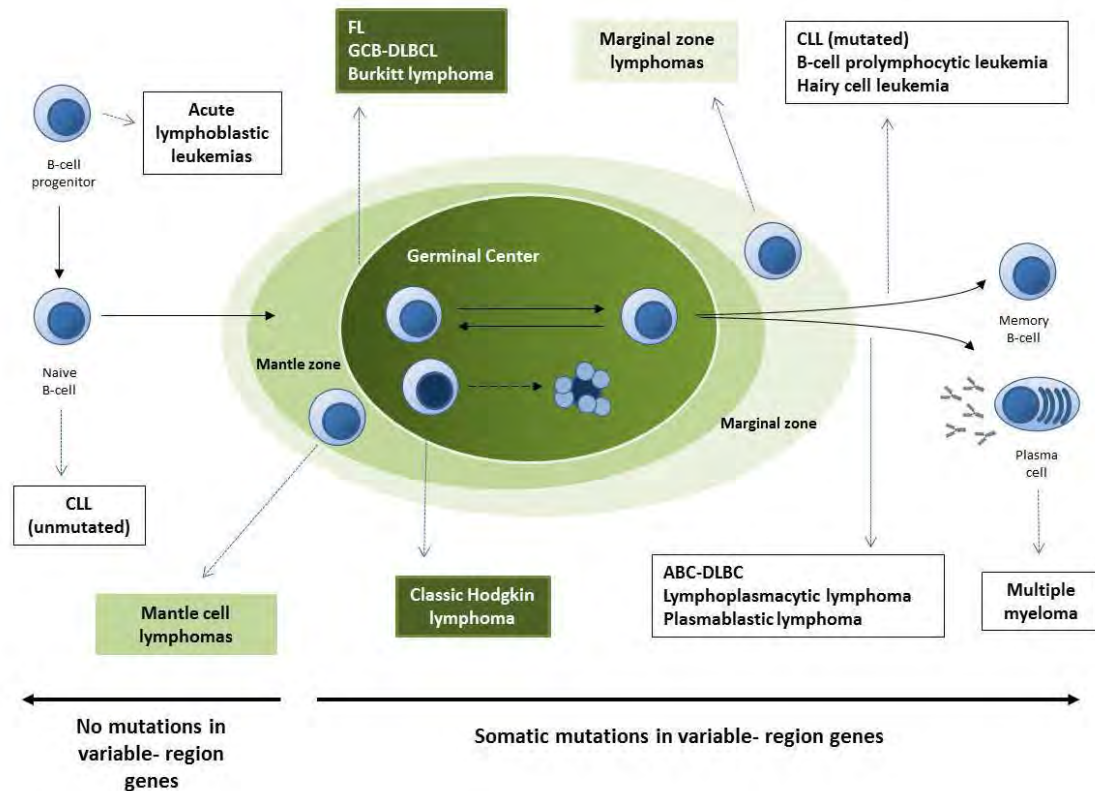
Figure 1.6: Assignment of some Human B-Cell Lymphomas and Leukemias to Their Normal B-Cell Counterparts. Naive B cells that recognize antigen with their antigen receptors establish germinal centers. Germinal-center B cells that acquire affinity-increasing mutations are positively selected and differentiate into memory B cells or plasma cells. Acute lymphoblastic leukemia represents a cancer of B-cell progenitors. Lymphomas/leukemias showing no somatic mutations in the variable region are derived from naïve B-cells that haven't undergo an antigen-based maturation and selection process in the germinal centres. Also, neoplasms with absence of intraclonal variable-region gene diversity might be derived from post-germinal-center B cells.

Additionally, some B-cell lymphomas such as FL retain the dependency to the expression of their BCR and to the germinal centre microenvironment for their survival and growth [45, 46]. In contrast, other lymphomas appear to have lost this dependency, showing a more aggressive phenotype such as is the case with DLBCL and

BL. Therefore, the diversity among different B-cell lymphoma types reflects not only the cell of origin but also the acquired somatic mutations of these cells [20]. Below we discuss we discuss some of the more common B-cell lymphomas in more detail.

1.3. Types of non-Hodgkin B-cell lymphomas

1.3.1 Follicular lymphoma

Follicular lymphoma (FL) is the most common indolent non-Hodgkin lymphoma (NHL) and the second most common NHL accounting for nearly 20% of all lymphomas in Western countries with an annual incidence of 3,3 new cases per 100,000 habitants [5]. The median age at diagnosis is 65 years old and it is slightly more frequent in women than men [5]. The incidence of FL, as with other non-Hodgkin lymphomas, is rising by 2% per annum and whilst the reason for this increase is unclear, might be a consequence of the increase in global life expectancy as the incidence of FL increases significantly with age [3, 47].

FL is classified as an indolent or low-grade lymphoma that commonly presents as a painless and slow progressive adenopathy. The initial symptoms of FL include painless swelling in one or more lymph nodes. However, some patients in advance stages can experience some symptomatology called B symptoms including fever of unknown origin (> 38ºC), night sweats and a quick weight loss. Even if the disease is highly responsive to therapy, it is generally considered incurable and patients normally experience relapse-remission cycles. The life expectancy for these patients varies considerably among individuals but, despite this "incurability", outcomes are normally good with median overall survival exceeding 12 years [48]. Even though, a proportion of FL patients have a high risk of death from the disease. Approximately 45% of FL patients will experience transformation to an aggressive lymphoma, most frequently DLBCL, resulting in a poorer prognostic outcome of 5 years [49].

1.3.1.1 Clinical management of FL

FL is characterized by the proliferation of neoplastic GC B cells, both centrocytes and centroblasts, that follow a characteristic follicular pattern [4]. The current histologic grading system for FL evaluates the proportion of centrocytes to centroblasts. Grade 1 is defined as 0-5 centroblasts per high-powered field, Grade 2 as 6-15 centroblasts per high-powered field and Grade 3 as > 15 centroblasts per high-powered field. Grade 3 is further classified as 3A and 3B with the latter characterised by an absence of centrocytes [4, 50]. It is widely accepted that FL3B represents a biologically distinct entity with a clinical course more similar to DLBCL. Indeed, FL3B is often negative for t(14;18) and CD10 expression [51, 52].

In addition to morphological features, FL diagnosis relies on the immunohistochemical positive expression of BCL2, BCL6, CD10, CD20, CD79a and a negative expression ofCD5 and CD3. Positive CD21 and CD23 staining are often used to allow identification of the follicular pattern and the proliferative marker Ki67 is useful as an indicator of low-grade lymphoma (generally <30% of positivity) [53]. When the immunophenotype is not clear the presence of BCL2 translocation t(14;18)(q32;q21) is detected by fluorescent in situ hybridization (FISH) [53].

FL is a very heterogeneous disease and the prognosis varies widely among individuals. Grading of FL can be helpful in determining prognosis and optimal therapy for patients. Apart from the histological grading, another way for staging FL disease widely used in the clinic is the modified Ann Arbor Classification system [54]. This staging system classifies FL patients in four groups according to the degree of dissemination of the disease:

Stage I   Single lymph node group

Stage II   Multiple lymph node groups on same side of diaphragm

Stage III   Multiple lymph node groups on both sides of diaphragm

Stage IV   Multiple extranodal sites or lymph nodes and extranodal disease

The majority of FL patients present with advanced disease at time of diagnosis, and only 10-20% present with early stage disease (Ann Arbor I or II) [55]. For treatment of

1. Introduction

limited stage or localized FL, local radiotherapy in the affected areas results in long term disease-free survival and potential cure. However, for patients with a life expectancy lower than 15 years a watch-and-wait strategy is often more appropriate. In contrast, in advanced stage FL, the treatment approach depends on presence or absence of specific symptoms including the tumor burden and age of the patient [48].

The addition of rituximab, an anti-CD20 monoclonal antibody, to conventional chemotherapy has significantly improved the outcome of FL patients [56, 57]. CD20 is specifically expressed in the membrane of pre-B cells and mature B-cells, both normal and tumoral. Treatment with this monoclonal antibody promotes an immune response against B-cells with the bound antibody. However, the optimal chemotherapy backbone remains undecided. The most commonly used chemotherapy backbone regimens are a combination of cyclophosphamide, vincristine and prednisone (CVP), the combination of cyclophosphamide, doxorubicin, vincristine and prednisone (CHOP), or only bendamustine [48].

The prognosis for an individual patient can be estimated using The Follicular Lymphoma International Prognostic Index (FLIPI) that subclassifies patients according to five independent prognostic factor [58]: number of involved nodal sites lactate dehydrogenase (LDH) levels, age >60 years, stage III or IV disease and haemoglobin levels (cut-off <12g/dL). The FLIPI provides a roughly equal distribution of patients across low-risk (0-1 factor), intermediate-risk (2 factors), or high-risk (>3 factors) categories. The 10-year OS rates calculated for the original cohort used to calculate the index were 71% (low risk), 51% (intermediate risk), and 36% (high risk) [58]. However, it should be borne in mind that the FLIP index was derived from FL patients (>4000) treated before the advent of routine rituximab treatment [59]. An alternative index, FLIPI-2 developed in the post-rituximab era includes, age > 60 years, elevated b2-microglobulin (B2M), haemoglobin <12 g/dL, bone marrow involvement, and lymph node diameter >6 cm as independent risk factors for poor progression-free survival (PFS) [60].
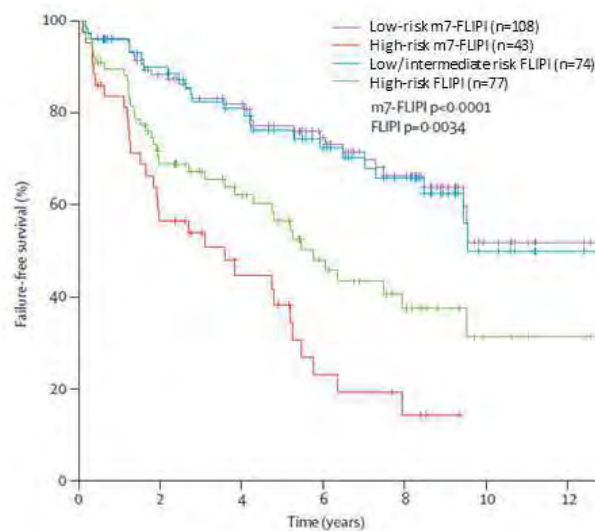
Figure 1.7: The clinicogenetic risk model m7-FLIPI versus FLIPI model. *Kaplan-Meier curves for failure-free survival by FLIPI and by m7-FLIPI. Modified from Pastore et al. 2015 [61].*

Also, more recently, a clinic-genetic model that integrates the mutational status of 7 genes (*EZH2, ARID1A, MEF2B, EP300, FOXO1, CREBBP,* and *CARD11*) was added to FLIPI (called m7-FLIPI), and was shown to be more efficient in identifying high risk FL populations (Figure 1.7). However, this system needs to be validated prospectively [61]. It has been observed that one of strongest predictor of long-term outcomes in FL is the length of first remission after a standard immunochemotherapy induction, with patients that experience relapse within 2 years having the worse prognosis [62].

1.3.1.2 Pathogenesis of FL

FL arises from the malignant transformation of normal GC B-cells. The first event of the oncogenic cascade is generally attributed to the chromosomal translocation t(14;18)(q32;q21), which is considered the molecular hallmark of this pathology and present in 85% of FL cases [63]. This somatic rearrangement is thought to be consequence of defective RAG-mediated VDJ recombination in the bone marrow during B-cell lymphopoiesis [64, 65]. The t(14;18) translocation leads to constitutive expression of the anti-apoptotic BCL2 protein as the protein encoding region of the gene comes under control of the highly expressed IGH promoter region. When these naive B cells carrying the t(14;18) exit the bone marrow, they colonize secondary lymphoid tissue and undergo a GC reaction they have a survival advantage due to their

constitutive expression of *BCL2*, which is not normally expressed in the germinal center [64, 66].

However, cells harbouring t(14;18) can be detected in the peripheral blood of >50% of healthy individuals, suggesting that the rearrangement itself is insufficient for malignant transformation and that secondary genetic alterations are required [67, 68]. These alterations are likely to be acquired during the SHM process in germinal centres [69]. In addition, some recurrent chromosomal abnormalities are also common in FL such as losses of 1p36 and 6q as well as gains of 7,18 and X [70, 71]. Analysis of genes contained in deleted regions has led to identification of critical tumour suppressors such as EPHA7 and TNFRSF14 [72, 73].

Mutations in histone-modifying genes have also been found to be characteristic of germinal centre derived lymphomas such as FL. Mutations in chromatin-modifying genes occur with high frequency, affecting histone methyltransferases, histone acetyltransferases or histone linker proteins, suggesting that these mutations act in concert, rather than independently, to promote lymphomagenesis. For example, truncating and missense mutations affecting *KMT2D* are observed in 60-70% of FL cases and loss of KMT2D results in decreased global H3K4 methylation levels and down-regulation of key sets of genes involved in immune signalling and B cell differentiation [74, 75]. Also, inactivating mutations of *CREBBP* and *EP300* acetyltransferases are frequently found in FL can have an impact on global gene expression [76]. Gain-of-function mutations of *EZH2* increase gene repression via enhanced H3K27 trimethylation and this transcriptional profile favours proliferation of lymphoma cells and results also in repression of plasma cell differentiation signatures [77]. EZH2 is a well characterized histone methyltranferase in lymphoma. It is the catalytic subunit of the Polycomb Repressive Complex 2 (PRC2), which is responsible in physiological conditions, for healthy embryonic development through the epigenetic maintenance of genes responsible for regulating development and differentiation[78]. *EZH2* has been shown upregulated in B cells during the germinal centre reaction promoting cell proliferation [79]. Activating somatic mutations in the catalytic SET domain of EZH2 have been identified in more than 25% of follicular lymphoma patients and have been suggested as early events in FL [80].

Similarly, frequent somatic mutations in *MEF2B* transcription factor have been also identified and result in an enhanced transcriptional activity. MEF2B has been described as a transcriptional activator of *BCL6*, therefore, this mechanism may contribute to lymphomagenesis as BCL6 plays a key role in germinal centres formation and maintenance and is frequently overexpressed in FL [81]. In addition, mutations affecting genes involved in immune modulation (*B2M, CD58, TNFRSF14)*, JAK-STAT signalling (*SOCS1* and *STAT6*), BCR/NF-κB signalling (*BCL10*, *CARD11*, *CD79B*) and mTOR signalling (RRAGC) have been reported [82, 83].

In addition to the tumour cells themselves, the microenvironment plays a crucial role in FL pathogenesis. This is highlighted by the fact that malignant FL cells are unable to grow *in vitro* in the absence of extrinsic survival signals [84, 85]. FL tumours are highly infiltrated by non-malignant cells that are normally present in germinal centres in physiological conditions. FL cells retain dependency on BCR expression. Therefore, cells from the microenvironment support their proliferation and survival. SHM in FL cells can introduce alterations in the variable region of BCR, making them more reactive and allowing the interaction with GC microenvironment cells even in the absence of antigen presentation [48]. In addition, it has been shown that the expression profile of infiltrating T lymphocytes is different between the ones present in FL tumours and in healthy individuals. T helper lymphocytes are significantly enriched in FL tumours stimulating CD40 signalling and secreating IL4 to promote tumor cell survival and growth [86-88] suggesting that tumor cells "re-educate" the tumor microenvironment.

1.3.1.3 High grade transformation

Approximately 45% of FL patients will experience histological transformation to a more aggressive type of lymphoma, termed transformed FL (tFL), drastically reducing their life expectancy from median overall survival of 12 years to just over 2 years [89]. Transformation is also associated with a more aggressive clinical course, and resistance to treatment. The occurrence of high-grade transformation is largely diagnosed on the basis of histological features. Most commonly this is defined by the proportion of large cells infiltrating the lymph nodes increasing leading to effacement of follicular architecture and becoming more characteristic of DLBCL morphology (Figure 1.8). The simultaneous presence of FL and DLBCL in the same initial diagnostic biopsy represents

1. Introduction

a composite histology which implies but does not necessarily confirm an early transformation of FL [89]. For this confirmation, a demonstration of a clonal relationship between the original FL and subsequent neoplasms is required. This can be established by analysing the variable region of Ig gene and showing that they share a backbone of common somatic mutations [89].
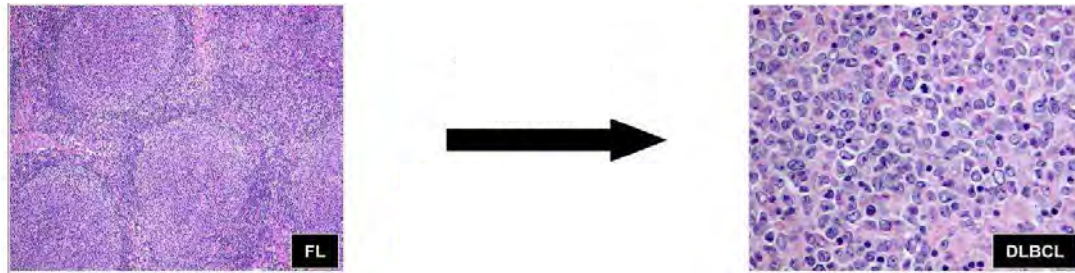
Figure 1.8: Histological transformation of FL to DLBCL. *The figure shows the typical follicular pattern of a FL tumour versus a diffuse architecture after histological transformation with a high proportion of infiltrating large cells. Modified from Lossos, et al. 2011 [89].*

FL transformation does not occur due to a single genetic or epigenetic event, but appears to arise through a subset of acquisitions of genetic aberrations. tFL tumors are typically genetically more complex than to their antecedent follicular lymphoma tumours [90]. FL transformation is associated with alterations that deregulate cell-cycle progression and DNA damage response genes such as *CDKN2A/B, TP53* and *MYC*. The oncogene *MYC,* in particular, can be deregulated in a number of ways including copy number amplifications, point mutations and translocations [91]. In addition, acquisition of somatic mutations affecting regulators of NF-κB signalling such as in *MYD88* and *TNFAIP3* has been associated with transformation [82]. These mutations are typically associated with an activated B-cell-like (ABC) phenotype [92, 93]. In contrast to this, the cell of origin of tFL is classified as a germinal centre B-cell-like (GCB) phenotype in approximately 80% of cases [94]. Aberrations in *BCL6, CIITA, PAX5, PIM1, B2M, EBF1* and *TNFRSF14* are notably increased in tFL [82, 91, 95], as well as mutations in *BCL2* that have been shown increased at transformation and correlate with a higher risk of

death [96]. Transformation has also been suggested to occur from the clonal expansion of minor subclones that become dominant under strong selective pressure [97]. According to this hypothesis, transformation results from divergent clonal evolution whereby both FL and transformed FL (tFL) dominant clones derive from a common progenitor cell through the independent acquisition of distinct mutations [91]. Currently there are no biomarkers to predict those patients that are at risk of transformation. The identification of such biomarkers could greatly improve the outcome for these patients by allowing them to be treated with more aggressive therapy upfront.

1.3.2 Diffuse Large B-Cell Lymphoma (DLBCL)

Diffuse Large B Cell Lymphoma (DLBCL) is the most common type of lymphoma accounting for 40% of all lymphomas in Western countries with an annual incidence of 8.3 new cases per 100,000 habitants. The median age at time of diagnosis is 70 years old with a slight male predominance. Similar to FL, the incidence of DLBCL also increases with age (Figure 1.9) [5].
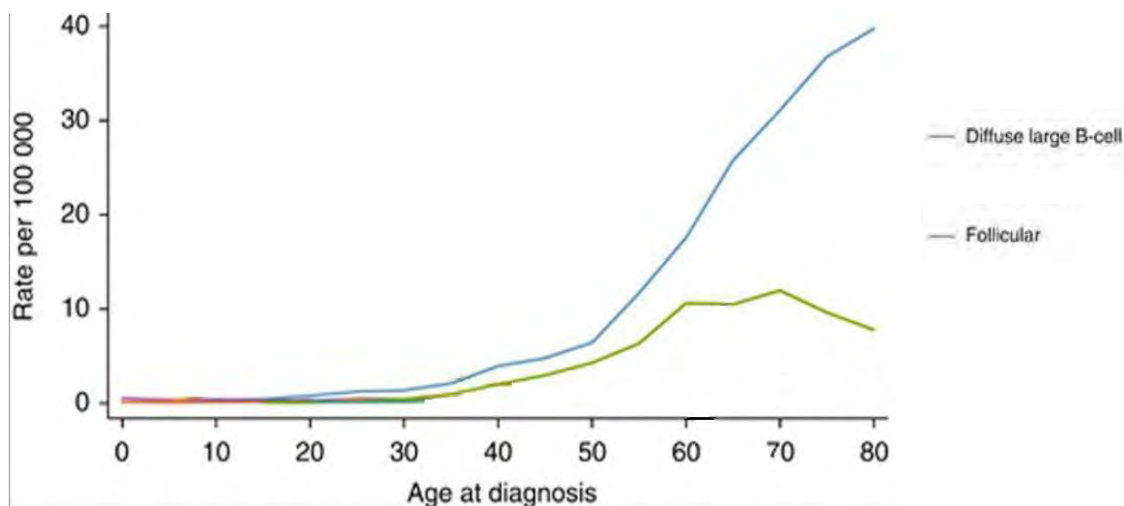
Figure 1.9: Age-specific incidence of FL and DLBCL. Incidence rates per 100,000 inhabitants. Data from Haematological Malignancy Research Network (HMRN) 2004–2012. Modified from Smith A, et al. 2015 [5].

Clinically, most patients present with a rapidly growing tumour mass involving one or more lymph nodes and extranodal sites (~40% of patients). About half of patients

present with stage I-II disease and the other half with advance disease or stage III-IV according to Ann Arbor staging system. In addition, approximately one-third of patients present with associated B symptoms (i.e. fever, weight loss and night sweats), and some patients with symptoms associated with organ(s) involvement. The frequency of bone marrow involvement in DLBCL is 10-20% and is associated with a worse overall survival [98]. Increased levels of serum lactase dehydrogenase (LDH) and beta-2-microglobulin (B2M) are often found in DLBCL patients [99].

The standard therapy for patients with DLBCL is chemotherapy, based on rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone (R-CHOP). Using this therapeutic approach around 60-70% of patients are cured. However, 30-40% of them will relapse or become refractory to R-CHOP treatment seriously compromising their outcome [100].

1.3.2.1 Diagnosis, molecular classification and prognosis molecular predictors

DLBCL is characterised at the morphological level by large neoplastic cells with a high proliferative index arranged in a diffuse pattern that totally or partially effaces normal lymph node architecture. Three morphological variants can be distinguished: centroblastic, immunoblastic and anaplastic, the most common being the centroblastic variant (80% of patients) [99]. DLBCL is diagnosed with a biopsy of an affected lymph node or from extraganglionar affected tissue that is immunophenotyped by either immunohistochemistry or flow cytometry using the B cell linage markers CD19, CD20, CD22 and CD79A and often the additional markers, CD5, CD10, BCL6, IRF4/MUM1, c-MYC, BCL2, ki67 (>40%) and CD30 [4, 101].

DLBCL is clinically a heterogenous disease that is reflected at the molecular level by gene expression profiling (GEP) microarray studies. The seminal work of Alizadeh *et al* in 2000 first revealed that DLBCL cases could be divided into two distinct molecular subtypes, germinal centre B-cell like (GCB) and activated B-cell like (ABC) as well as a small unclassifiable subgroup [102]. These two molecularly distinct forms of DLBCL have gene expression patterns indicative of different stages of B-cell differentiation. Furthermore, this study showed that the molecular subtypes were also prognostically

significant and that patients with GCB-derived disease had a better overall survival than non-GCB patients. On the basis of this study and the many confirmatory studies that followed, the revised WHO classification recommends assessing the cell of origin for all DLBCL cases [1]. However, this information is not yet clinically relevant as, with the exception of clinical trials, the same therapeutic strategies are used independently of the molecular subtype. The high costs and necessity of fresh or fresh frozen material for microarray technology means that GEP molecular classification is out of reach for most routine clinics. As a result, various immunohistochemical (IHC) algorithms have been developed as surrogates for GEP to predict the cell-of-origin of DLBCL cases. Among those, the Hans algorithm is the most commonly used [103]. This system employs CD10, BCL6 and MUM1 expression using 30% positivity cut-offs to classify between GCB and non-GCB tumours according to the following schema (Figure 1.10):
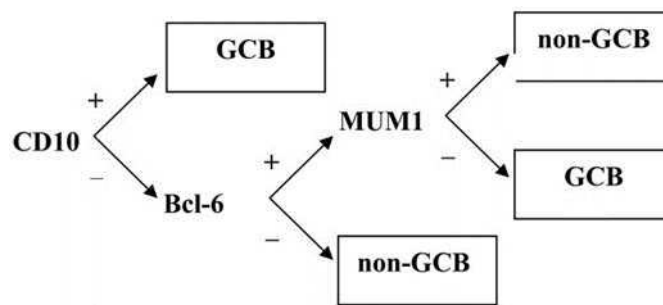
Figure 1.10: Hans algorithm, according to CD10, BCL-6 and MUM1 expression.

The Hans algorithm however is not particularly accurate, with a concordance of 76% with cell of origin detected by GEP. Consequently, other algorithms have been proposed with improved concordance to GEP classification of up to 93%. These algorithms use additional antibodies against of the GC-marker SERPINA9 (GCET1) and ABC-marker FOXP1 [104, 105]. However, these alternatives are not in widespread use yet probably as a result of the associated additional costs of the extra staining required.

DLBCL tumours usually present a complex karyotype and the FISH technique is generally the method of choice evaluating the presence of chromosomal rearrangements. Most commonly, paraffin embedded tissue sections are used for assessing rearrangements in *BCL6*, *MYC* and *BCL2* which are the major translocations in DLBCL. In about 30% of cases *BCL6*, encoded at chromosome 3q27 is juxtaposed with

1. Introduction

IGH located on chromosome 14q32, although *BCL2* can also be transposed to other loci. For example, *IGH-BCL2* translocation t(14;18)(q21;q32) is also observed in 20-30% of cases as well as translocations involving *MYC* at 8q24 in 10-15% of cases, that is associated with high-grade morphological features and bad prognosis [106]. Similarly, some studies have shown that *MYC* amplifications have a negative impact on overall survival even if the degree of negative prognosis is less than with the rearrangement [107, 108]. *BCL2* rearrangements also seem to correlate with poorer prognosis in patients with GCB subtype, but not in patients with non-GCB subtype (where the frequency is very low) or in the overall DLBCL patient group [109].

Those diseases with *MYC* and another oncogene rearrangement, usually *BCL2* but less often *BCL6* and rarely *BCL3* or *CCND1*, are known as double hit lymphomas (DHL). Cases with rearrangements of *MYC, BCL2*, and *BCL6* are known as triple hit lymphoma (THL) [110]. Paradoxically, even if they show aggressive clinical courses, nearly all cases of *MYC/BCL2* DHL and *MYC/BCL2/BCL6* THL have a GCB cell-of-origin [111]. Also, some DLBCL cases without translocations can overexpress these oncogenes. Double expressor lymphomas (DEL) are defined as lymphomas in which *MYC* and *BCL2* are overexpressed, but not due to translocation and represent 25-35% of all DLBCL cases. They are also associated with poor prognosis and they are more common in ABC/non-GCB group [112, 113]. Other biomarkers for prognosis in DLBCL are the expression of CD5, which is present in 5–10% of cases, and it is often associated with aggressive clinical behaviour and poor survival when treated with R-CHOP therapy. These patients have often a complex karyotype and are usually an ABC/non-GCB type [114]. Also, mutations in *TP53* have been shown to be an independent predictor of poorer prognosis. These mutations are present in about 20% of cases of DLBCL and occur in a similar frequency between GCB and ABC/non-GCB type [115]. On the other hand, CD30 expression (10–15% of cases) has been shown to correlate with better prognosis [116].

Overall, for all these reasons, the 2016 updated version of WHO classification recommends assessing in the immunohistochemistry panel at the time of diagnosis the following: (1) the type of lymphoma according to the cell of origin (GCB vs non-GCB) using Hans algorithm as previously explained, (2) the expression of MYC and BCL2 (double expressors) as an adverse prognostic factor, and (3) the expression of CD5 as

an adverse prognostic factor. In addition, TP53 and CD30 expression can also be evaluated as prognosis factors. In those cases with c-MYC overexpression, assessing the presence of translocations of *c-MYC* using FISH is recommended. If the result is positive, the presence of rearrangements in *BCL2* and *BCL6* should also be assessed [1].

1.3.2.2 Prognostic assessment and treatment

In 1994 the international prognostic index (IPI) was developed to stratify risk for patients diagnosed with DLBCL. The IPI includes five variables which are adverse prognostic factors: disease stage, age, performance status, serum LDH levels and the presence or absence of multiple extranodal sites affected. The IPI score divides DLBCL patients into four risk groups: low risk (0-1 factors), low/intermediate risk (2 factors), intermediate/high risk (3 factors) and high risk (4-5 factors) [117]. However, rituximab has greatly improved the outcome of patients with DLBCL and as a result the IPI score lost some of its prognostic power. The National Cancer Centre Network (NCCN) proposed an enhanced IPI system which stratifies risk in DLBCL patients more accurately in the chemoimmunotherapy era (Figure 1.11). The NCCN-IPI refined categorization of age and LDH, and the identification of disease involvement at specific extranodal sites. This system has a maximum total score of 8 points and can stratify DLBCL patients into four distinct risk groups: low (0-1 point); low-intermediate (2-3 points); high-intermediate (4-5 points); and high (≥6 points) (Table 1.3) [118].

IPI risk factors
(1) Age>60 years
(2) Ann Arbor stage III/IV
(3) More than 1 extranodal site
(4) LDH level above normal
(5) (*)ECOG performance status ≥2

NCCN-IPI risk factors	
Parameters	Score
Age, years	
>40−60	1
>60−75	2
>75	3
LDH, normalised ratio	
>1−3	1
>3	2
Ann Arbor stage III−IV	1
Extranodal disease (marrow, CNS, lung, liver/GI tract)	1
Performance status ≥2	1

Table 1.3: Prognosis stratification of DLBCL according to IPI and NCCN-IPI scores. *LDH (lactase dehydrogenase), CNS (central nervous system), GI (gastrointestinal). (*) ECOG performance status measures how disease affects the daily living abilities of patients [118].*

1. Introduction
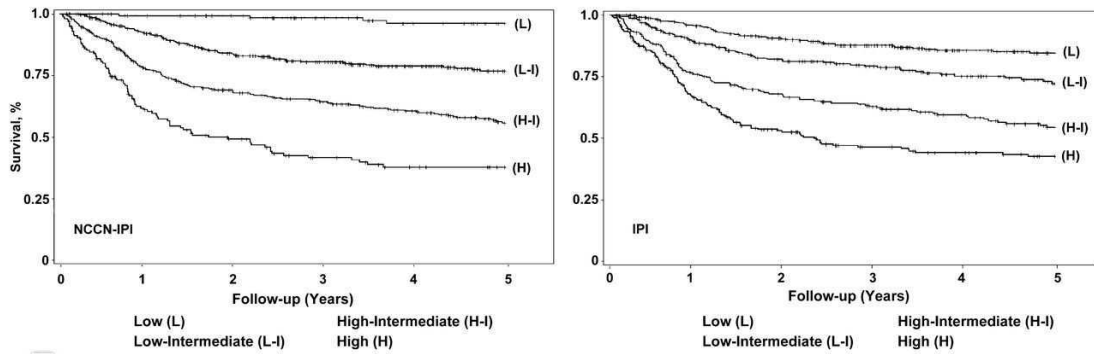
Figure 1.11: Risk stratification power of the two risk factor indexes (NCCN-IPI vs IPI). *Compared with the IPI, the NCCN-IPI better discriminated low- and high-risk subgroups (5-year overall survival [OS]: 96% vs 33%) than the IPI (5-year OS: 90% vs 54%), respectively. (n=1138 patients). NCCN-IPI is more powerful than the IPI for predicting survival in the rituximab era. Reproduced from Zhou, et al. 2014 [118].*

As previously mentioned, the standard frontline therapy for patients with DLBCL is chemotherapy with R-CHOP. The current management protocol of patients is based on the risk factors including the IPI score, along with the degree of dissemination of the disease. In approximately 30% of cases, DLBCL is presented as limited-stage disease, typically defined as Ann Arbor stage I or II, non-bulky and without B-symptoms. Historically, therapy relied on a combined modality of abbreviated chemotherapy followed by involved-field radiotherapy (IFRT). However, the use of radiotherapy didn't improve overall survival for these patients and therefore, sole use of chemotherapy treatment is recommended [119]. Three to four cycles of R-CHOP followed closely by IFRT, or six cycles of R-CHOP chemoimmunotherapy are the recommended regimen for these patients [120]. When disseminated disease is diagnosed, systemic chemoimmunotherapy (R-CHOP) administered every 21 days is the standard treatment [121].

The CD20 antibody Rituximab has dramatically improved the outcome of DLBCL patients, and when combined with CHOP chemotherapy (R-CHOP), patients in all risk groups have demonstrated benefit in overall survival [122]. However, treatment failure still occurs. For those patients with relapsed/refractory DLBCL, representing 30-40% of patients, high-dose chemotherapy and autologous stem cell transplantation (ASCT) is the standard approach. This second-line chemotherapy often goes in combination with

rituximab [123]. For patients who are not candidates for ASCT because of age or comorbidities, palliative chemotherapy should be considered. Alternatively, enrolment in clinical trials of new targeted therapies might be indicated.

Therefore, there is a clear need of identifying new biomarkers to promote a better risk stratification of DLBCL patients as well as biomarkers for response to therapy, to know which patients will do well with R-CHOP versus which ones will benefit from more aggressive treatment regimens. For example, dose adjusted rituximab, etoposide, prednisone, vincristine, cyclophosphamide, and doxorubicin (R-EPOCH) compared with standard R-CHOP as a first line immunochemotherapy resulted in improving the outcome in DHL patients [124]. On the other hand, investigating the biology of the disease is crucial to develop novel targeted agents against specific molecular alterations. Altogether, having a better understanding of the pathogenesis of DLBCL will be translate into developing better tools to facilitate treatment decisions allowing a transition towards a more personalized clinical management of patients as well as improving therapeutic strategies.

1.3.2.3 Pathogenesis DLBCL and novel therapies

DLBCL can be classified into two different molecular subgroups according to their cell of origin and gene expression profile. Therefore, it is not surprising that these subgroups differ in the molecular mechanisms leading to lymphomagenesis. GCB-type DLBCL, in common with FL, originates from GC B-cells, so its pathogenesis is intimately linked to the biology of the GC reaction. Somatic mutations common to FL are also found in GC-type DLBCL such as epigenetic modifier proteins *EZH2, EP300, CREBBP* and *KMT2D* [80, 82].

Under physiological conditions, EZH2 plays an essential role in the GC formation. This process favours the proliferation of centroblasts and the blockade of terminal differentiation. This histone methyltransferase is upregulated in centroblasts within the GC promoting transcriptional silencing of the genes required for memory and plasma cell differentiation such as *IRF4* and *PRDM1* as well as of checkpoint regulators such as *CDKN1A* and *CDKN1B* [77]. Missense mutations in *EZH2* occur in 21.7% of patients with GCB-DLBCL but are absent in ABC-DLBCL [125]. This might be because EZH2 represses

the germinal centre B-cell exit. These mutations are always heterozygous and are located in the EZH2 catalytic SET domain. The vast majority of them affect Y641 codon [125]. Whereas wild type EZH2 mediates mono- and di-methylation of H3K27, the Y641 mutant is more efficient at adding a third methyl group. For this reason, these mutations are considered activating mutations [126, 127]. Similarly, somatic missense and nonsense mutations of *KMT2D* histone methyltransferase occur in 23-27% of DLBCL patients. These mutations result in loss of function and may disrupt the catalytic SET domain of this protein [128, 129]. In addition, EP300 and CREBBP histone acetyltransferases are commonly mutated in GCB DLBCL, with mutation rates of 5-10% and 18-23% respectively [128]. *EP300* mutations are usually heterozygous and generally disrupt the HAT domain promoting a catalytically inactive enzyme. EP300 is a direct target for BCL6 so is hypothesised to function as a tumor suppressor protein in DLBCL [130]. Similar to *EP300*, *CREBBP* mutations are usually inactivating mutations that perturb the HAT domain [76].

In contrast, NF-κB activation through the classical pathway is a hallmark of the ABC-DLBCL subtype [131]. In B cells NF-κB signalling occurs downstream of the activation of various receptor molecules including BCR, CD40 and some Toll-like receptors (TLRs) [132, 133]. The NF-KB pathway antagonizes the apoptotic action of chemotherapy explaining why ABC is more refractory to chemotherapy [134]. Mutations promoting the activation of this signalling pathway are common features of ABC-DLBCL. For example, *CARD11*-activating mutations are present in approximately 10% of ABC-DLBCL tumors [135, 136]. CARD11 is a member of CBM complex which is required for the activation of the classical NF-κB pathway. These mutations confer hypersensitivity to upstream signals. In addition, mutations affecting BCR signalling (upstream to NF-κB signalling) are frequent with 20% of ABC-DLBCL cases containing mutations in the BCR receptor subunits, *CD79A* and *CD79B* [135]. Upon antigen encounter and BCR clustering, CD79A and CD79B transmit activation signals to a variety of downstream signalling pathways leading to cell survival and proliferation. Mutations in these genes promote chronic activation of the BCR receptor. Similarly, recurrent gain-of-function mutations in *MYD88* (adapter protein of TLR) are present in 30% ABC tumors and result in activation of NF-κB activity [92]. Genetic and functional data suggest that chronically

active BCR signalling and MYD88-dependent signalling cooperate to sustain the survival of ABC-DLBCL tumor cells [92]. In addition, ABC-DLBCL is a post-GC lymphoma so the plasmacytic differentiation is blocked at plasmablast stage. This blockade is achieved by inactivating BLIMP1 transcriptional factor which is necessary for driving mature B cells into terminal differentiation [137].

Somatic mutations common to both DLBCL subtypes include inactivating mutations of *TP53*, genes involved in immunosurveillance (*B2M, CD58*), and oncogenic activation of *BCL6* [1]. Apart from somatic point mutations chromosomic rearrangements are also common feature in DLBCL. Overexpression of the transcriptional factor *c-MYC* both by gene amplification or translocation and often confers aggressive clinical behaviour as previously mentioned. In GCs, *c-MYC* is expressed in the light zone but completely suppressed in proliferating GC centroblasts in the dark zone. Expression of *c-MYC* is required for re-entering the dark zone for additional rounds of positive selection within the GC reaction. In indolent lymphomas such as FL, acquisition of *c-MYC* rearrangement usually results in transformation into highly aggressive lymphomas [138]. Similar to FL, other oncogene rearrangements are also common in DLBCL such as *BCL6* and *BCL2* rearrangements, the latter being more frequent in GCB-DLBCL than FL [139].

The discovery of these biological features has promoted the development of targeted drugs and their introduction for the treatment of DLBCL currently restricted to clinical trials. Drugs targeting components of the NF-κB pathway, constituatively active in ABC-type DBCL, include Bortezomib or Ibrutinib. Bortezomib is a proteosome inhibitor that enables the arrest of IkB degradation and consequently the NF-κB pathway [140]. Ibrutinib, in contrast, targets BTK a specific tyrosine kinase of the NF-κB pathway [141]. Additionally, lenalidomide, an analogue of thalidomide, is an oral immunomodulatory drug that inhibits the production of certain pro inflammatory cytokines such as TNF-α, IL-1, IL-6 and IL-12 [142]. Lenalidomide has been approved for myelodysplasic syndrome (MDS) and myeloma and has shown promising data of efficacy and safety in relapse/refractory DLBCL patients [143]. Other promising drugs under study include Everolimus, an mTOR pathway inhibitor, Ventoclax, a BCL2 inhibitor, and tazemetostat, an EZH2 inhibitor (reviewed in [144]). Also, other promisingg therapies for DLBCL

treatment include immunotherapy such as therapeutic blockade of immune checkpoints [in particular cytotoxic T lymphocyte-associated protein 4 (CTLA-4) and programmed cell death protein 1 pathway (PD-1/PD-L1] as well as CAR T cell therapy [145]. Also, therapies based on microRNA targeting seem promising strategies [146-149].

1.4. MicroRNAs

1.4.1 History

The history of microRNAs (miRNAs) started in 1993 when two independent groups that were investigating *lin4-lin14* interactions in neural development of *Caenorhabditis elegans* made some interesting observations. Lee and colleagues demonstrated that the genomic locus containing *lin-4*, although functional and encoded for two small transcripts of 22 and 61nt, did not encode for a protein [150]. At the same time, the Ruvkun lab identified complementary regions to the *lin-4* transcripts in the 3'untranslated region (UTR) of the *lin-14* gene [151], establishing a novel post-transcriptional regulatory mechanism involving non-coding RNAs that can modulate gene expression by direct binding to the 3' UTR region of genes.

In 2000, another gene in nematodes, *lethal-7* (*let-7*) related to development was discovered which encoded a 21 nucleotide non-coding RNA rather than a protein product [152, 153]. Later that year, Pasquinelli et al. identified *let-7* gene homologues in other many animal species, including humans, and that the developmental regulation mediated by this miRNA was highly conserved [154]. Subsequent studies supported the idea that miRNAs were indeed conserved through evolution suggesting that they could be a general gene regulatory mechanism in eukaryotes [155-157].

1.4.2 Biogenesis

MiRNA biogenesis is a controlled and sequential process covering different steps starting at primary transcripts and ending in a mature and functional miRNA. MiRNA genes are widely located throughout the genome. It is estimated that approximately

50% of miRNAs are expressed from non-protein coding transcripts. It is frequent that clusters of miRNAs are transcribed polycistronically [158]. Also, many miRNA genes are located in intronic regions of protein coding genes and are generally co-transcribed with their host genes and processed separately [159].

MiRNAs are initially transcribed by RNA polymerase II as long (several hundred nucleotides) primary miRNAs (pri-miRNAs) containing stem-loop structures with a 5'-guanosine cap and a 3' polyadenylated tail and. This pri-miRNA is then processed into 70-120 nucleotide-long precursor miRNA (pre-miRNA) with a hairpin structure. This processing takes place in the nucleus by an RNase III enzyme called Drosha [160]. Drosha dimerizes with the cofactor protein DGCR8 to form a functional complex [161]. The pre-miRNA, with a 5' phosphate and a short 3' overhang, is then exported into the cytoplasm by exportin 5 (XPO5) in a Ran-GTP dependent manner [162, 163]. In the cytoplasm, pre-miRNAs are processed into mature 22 nucleotide-long miRNA asymmetric duplexes by another RNase III enzyme called Dicer [164, 165]. Similar to Drosha, cofactors such as TRBP and PACT are necessary for Dicer activity [166]. The duplex which is formed by the mature miRNA sequence, or guide strand, and the antisense passenger strand, also known as miRNA* is then associated to Argonaute (Ago) proteins. The antisense strand is usually degraded, despite there are some studies that report antisense guide can be also functional [167, 168]. At this point, Ago, TRBP and Dicer proteins together with the miRNA are assembled to create the RNA-induced silencing complex (RISC) where Ago proteins are the catalytic component of the complex [169, 170]. This complex is guided by the miRNA sequence to target sequences which are usually located in the 3'UTR of the target gene mRNA but miRNAs can also target 5'UTR sequences and sequences in coding regions [171, 172] (Figure 1.12).

1. Introduction
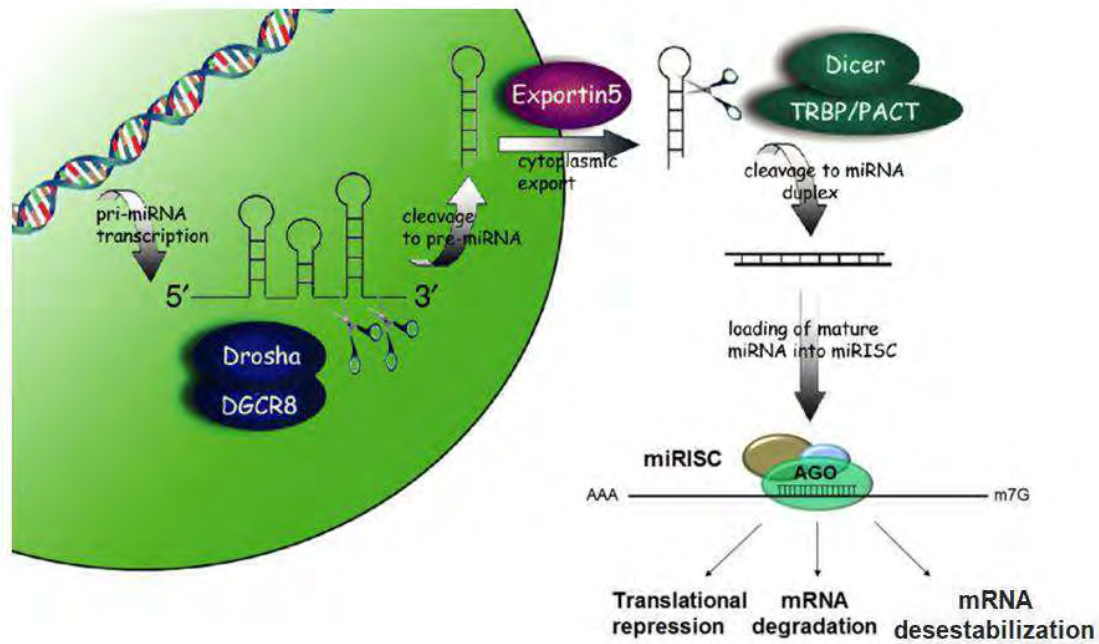
Figure 1.12: MiRNA biogenesis. *Pri-miRNAs are transcribed and cleaved to pre-miRNA trough Drosha-DGCR8 complex. Pre-miRNAs are then exported to the cytoplasm by Exportin 5 and further cleaved by Dicer. Finally, mature sequences are biologically functional to modify gene expression. Modified from Lawrie CH, et al. 2007 [173].*

Although most miRNAs are generated by the canonical pathway, there are also alternative miRNA biogenesis pathways. The majority of these pathways are Drosha/DGCR8-independent which can generate pre-miRNA-like hairpins that can be recognized by Xpo5 and Dicer. This is the case with *mirtrons* which are short intronic hairpins excised by splicing and then debranched. Because of their small size (derived from small introns) they bypass the Drosha-processing step even if they usually need to be trimmed by nucleases before being export into the cytoplasm [174, 175]. Similarly, other small RNA species such as snoRNA-derived miRNAs or tRNaseZ-derived miRNAs can also skip Drosha processing and enter in the canonical biogenesis pathway [176]. A Drosha/DGCR8-dependent but Dicer-independent biogenesis strategy has been reported for the conserved vertebrate miRNA *miR-451*. Similar to conventional miRNAs, the *pri-miR-451* is subjected to Drosha/DGCR8-processing, resulting in a *pre-miR-451* hairpin structure with a short ~18 bp length, which is directly incorporated into Ago2 resulting in a Drosha-dependent and Dicer-independent non-canonical biogenesis pathways [177]. The agotrons constitute one of the most recently

44

discovered classes of evolutionary conserved RNA-species that associate with Ago proteins [178]. In contrast to all the Ago-associated RNAs that have been described to date, agotrons bypass the canonical miRNA biogenesis pathway entirely, and thereby provide the first endogenous example of a both Drosha/DGCR8- and Dicer-independent biogenesis strategy. Similarly, to mirtrons, the agotrons are transcribed from host gene introns, but they escape Dicer processing. Agotrons are then exported to the cytoplasm by a hitherto unknown mechanism, possibly by Xpo5, and loaded onto Ago as unprocessed, full-length ~100 nt, debranched introns. Albeit to a lesser extent than conventional miRNA species, agotrons have also been shown to be capable of functioning as post-transcriptional regulators of gene expression in a seed-dependent manner despite their long mature length and much more complex secondary structure [178] (Figure 1.13).
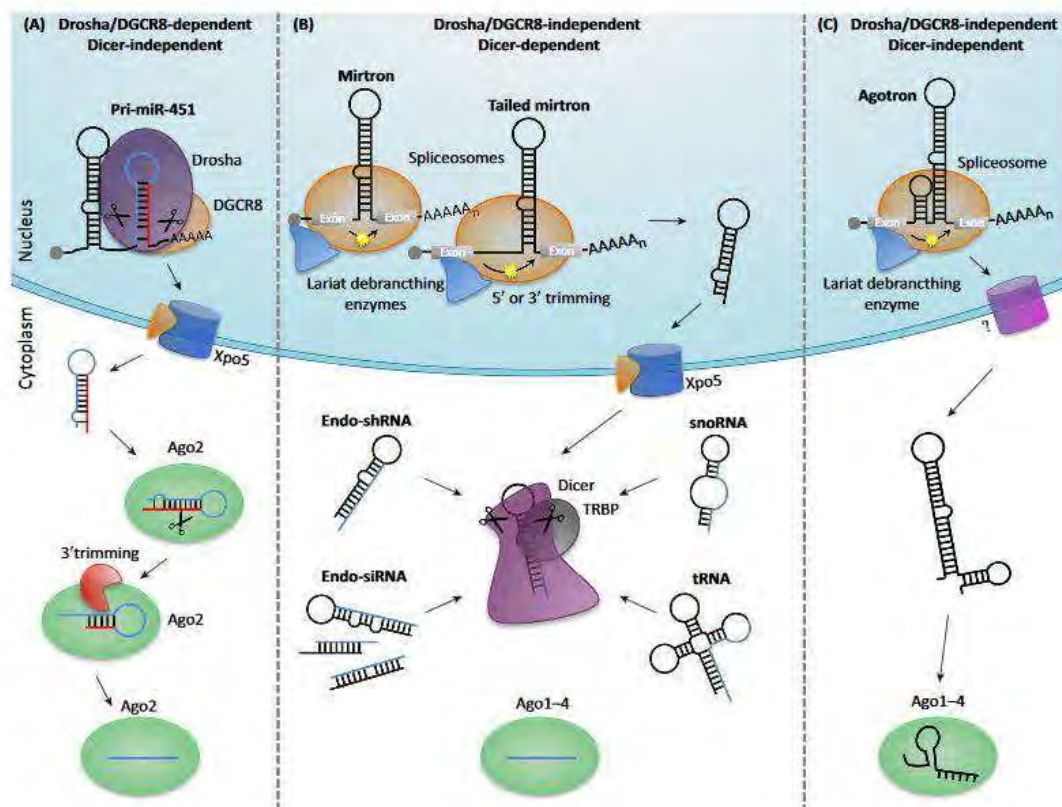
Figure 1.13: Alternative miRNA biogenesis pathways. *Drosha/DGCR8-dependent and Dicer-independent pathway. B) Drosha/DGCR8-independent and Dicer-dependent pathway. C) Drosha/DGCR8-independent and Dicer-independent pathway. Modified from Daugaard I, et al. 2017 [176].*

1.4.3 Function

MicroRNAs primarily function as post-transcriptional regulators of gene expression [179]. The loaded miRNA bound to Ago proteins acts as guides for the RISC complex towards target sequences in mRNAs promoting repression of gene expression. The consequence of this target site recognition has been described to negatively interfere with the translational machinery altering the protein production [180]. More recent studies have revealed that, in addition to repression of the translation, miRNAs can also regulate gene expression by mRNA deadenylation, destabilization and degradation, all of which result in a decrease of mRNA expression levels [181-183]. To directly cleave the target mRNA, a catalytically active Ago protein is required. In vertebrates, among Ago proteins, only Ago2 has slicer activity [184]. MiRNAs can have pleiotropic effects, targeting multiple genes and signalling pathways [185] while genes can be targeted by multiple miRNAs [186]. *In-silico* predictions suggest that more than one-third of all human genes are targets of miRNAs [187].

Canonical miRNA binding involves base pairing of the miRNA seed region to complementary target sites. This seed region corresponds to the first nucleotides 2–8 in the 5' end of miRNAs [188]. The remaining sequence does not pair perfectly with the target in mammals. MiRNAs with the same seed sequence usually share targets and the seed sequence is often used for the classification of miRNAs into families [179]. However, several functional non-canonical miRNA-target sites have been reported [189-191].

Alternative functions for miRNAs have been also described. For example, enhanced mRNA translation due to miRNA targeting has been reported [192]. Apart from mRNAs they can also target other RNA molecules such as long non-coding RNAs (lncRNAs) which are also regulators of gene expression and can function as miRNA decoy systems [193]. Some studies have reported that miRNAs might have a regulatory role in the nucleus activating gene expression by targeting promoter regions [194-196]. Also, they have been shown to recognize and guide transcription factors to their gene promoters [197], as well as interact with other proteins by sequence-specific interaction in a RISC-independent manner and inhibit their activity [198, 199].

1.4.4 MiRNAs and cancer

MiRNA expression is regulated by multiple transcription factor networks as well as epigenetic machinery. In addition, miRNAs can themselves function as repressors of epigenetic factors, generating regulatory circuits that can have a significant impact in the transcriptional landscape of the cell. In the last few years it has been shown that miRNAs play key regulatory roles in many physiological processes including cell cycle control [200], stem cell differentiation and self-renewal [201] and organogenesis [202], among many other functions. However, deregulation of these processes can result in pathogenesis. Multiple studies have correlated changes in miRNA expression profiles with different types of cancer [186, 203-205] (Figure 1.14). Moreover, an active role of miRNAs in cancer was first demonstrated in B-cell lymphoma and has been later demonstrated in many studies [206-208].
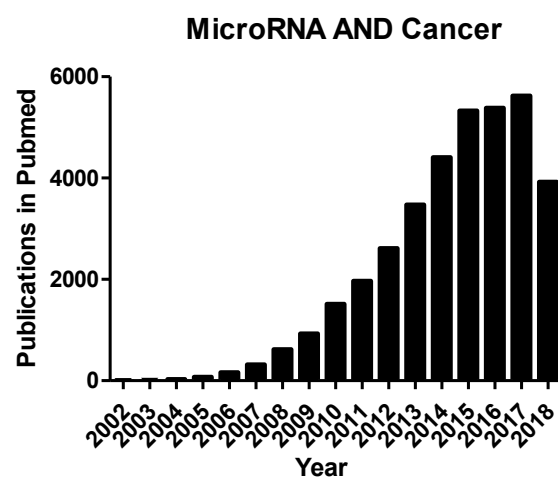
Figure 1.14: Number of publications indexed in Pubmed until 2018 per year about microRNAs and cancer.

Several miRNAs have been shown to play a role as tumour promoters which have been termed oncomiRs. For example, the *miR-17~92* cluster targets the cell cycle regulator *PTEN* in lymphocytes increasing cell proliferation [209]. Similarly, *miR-221/miR-222* can reduce CDKN1B expression and promote cell proliferation in different tumour types [210, 211]. *MiR-10b* can help tumour cells evade immune system surveillance through inhibition of class I polypeptide-related sequence B (MICB), a stress-induced ligand of the activating natural killer cell receptor KLRK1 (NKG2D) [212]. Also, *miR-10b* has been

shown to promote metastasis both *in vitro* and *in vivo* [213]. The hypoxia regulated *miR-210* is a key miRNA in angiogenesis, as it can regulate the expression of Vascular Endothelial Growth Factor (VEGF), promoting angiogenesis [214]. Deregulation of other miRNAs, such as *miR-126* or *miR-378* can also induce angiogenesis by induction of VEGF expression [215, 216].

In contrast, other miRNAs have been shown to be downregulated in cancer as they play functions as tumour suppressors. For example, *miR-34* family members are frequently downregulated in various tumours. Their expression is positively regulated by p53 and these miRNAs directly repress the expression of several targets involved in the regulation of cell cycle and in the promotion of cell proliferation and survival [217, 218]. Also, *miR-31* is able to reduce breast cancer metastasis by inhibiting extravasation and metastatic colonization of breast cancer cells [219].

MiRNAs seem to have a crucial role in B-cell differentiation and function. Indeed, the same miRNAs involved in normal B-cell physiology seem to be also involved in B-cell malignancies (see next section).

1.4.4.1 MiRNAs in B-cell development

Early B-cell development is regulated for multiple miRNAs. Many of them regulate important transcription factors for B-cell commitment. The *miR-17~92* cluster is essential for pro-B to pre-B cell transition as it downregulates the proapoptotic protein BCL2L11 (BIM) [220]. The *miR-17~92* cluster consists of six miRNAs *(miR-17, miR-18a, miR-19a, miR-20a, miR-19b-1, and miR-92-1)* processed from the same polycistronic transcript. Similarly, *miR-181a* enhances B-cell differentiation also by *BIM* targeting [221, 222]. In contrast, *miR-212/132* cluster and *miR-34a* suppress B-cell differentiation at pro-B to pre-B transition stage by repressing *SOX4* [223] and *FOXP1* transcription factors respectively, required for early B-cell development [224, 225]. Similarly, *miR-150* impairs B-cell development by targeting *c-Myb* transcription factor [226]. In the control of hematopoiesis, *miR-125b* represses *LIN28A* and directs hematopoiesis towards myeloid linage [227].

Regarding B-cell maturation in follicles, some studies revealed that *miR-155* expression is necessary for GC formation. *MiR-155* deficient mice present a decreased number of

GCs [228], suggesting that *miR-155* might have an impact on several target genes in B-cells. However, the majority of these *miR-155*-targeted genes still unknown. Some validated targets are SPI1 that encodes the transcription factor PU.1, relevant not only in B-cell commitment but also in late B-cell development. Another direct target of *miR-155* is AID enzyme, responsible for class-switch recombination (CSR) and SHM during the GC reaction [229]. Also, *miR-181b* has been shown to target AID enzyme [230]. *MiR-30* and *miR-9* seem to be necessary for GC maintenance as they target the transcription factor *PRDM1*, critical for plasma cell differentiation [231]. Similarly, *miR-125b* also represses plasma cell differentiation [232].

The *miR-17~92* cluster seems to have a potential role in plasma cells regulating the homing of these cells to the bone marrow through the downregulation of S1PR1 receptor. Homing to bone marrow allows a long-term longevity of plasma cells enabling efficient immune responses [233, 234].
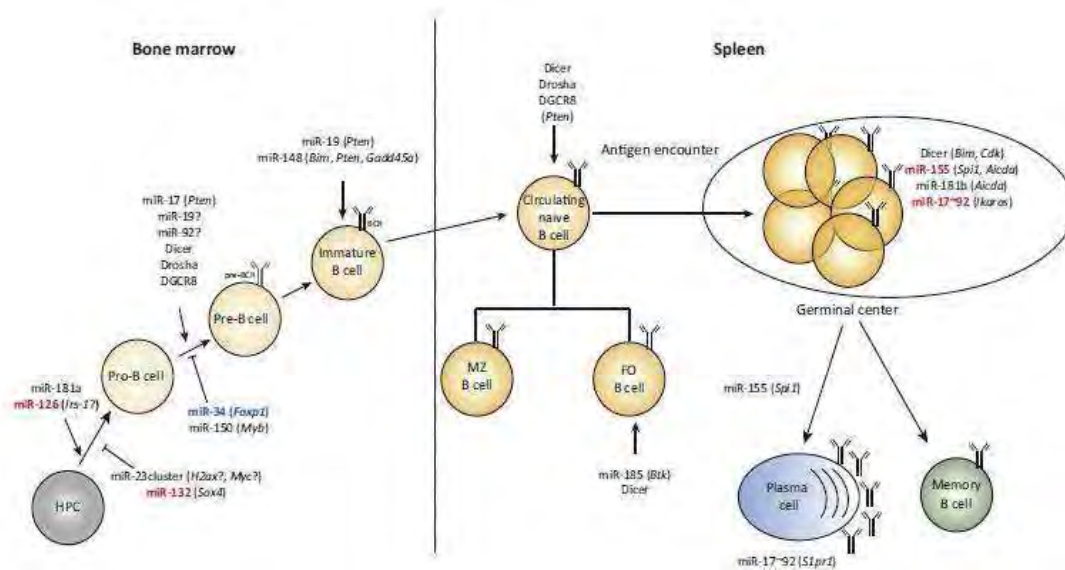
Figure 1.15: Examples of miRNA regulation in B-cell development and activation. Modified from Coffre, et al. 2017 [235].

1.4.4.2 MiRNAs in B-Cell Malignancy

Deregulation of miRNA expression has been shown to be a ubiquitous feature of cancer, and miRNA expression signatures can classify different tumour types and subtypes, including haematological malignances, to a higher degree of accuracy than

traditional GEP [236]. The first evidence of miRNAs being dysregulated in a cancer came in 2002, when a region frequently deleted in chronic lymphocytic leukaemia (CLL), the 13q14 locus, was shown to encode for the miRNAs *miR-15a* and *miR-16-1* [237]. Soon after, miRNAs were demonstrated to act as tumour promoters or so called oncomiRs, in 2005 by He *et al.*, who demonstrated that over-expression of the *miR-17~92* cluster lead to an increased tumour progression in a *c-Myc* induced B-cell lymphoma mouse model [238]. The functionally important targets of *miR-17~92* comprise pro-apoptotic proteins *PTEN* and *BIM*, which might contribute to the lymphoproliferative phenotype [209].

Apart from the *miR-17~92* cluster, the expression of other miRNAs involved in pro-B to pre-B transition has been found altered in B-cell lymphomas: low levels of *miR-34a* are correlated with poor prognosis in CLL, DLBCL and MCL. This low expression might be mediated by MYC [239, 240]. Downregulation of *miR-34a* induces the expression of *FOXP1* resulting in enhanced *in vitro* proliferation in DLBCL [239] and it is also associated with increased resistance to doxorubicin [241]. Similarly, *miR-150* acts as a tumor suppressor and also targets *FOXP1* in mature B cells [242]. In progenitor cells targets the transcription factor c-Myb is required for normal haematopoiesis and B cell development [226].

MiR-155 is one of the most studied miRNAs in B-cells. Transgenic mice overexpressing this miRNA developed high-grade lymphoma demonstrating its oncogenic function in these cells [243]. It can regulate several signalling pathways to promote cell proliferation such as TGFβ signalling pathway by targeting *SMAD5* [244], BCR signalling by targeting *INPP5D,* a negative regulator of the downstream tyrosine kinase signalling pathway [245] or JAK-STAT3 signalling pathway by targeting the negative regulator *SOCS3* [246]. *MiR-21* is highly expressed in non-Hodgkin lymphomas and its expression is correlated with poor prognosis [247, 248]. One of the targets of *miR-21* is the tumour suppressor *PTEN* [249]. Epigenetic silencing of *miR-125b* is required for normal B-cell development and upregulation of this miRNA might induce myeloid and lymphoid leukemias [250, 251].
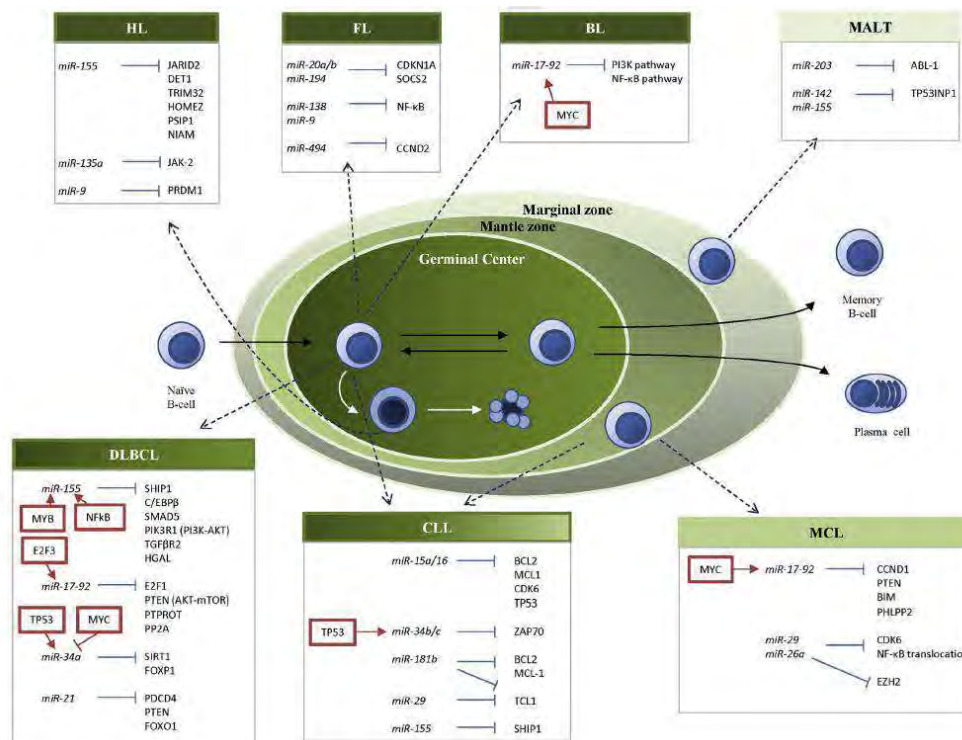
Figure 1.16: Some examples of aberrantly expressed miRNAs and their targets in B-cell lymphomas. *These miRNAs play key roles in germinal center B-cell maturation pathway under physiological conditions. Many of the dysregulated miRNAs target key molecules in this pathway and many key molecules in lymphomagenesis. The diagram demonstrates the links between lymphoma type, commonly dysregulated miRNAs and associated target gene and their respective cell of origin and position of cell of origin in the B-cell developmental pathway. Reproduced from Sole et al. 2017 [252].*

1.4.5 Factors affecting miRNA regulation

The reasons for aberrant miRNA expression in B-cell lymphoma (and other cancers) are numerous, and can include chromosomal aberrations, epigenetic deregulation, aberrant expression of transcription factors that regulate promoter regions of miRNAs, mutations in miRNA sequences and in miRNA binding sites, and the factors that change the balance of miRNA biosynthesis or function [253]. For example, *miR-17~92* cluster is transactivated by MYC, a transcription factor frequently dysregulated in cancer [254]. Dysregulation of *MYC* by chromosomal translocation and juxtaposition to immunoglobulin enhancer is a common feature in BL [255].

In addition, genetic variations affecting miRNA regulatory pathways have been described to have important implications in miRNA activity. These polymorphisms may

affect miRNA biogenesis pathways as well as their targeting capacity and are associated with pathogenesis of various cancer types [256]. Several polymorphisms affecting genes encoding components of miRNA biogenesis such as *DROSHA, DGCR8, XPO5, DICER* and *AGO2* have been frequently identified in some cancer diseases associated with risk of cancer [257], poor prognosis [258], cancer development and progression [258] and response to chemotherapy [259]. Single nucleotide polymorphisms (SNPs) have been also identified in miRNA sequences. These variations, including pri-miRNAs, pre-miRNAs and mature miRNAs, have the potential of affect the processing efficiency and/or target selection of miRNAs, leading to aberrant expression of hundreds of genes in different biological pathways [260]. However, as miRNAs are highly conserved, SNPs in miRNA genes are relatively rare.

The disruption of miRNA-dependent regulation by SNPs in the miRNA binding site of target mRNAs has been confirmed as a mechanism for altered gene expression in cancer. In contrast to the miRNA-polymorphisms in the miRNA biogenesis pathway, the polymorphisms located at the 3'-UTR of a miRNA target gene are more abundant in the human genome, and affect only the expression of the target gene and its downstream effectors, resulting in a more defined and limited range of effects [261]. These SNPs can disrupt putative miRNA-binding sites as well as create novel ones. Some examples of SNPs located in the 3'-UTR of target mRNAs and their clinical significance, are reviewed in [256].

1.4.6 *In-silico* target prediction tools

While it is interesting to identify aberrantly expressed miRNAs in cancers, in order to fully understand the functional consequences of miRNA dysregulation it is crucial to accurately identify the target genes within a defined cellular context. However, this is far from straightforward as in contrast to plants, a near-perfect Watson-Crick base pairing of miRNA to its target is rare in animals, making a challenge to predict target sites [179]. For this purpose, several different miRNA target gene predictive algorithms have been developed and take a number of different approaches to predicting the miRNA:mRNA interaction (Table 1.4).

Tool	Characteristics	Seed match	Conservation	Free energy	Site accessibility	Target-site abundance	Ref
TargetScan	Stringent seed pairing Low sensitivity Only considers targets in 3'UTR Scores the context and site conservation Includes different mRNA isoforms	x	x		x		[262] [263]
Diana Tools	Moderately stringent base pairing Searches targets in 3'UTR and CDS Considers site conservation and thermodynamics Compares results with other tools it doesn't differentiate isoforms Considers target-site abundance	x	x	x	x	X	[264] [265]
miRanda	Moderately stringent base pairing Highly sensitive Searches targets in 3'UTR and CDS Considers site conservation and thermodynamics Analyzes non-conserved sites It doesn't differentiate isoforms	x	x	x			[266]
PITA	Target-site accessibility as the major feature Highly considers thermodynamics Considers site conservation Flexible settings (also for seed matching)	x	x	x	x	X	[267]
PicTar	Requires an exact seed match Requires binding sites coregulated by multiple miRNAs across species Considers thermodynamics Considers co-expressed miRNAs targeting 30% estimated false positive rate	x	x	x			[268]
RNA22	Moderately stringent base pairing Highly considers thermodynamics It doesn't consider the site conservation Includes interactions in the whole mRNA Risk of potential false positives	x		x			[269]
RNAhybrid	Considers the free energy between a miRNA and an mRNA with a user-defined seed region Considers target-site abundance	x		x		X	[270]

Table 1.4: Characteristics of some of the most relevant web-based miRNA target prediction tools.

1.4.6.1 Seed-sequence matching parameter

The majority of the tools include the seed region as the key biological element for miRNA-target prediction [271]. Functional miRNA-target interactions can be formed from as few as 6-nt matches in a region often termed the seed region [179]. There are three 6-mers possible match positions in the seed (1-6, 2-7 and 3-8 miRNA positions) as well as two 7-mers (1-7 and 2-8 miRNA positions) and one 8-mers (1-8 miRNA position) (Figure 1.17). These canonical seed-pairing rules are the basis for the bioinformatic prediction approaches [271]. However, seed matching of such small sequences is very likely to occur by chance, so such strategies still suffer from both false-positive (40-66%) and false-negative predictions (50-70%) [272]. Furthermore, experimental methods capable of recovering RNA directly targeted by miRNAs (see next section) have observed that a great amount of the identified direct target interactions don't follow canonical seed-pairing rules [191, 273-275]. These non-canonical interactions include nucleation bulges, centered sites and seed-like motifs (mismatches in seed pairing, deletions or G:U wobble pairing) (Figure 1.18) [191, 274, 275]. Prediction models such as transitional nucleation model have been developed and implemented to include these experimentally observed non-canonical interactions (Figure 1.19) [273]. These interactions have been functionally validated and stability of Ago-complex have been confirmed [276]. These makes even more challenging miRNA target prediction also because non-canonical sites seem to have a modest sequence conservation across species [273-275].
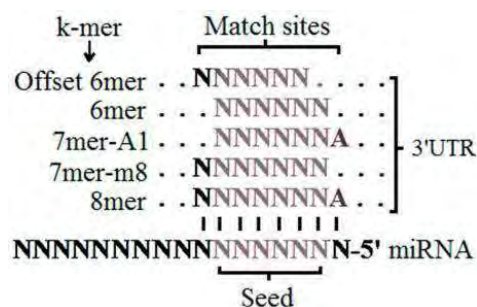
Figure 1.17: Canonical seed-matching. *Including all k-mer interactions: 8mer, 7mer-m8, 7mer-A1, 6mer and offset 6mer. Reproduced from Vlachos IS, et al. 2015 [277].*

Figure 1.18: Non-canonical miRNA target sites. *Key binding regions are highlighted in red and subtly contributing regions are in purple. Solid lines indicate Watson-Crick base pairing and dots indicate G:U wobble pairs. Modified from Seok, et al. 2016 [278].*
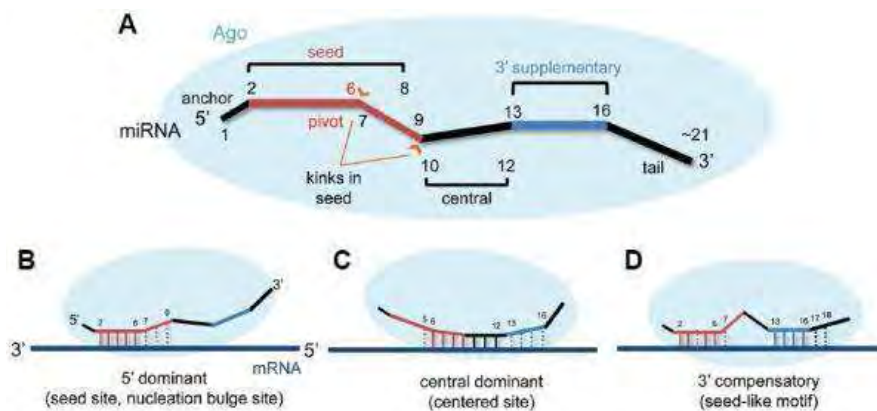
Figure 1.19: Modes of miRNA-target recognition. *(A) Schematic diagrams of functional miRNA domains structured by Ago. (B–D) Possible initial target recognition models for 5' dominant binding (seed site or nucleation bulge site, B), central dominant pairing (centered site, C), and 3'-compensatory interactions (seed-like motifs, D). Transitional nucleation pairings are indicated by bold lines. Non-Watson-Crick base pairing, mismatches, and deletions are all indicated by dotted lines. Reproduced from Seok, et al. 2016 [278].*

1.4.6.2 Conservation parameter

The degree of evolutionary conservation of the target site is an important parameter considered by many miRNA target prediction algorithms. The premise for this importance in predictive algorithms is that sites that are conserved more than would be expected by chance might be under selective pressure and therefore biologically important [262].

1.4.6.3 Free energy parameter

Thermodynamic criteria of potential miRNA:mRNA interactions are also considered by many of these computational tools. The lower the free energy, the greater the RNA:RNA binding, which increases the likelihood that this interaction occurs [279]. To measure this, most prediction tools use Vienna RNA package [280].

1.4.6.4 The context parameter

The context parameter is also considered in these prediction algorithms and refers to those factors that may affect the miRNA-mRNA interaction or accessibility to the binding site such as the binding of proteins to the RNA (for example the translational machinery in the coding regions) [281] and secondary structures of the RNA [267]. To asses this, some parameters are considered. For example, the number of AU nucleotides flanking the seed region is considered by some prediction tools such as TargetScan or miRanda and associated with a weaker mRNA secondary structure near the site, improving the accessibility to the seed site. Also, the position within the 3'UTR is important. The UTR quartiles near the stop codon and near the poly-A tail are most susceptible for an effective targeting. Secondary RNA structures are more likely to occur in the center of long 3'UTRs [282].

1.4.6.5 TargetScan

TargetScan was the first algorithm to predict miRNA targets in vertebrates and the most stringent of the sequence-based tools but with a high false negative rate [271]. It only considers seed regions at the 3'UTR and does not support mismatches except when a mismatch in the seed is compensated with a interaction in the 3' site of the miRNA [262, 282]. Also, it strongly prioritizes the conservation level of mRNA:miRNA

interactions and positively rates additional pairing of 3' portion of the miRNA and proximity of other miRNA binding sites. It also incorporates a phylogenetic tree based on UTR genomic regions to quantitatively evaluate motif conservation [262] and can give a cell type specific score for each site [283]. The parameters and values used by TargetScan are listed in Table 1.5.

Parameter	Value Range	Meaning
Site type	8mer > 7mer > 7mer-A1 > 6mer	The matching sites in the seed region (nucleotides 2 to 8 from 5' of miRNA that have perfect WC pairing with the 3' UTR), from the strictest to the least strict.
Context++ score	From 1 to −1	The sum of the contribution of 14 features for each of the four site types, the more negative the score, the greater the repression.
Context++ score percentile	From i to 100 − i;	Percentage of sites for the miRNA with a less favorable context++score.
Weighted context++ score	From 1 to −1	The scores with a lower negative value indicate a greater prediction of repression.
Cumulative weighted context++ score	$C(i{-}1) + (1 - 2CSi)(AIRi{-}C(i{-}1))$	This score estimates the total repression expected from multiple sites of the same miRNA, for each mRNA target predicted.
Branch-length score	8mer: 1.8; 7mer-m8: 2.8; 7mer-A1: 3.6; 6mer: NA	This score is the sum of phylogenetic branch lengths between species that contain a matching site.
P_{CT} score	Between 0 and 1	The higher the score, the greater the conservation and the greater mRNA destabilization expected.
Aggregate P_{CT}	Value = $1 - ((1 - P_{CT})$ site1 $\times (1 - P_{CT})$ site2	For each miRNA, this parameter includes the conserved 3' UTR targets with multiple sites that were missed in the human 3' UTR annotation but were present in the mouse annotations.
Conserved sites	≥0	Number of conserved sites identified.

Table 1.5: Description of the parameters and values shown in the results delivered by TargetScan [271].

1.4.6.6 miRanda and other tools

The miRanda algorithm is more flexible and sensitive, but with a higher false positive rate. Its scoring system considers the energy needed to access the binding site calculated using the Vienna package, while TargetScan considers the context. miRanda tool also considers the context (accessibility of RNA due to secondary structures, position of the target site, evolutionary conservation etc). However, miRanda can predict target sites of non-canonical and non-conserved sites [266, 271]. Similarly, DIANA-microT and RNA22 tools can also analyze non-conserved sites as well as located in coding regions and all along the mRNA in case of RNA22 [271]. Other programmes including seed searching in 5'UTRs and coding sequences are miRTar and TargetS tools [269, 284, 285]. The parameters and values used by miRanda and DIANA-microT tools are listed in Table 1.6.

Tool	Parameter	Value Range	Meaning
miRanda	mirSVR score	<0	This score is an estimate of the miRNA effect on the mRNA expression level. The more negative the score, the greater effect.
	PhastCons score	From 0 to 1	This measures the conservation of nucleotide positions across multiple vertebrates.
DIANA Tools	miTG score	From 0 to 1	This is a general score for the predicted interaction, the closer to 1, the greater the confidence.
	Also Predicted	red, blue and green	This compares with other tools; miRanda in red, TargetScan in blue and TarBase in green.
	Region	UTR3, CDS	Region of the mRNA where the interaction occurs.
	Binding Type	6mer; 7mer; 8mer; 9mer; miRNA bugle	The matching sites between the miRNA and the mRNA.
	Score	From 0 to 1	It is the site contribution score in the miTG score.
	Conservation	≥0	Number of species in which the predicted interaction is conserved.
	Signal-to-noise ratio (SNR)	>0	This score is a measure of the "signal to noise" ratio, which enables the identification of the miTG score of each interaction without background noise.
	Precision	From 0 to 1	This score is an indicator of the false-positive rate in a miTG interaction.

Table 1.6: Description of the parameters and values shown in the results delivered by miRanda and DIANA-microT tool [271].

1.4.6.7 Limitations of miRNA target gene prediction tools

These tools, even if based on common features, show differences when predicting miRNA binding sites. The reasons can be the use of different databases for mRNA and miRNA sequences. However, other reasons are intrinsic to the prediction algorithm and their scoring system (Figure 1.20).
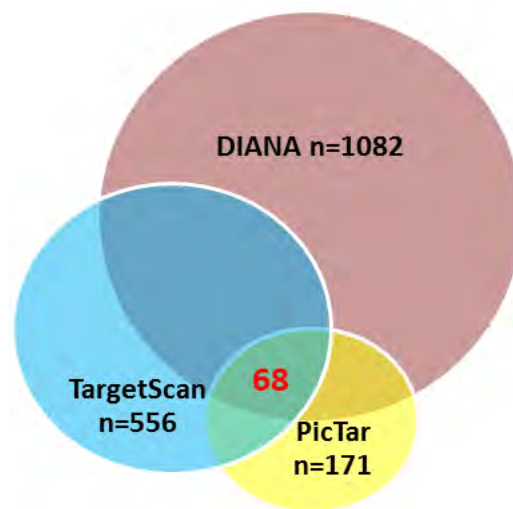
Figure 1.20: Number of miR-155 target genes according to three different miRNA target prediction tools: The figure shows the number of has-miR-155 target genes identified using the default parameters with TargetScan v.7.2, DIANA tools (microT-CDS algorithm v.5.0) and PicTar (last update: March 26, 2007)

As figure 1.20 shows, the sensitivity and specificity is very variable when comparing different target prediction tools. For some applications they might be a rapid and useful method for exploring putative miRNA binding sites. However, the experimental validation of these predictions often fails as the exact rules that govern miRNAs-binding mechanism have not been completely elucidated. Also, the targets of miRNAs among different cells lineages and tissues can vary making *in-silico* prediction more difficult [283, 286]. Therefore, predictions really need to be experimentally validated.

1.4.7 Experimental identification of miRNA target sites

MiRNA deregulation is a commonly observed feature in multiple diseases including cancer. However, in order to understand the meaning of this aberrant expression is necessary to know the specific targets of this miRNA within a specific cellular context. Multiple high-throughput experimental approaches have been used for this purpose, from microarray analysis to RNA sequencing and proteomic analysis [287, 288]. However, these techniques are not able to differentiate between indirect and direct miRNA targeting. Therefore, methods to experimentally identify direct miRNA targets have been developed. These methods identify direct protein-RNA interactions.

One of the first techniques used for this purpose was RNA immonoprecipitation (RIP). This method is based on the immunoprecipitation (IP) of a specific RNA-binding protein (RBP) under conditions that preserve ribonucleoprotein complexes (RNPs), and the subsequent analysis of the purified RNAs that were interacting with the RBP using microarrays (RIP-Chip) or high-throughput sequencing (RIP-Seq) [289, 290]. However, this methodology was not suitable for direct protein-RNA contacts as it preserves protein-protein interactions and thus can purify multiple RBP and the RNA bound to any of these proteins [291].

Later on, a more accurate method for the identification of direct protein-RNA interaction called CLIP (crosslinking and immunoprecipitation) was developed. The CLIP method relies on the purification of short RNA fragments that crosslink to a specific protein and then identifies these fragments using sequencing. This method and similar techniques have been widely used for the study of RNA and interactions with various RBPs [292, 293].

Most variants of CLIP use ultraviolet (UV) light to form covalent bonds between RBPs and their cognate bound RNA. UV crosslinking requires direct contact between an amino acid and a nucleic acid, and therefore ensures that only direct protein-RNA interactions are preserved and allows more stringent purification steps to be used as the covalent bond is retained throughout the method [294]. The original CLIP protocol and most of its variants use UV-C wavelength (254 nm) [295]. However, a CLIP variant called PAR-CLIP (Photoactivatable ribonucleoside-enhanced CLIP) introduces a variation

on this crosslinking strategy. In PAR-CLIP, cells are pre-incubated with a photoactivatable ribonucleoside, 4-thiouridine (4-SU) or 6-thioguanosine (6SG), which is incorporated into the nascent nucleic acids, and enables protein-RNA crosslinking with UV-A wavelength (365 nm) [296].

For the identification of miRNA-binding sites, IP against Ago proteins are performed [297]. In 2009, Chi et al reported in Nature journal the first work where Argonaute HITS-CLIP (High-throughput sequencing of RNA isolated by CLIP) was applied in order to analyze interactions between miRNAs and mRNA [298]. Since then, multiple works have been published studying interactions between miRNAs and their target sequences using these technologies such as CLIP-Seq, HITS-CLIP and PAR-CLIP [278, 299, 300]. Existing CLIP related protocols are listed in Table 1.7.

Acronym	Full name	Ref
CLIP	(UV) Crosslinking and immunoprecipitation	[301]
Fractionation CLIP	CLIP from nucleus, cytosol. and polysomes	[302]
HITS-CLIP	High-throughput sequencing of RNA isolated by CLIP	[303]
CLIP-Seq	CLIP coupled with high-throughput sequencing	[304]
CRAC	UV crosslinking and analysis of cDNAs	[305]
PAR-CLIP	Photoactivable ribonucleoside-enhanced CLIP	[306]
iCLIP	Individual-nucleotide resolution CLIP	[307]
CLAP	Crosslinking and affinity purification	[308]
4-SU-iCLIP	4-SU-mediated crosslinking followed by iCLIP	[309]
urea-iCLIP	iCLIP with denaturing purification	[309]
BrdU CLIP	Bromodeoxyuridine UV CLIP	[310]
FAST-iCLIP	Fully automated and standardized iCLIP	[311]
irCLIP	Infrared-CLIP	[312]
eCLIP	Enhanced CLIP	[313]
seCLIP	Single-end eCLIP	[314]
uvCLAP	UV crosslinking and affinity purification	[315]
FLASH	Fast ligation of RNA after some sort of affinity purification for high-throughput sequencing	[315]
Fr-iCLIP	Fractionation iCLIP	[316]
sCLIP	Simplified CLIP	[317]
dCLIP	Denaturing CLIP	[318]
CLASH	Crosslinking, ligation, and sequencing of hybrids	[319]
hiCLIP	RNA hybrid and iCLIP	[320]
PAPERCLIP	Poly(A) binding protein-mediated mRNA 30 end retrieval by CLIP	[321]
cTag-PAPERCLIP	''Conditionally'' tagged PAPERCLIP	[322]
m5C-miCLIP	Cytosine-5 methylation iCLIP	[323]
m6A-miCLIP	N6-methyladenosine iCLIP	[324]

Table 1.7: List of CLIP and related protocols ordered by their historical development. Updated publications introducing important variations to the same method are grouped with the initial publication. Protocols that are not aimed at studying the specificity of an RBP but that apply the CLIP technology to a new purpose are listed at the end [294].

The PAR-CLIP protocol in particular has several advantages over other CLIP techniques using UV 254 nm crosslinking in absence of photoreactive ribonucleoside analogs. PAR-CLIP facilitates RNA-protein crosslinking and transcriptome-wide identification of RBP and RNP binding sites as it gets a higher yield of crosslinked RNA using similar radiation intensities. Also, and more importantly, the UV 365 nm crosslinking reaction between

the photoactivatable nucleoside and the interacting proteins promotes a specific mutation (T/C using 4-SU and G/A using 6SG) during the cDNA synthesis. These mutations can be identified by NGS facilitating the identification of specific crosslinked regions which is useful to discriminate background noise (Figure 1.21) [296, 306].
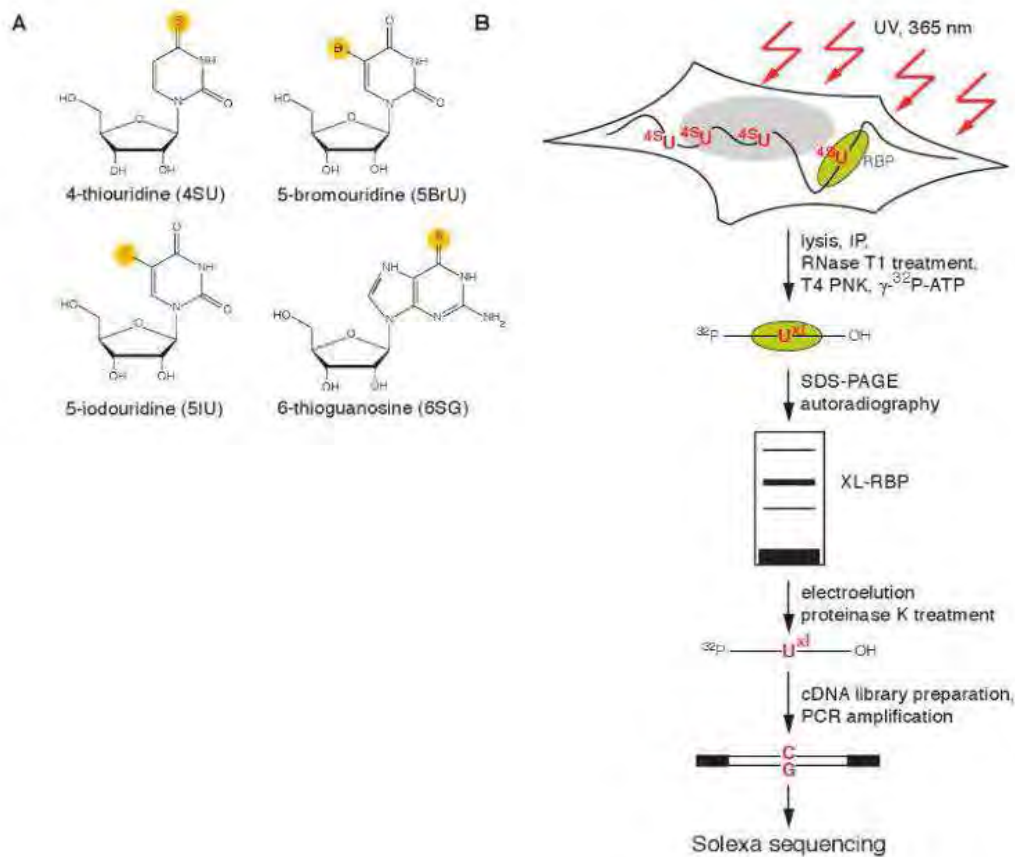
Figure 1.21: PAR-CLIP methodology. (A) Structure of photoactivatable nucleosides (B) Diagram showing PAR-CLIP protocol. 4-SU-labeled transcripts were crosslinked to RBPs and partially RNase-digested RNA-protein complexes were immunopurified and size-fractionated. RNA molecules were recovered and converted into a cDNA library and deep sequenced. Modified from Hafner, et al. 2010 [296].

Overall, CLIP methodologies have given the opportunity to better understand the function and the mechanism of these regulatory molecules. Some studies have analyzed the whole targetome *in vivo* and *in vitro* [325, 326]. Others have experimentally identified targets of specific miRNAs in defined cellular contexts [274] as well as analyzed motives of miRNA-binding [296] allowing to increase the knowledge about the binding mechanism of these molecules to specific targets.

2. HYPOTHESIS AND OBJECTIVES

2.1 Hypothesis

The hypothesis of this thesis is that the binding sites of miRNAs are important in the pathogenesis of non-Hodgkin B-cell lymphomas. We will test this hypothesis in two ways; firstly, by identifying mutations in the miRNA-binding sites of tumour material and investigating the functional relevance of these mutations. Secondly, we will use the PAR-CLIP technique to elucidate the lymphoma-specific targetome associated with a commonly dysregulated miRNA that is central to lymphomagenesis (i.e. *miR-155*).

2.2 Objectives

• Identify and analyze the functionality of mutations in miRNA binding sites as a possible pathogenic mechanism of FL

• Determine the biological relevance of mutations in miRNA binding sites and their association with FL transformation to high grade disease

• Assess whether mutations in miRNA binding sites could be useful as prognosis biomarkers to identify FL patients at risk of transformation to high grade disease

• Identify the targetome of *miR-155* in GCB-DLBCL and ABC-DLBCL subtypes in order to determine the biological relevance of this distinct miRNA expression pattern.

• Identify biomarkers for classification of DLBCL patients in the GCB and ABC molecular subtypes through the identification of the true targetome of *miR-155*

3. MATERIALS AND METHODS

3.1 Extraction of nucleic acids

3.1.1 Genomic DNA extraction

Cultured cells: DNA of lymphoid cell lines was extracted from $2x10^6$ to $5x10^6$ cells pellets with Nucleospin® Tissue Kit (Macherey-Nagel) according to manufacturer's instructions. This kit uses SDS and proteinase K for lysis and DNA binding is based in silica membrane technology. DNA was eluted in 50 µL of DNase/RNase free H_2O, quantified with NanoDrop®ND-1000 spectrophotometer (ThermoFisher Scientific, DE, USA) and stored at -20ºC.

Frozen tumour biopsies: DNA was extracted from frozen tumour material preserved in O.C.T compound. 3 cuts of 10 µm thick were cut from each sample and DNA was extracted with QIAmp DNA Mini Kit (QIAGEN, Venlo, Netherlands) according to manufacturer's instructions. This kit uses enzymatic lysis and DNA binding is based on silica-gel membrane technology. DNA was eluted in 200µL H_2O and quantified with NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific, DE, USA). DNA was stored frozen at -20ºC.

Tumour biopsies embedded in paraffin: Slices of tumour material embedded in paraffin were cut manually with HM355s microtome (Thermo Scientific, DE, USA). First, cuts from paraffin samples were discarded in order to avoid RNA degraded material. Then, 3 cuts of 10 µm thick were cut from each sample and collected. Total genomic DNA was isolated with QIAmp® DNA FFPE Tissue kit (QIAGEN, Venlo, Netherlands) according to manufacturer instructions. DNA was eluted in 30 µL of DNase/RNase free water (Thermo Fisher Scientific, DE, USA). DNA concentration and purity were determined with NanoDrop® ND-1000 spectrophotometer (ThermoFisher Scientific, DE, USA) and stored at -20ºC.

3.1.2 Plasmid DNA extraction

Plasmid DNA isolation for luciferase cloning was performed using the miniprep method. This method consists of a first bacterial lysis and elimination by centrifugation of bacterial walls and membranes together with the bacterial chromosomic DNA followed by plasmid DNA isolation from the bacterial lysate. For this purpose, 5 mL of

LB broth with 100µg/mL ampicillin were inoculated and grown at 37ºC 180rpm overnight (o/n). Bacterial cells were pelleted and plasmids were extracted using QIAGEN Plasmid Mini Kit (QIAGEN, Venlo, Netherlands) following manufacturer's instructions. DNA was eluted in 50 µL of DNase/RNase free H_2O, quantified with NanoDrop® ND-1000 spectrophotometer (ThermoFisher Scientific, DE, USA) and stored frozen at -20ºC. Similarly, for isolating the enveloping plasmids for lentiviral production a Maxiprep method was used which has the same principle as Miniprep method but optimized for getting a higher yield. This time QIAGEN Plasmid Maxi Kit (QIAGEN, Venlo, Netherlands) was used, starting from an o/n grown bacterial culture of 250 mL following manufacturer's instructions.

3.1.3 Total RNA extraction

Cultured cells: Total RNA was extracted using Tri Reagent® (Ambion). Cell pellets were resuspended in 1 mL of Tri Reagent® and incubated 10 minutes at room temperature (RT). 200 µL of chloroform were then added to each samples and vortex was applied for around 10 seconds. Samples were incubated 10 minutes at RT and centrifuged at 12000g 4ºC for 10 minutes. Aqueous phase was transferred to new 1.5ml tubes and 1µL of glycogen (5 µg/µL) followed by 500µL of isopropanol were added and mixed by inversion. After 10 minutes of incubation samples were centrifuged again at the same conditions as in the previous centrifugation. Resulting RNA pellet was washed with cold 75 % ethanol and after ethanol removal samples were left to dry for 10 minutes applying vacuum. Finally, pellets were resuspended in 30 µL of DNase/RNase free H_2O, quantified with NanoDrop® ND-1000 spectrophotometer (ThermoFisher Scientific, DE, USA) and stored at -80ºC.

3.2 PCR amplification

For the PCR amplifications primers were design using Primer3 Input (version 0.4.0) online tool. For the PCR amplification reaction 50-100 ng of template DNA, depending on the quality of the sample, were added to a final volume of 25uL of the PCR reaction mix (2.5µL Buffer 10X, 1µL $Mg^{2+,}$ 0.5µL dNTPs, 0.5µL Fwd primer (10µg/µL), 0.5µL Rev

primer (10µg/µL), 0.2µL BIOTAQ™ DNA polymerase and DNase/RNase free H$_2$O up to 25µL) (BIOLINE, London, UK). PCR reaction was carried out in a Verity Thermal Cycler (Applied Biosystems™). The PCR programme slightly varied according to the pair of primers used but for all the cases followed this scheme.

Denaturalization	95ºC 5 minutes	
Denaturalization	95ºC 5 minutes	
Annealing	50-66ºC 30 seconds	30-35 cycles
Elongation	72ºC 1 minute	
Final elongation	72ºC 10 minute	
	4ºC ∞	

3.3 DNA gel electrophoresis

Agarose gels at varying concentrations of 0.8-1.5 % were prepared to effectively separate DNA fragments according to their size. Agarose (Ecogen, Barcelona, Spain) was mixed with 1xTBE buffer (Bio-Rad, California, USA) and heated into a microwave until agarose was completely dissolved. Then, Gel Red dye (Biotium, California, USA) was added to the gel (1:10000 dilution) to enable visualization of DNA bands under UV light. Along with loaded samples, ΦX174 DNA-Hae III ladder (New England Biolabs, MA, USA) was used for expecting bands up to 1kb and 1kb DNA Ladder (New England Biolabs, MA, USA) for higher bands.

3.4 Sanger Sequencing

3.4.1 PCR purification

After amplifying the desired fragments by PCR, products were cleaned with ExoSAP-IT™ (Thermo Fisher Scientific, DE, USA). Briefly, 5µL of PCR product were mixed with 2µL ExoSAP and incubated 15 minutes at 37ºC followed by 15 minutes at 80ºC. Purified products were diluted 1:10 adding 63 µL of DNase/RNase free H$_2$O before proceeding

with Sanger Sequencing reaction. Alternatively, for some cases PCR clean-up was made by extracting the DNA fragment directly from the gel and performing a silica-membrane-based purification using QIAquick Gel Extraction Kit (QIAGEN, Venlo, Netherlands).

3.4.2 Sanger Sequencing reaction

For Sanger sequencing reaction 7μL of the purified PCR product, or alternatively 100ng of an isolated plasmid, were mixed with 3,25μL of a reaction mix containing 1,5 μL of the sequencing primer, 1,25 μL of the reaction buffer and 0,5 μL of the Big Dye (Thermo Fisher Scientific, DE, USA). The PCR conditions used were: 3 minutes at 94ºC, 25 cycles of 10 seconds at 96ºC, 5 seconds at 50ºC and 4 minutes at 60ºC and a final cooled down to 4ºC. Then, the sequencing reaction product was purified using Optima DTR™ (EdgeBio, Maryland, USA) and finally samples were injected in an ABI PRISM 3130 genetic analyzer (Thermo Fisher Scientific, DE, USA) for the capillary electrophoresis. The analysis was made using Mutation Surveyor® (SoftGenetics, PA, USA) software.

3.5 Reverse Transcription (RT)

3.5.1 miRNA RT

Complementary DNA (cDNA) was obtained with TaqMan®MicroRNA Reverse Transcription Kit (Applied Biosystems™, California, USA starting from 50ng of total RNA using specific miRNA-RT oligonucleotides, accordingto manufacturer's instructions. RT reaction was carried out in a Verity Thermal Cycler (Applied Biosystems™, California, USA).

3.5.2 Total RNA RT

Complementary DNA (cDNA) was obtained with High Capacity Reverse Transcription Kit (Applied Biosystems™, California, USA) starting from 500ng of total RNA using random primers, according to manufacturer's instructions. RT reaction was carried out in a Verity Thermal Cycler (Applied Biosystems™, California, USA).

3.6 Quantitative polymerase chain reaction (qPCR)

3.6.1 miRNA qPCR

The amplification reactions were performed in triplicate using TaqMan specific miRNA probes in a CFX Connect Real-Time PCR Detection System (Bio-Rad, California, USA) qPCR machine. Relative expression quantification was calculated using 2-ΔCt method normalizing Ct values of genes with Ct values of housekeeping controls (RNU6b and RNU48).

3.6.2 mRNA qPCR

The amplification reactions were performed in triplicate using 20 ng of cDNA as template with TaqMan specific miRNA probes in a CFX Connect Real-Time PCR Detection System (Bio-Rad, California, USA) qPCR machine. Relative expression quantification was calculated using 2-ΔCtmethod normalizing Ct values of genes with Ct values of the housekeeping control TATA binding protein (TBP).

3.7 Production of competent bacteria

Competent bacteria were produced in-house from XL1-blue Escherichia Coli strain using a $CaCl_2$ and glycerol-based treatment. To do that, XL1-blue bacteria were grown o/n in 6mL LB broth (Conda, Madrid, Spain) at 37ºC and 180rev/min in a shaker Incubator OPAQ (OVAN, Barcelona, Spain). The following day the 6mL of bacterial culture were added to 300mL of LB broth and were grown until the OD_{600} was between 0.5 and 0.8 so they still in the exponential growing phase. To perform these measurements 1 mL of the culture was removed and measured in a GeneQuant pro spectrophotometer (Amersham, Little Chalfont, UK). Consecutively, bacteria were poured in 50mL disposable centrifuge tubes and kept on ice for 10 minutes. Bacteria were then centrifuged at 2600 rpm, at 4ºC for 10 minutes and supernatant discarded. Each 50mL tube with bacteria were resuspended in 2mL of sterile 100mM $CaCl_2$ and

15% glycerol (Panreac, Barcelona, Spain) and 350µL were aliquoted in 1,5 centrifuge tubes. Finally, they were stored at -80ºC.

3.8 Transformation of plasmid DNA into competent bacteria

For plasmid transformation into competent bacteria, XL1-Blue competent bacteria were defrosted in ice, mixed with the plasmid or ligation reaction and incubated for 30 minutes on ice. To promote the transformation a heat shock of 45 seconds at 42ºC was applied to the bacterial vial using a block heater (Thermo Fisher Scientific, DE, USA). Bacteria were then immediately placed on ice. 250µL of warm LB broth were added to bacteria and they were incubated for one hour at 37ºC in agitation (200rev/min) in a shaker Incubator OPAQ (OVAN, Barcelona, Spain). Finally, bacteria were seeded in LB agar (Conda, Madrid, Spain) plates with 100µM ampicillin (Gibco™, Thermo Fisher Scientific, DE, USA) and incubated o/n at 37ºC in an incubator (Selecta, Barcelona, Spain).

3.9 Glycerol Stocks

For long storing of the desire bacterial clones glycerol stocks were maid and stored at -80ºC. Bacteria clones were individually picked from the LB agar plates and grown o/n at 37ºC in agitation in LB broth (Conda, Madrid, Spain) with 100µM ampicillin (Gibco™, Thermo Fisher Scientific, DE, USA). 500 µL of the o/n culture were mixed with 500 µL of sterile 50% glycerol (PanReac AppliChem, Barcelona, Spain) in a 2 mL cryovial and placed immediately at -80ºC for storage.

3.10 Cloning into luciferase psiCHECK-2 vector

PCR amplicons were purified using NucleoSpin Gel and PCR clean-up (Macherey-Nagel) kit and eluted in 25uL of RNase and DNase free water. DNA fragments along with the psiCHECK-2 vector were digested separately with PmeI enzyme (New England Biolabs) for 1.5 hours and NotI enzyme (New England Biolabs) for 1.5 hours. Digested DNA

fragments were purified by running them in an agarose 1.5X electrophoresis gel followed by an extraction of the DNA from the gel (NucleoSpin Gel and PCR clean-up Macherey-Nagel kit). DNA was eluted in 20uL of RNase and DNase free water. The digested fragments and vector were ligated using T4 DNA Ligase (New England Biolabs) for one hour at room temperature as follows:

Digested DNA Template	15µL
Digested psiCHECK-2 plasmid	2µL
Buffer 10X	2µL
T4 DNA Ligase	1µL
	20uL total volume

Half of the ligation product (i.e. 10uL) was transformed into competent bacteria (see section 3.8) and between 50-200uL of these bacteria were plated into LB plates with 100µg/mL of Ampiciline and incubated at 37ºC O.N. Resistant colonies were transferred onto a new LB Amp plate and incubated again at 37ºC O.N. The presence of the correct insert was checked by colony PCR. Colonies were picked, put into 50uL of distilled water, and then heat boiled at 95ºC for 5 minutes. 2uL were used as the template for the PCR using vector-specific primers (Table 3.1). These primers were design covering the cloning region of the vector so that the presence or absence of the insert can be distinguished by the size of the insert.

Oligo name	Sequence
Psicheck-2-Fwd	5'- AGGACGCTCCAGATGAAATG-3'
Psicheck-2-Rev	5'- GCGAGGTCCGAAGACTCAT-3'

Table 3.1: Primers for colony PCR.

3.11 Cell culture

3.11.1 Adherent cell lines

HEK293 cell line (human embryonic kidney cell line) was cultured in order to perform luciferase assays and for lentiviral production and titration. These cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% of Fetal Bovine Serum (FBS), 1% (or 2mM) L-Glutamine and 1% Penicillin-Streptomycin (Gibco™, Thermo Fisher Scientific, DE, USA).

For the PAR-CLIP optimization, FLAG-HA-Ago2 HEK293 engineering modified cell line was kindly facilitated by Markus Landthaler from RNA Biology and Posttranscriptional Regulation group, Max-Delbruck-Centre for Molecular Medicine, Berlin, Germany. This cell line that was modified to have an inducible Ago2 expression system was cultured in high glucose Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% of Fetal Bovine Serum (FBS), 1% (or 2mM) L-Glutamine and 1% Penicillin-Streptomycin (Gibco™, Thermo Fisher Scientific, DE, USA) together with blasticidin (15µg/mL) and hygromycin B (100µg/mL) (Sigma-Aldrich, Missouri, USA). For inducing Ago2 exogenous expression cells were cultured o/n supplementing the media with 1µg/mL of Doxycycline (Sigma-Aldrich, Missouri, USA).

3.11.2 Suspension cell lines

Diffuse Large B-cell Lymphoma (DLBCL) cell lines were cultured to use them as *in vitro* lymphoma models using Roswell Park Memorial Institute Medium (RPMI) supplemented with 10% of Fetal Bovine Serum (FBS) 1% (or 2mM) L-Glutamine and 1% Penicillin-Streptomycin (Gibco™, Thermo Fisher Scientific, DE, USA). These cells are: RIVA (ABC subtype), SU-DHL-10 (GCB subtype), HT (GCB subtype) and WSU-DLCL2 (GCB subtype, transformed from follicular).

3.11.3 Cryopreservation

For cell cryopreservation 2-5 million cells (depending on the cell type) were spin down, resuspended in 1mL Fetal Bovine Serum (FBS) with 10% DMSO and disposed in cryogenic tubes. Cells were cool down up to -80ºC decreasing temperature progressively, at a cooling rate of 1ºC per minute. After reaching this temperature cells were long-term stored in liquid nitrogen (-180ºC).

3.12 Lentiviral production and titration

For lentiviral production HEK293T cells were seeded in T75 flasks. At a 70% confluence they were transfected with the packaging plasmids together with the transfer vector using TurboFect™ Transfection Reagent (ThermoFisher Scientific, DE, USA). After 16-18 hours media was replaced with 10mL of fresh media and incubated until the following day when the first harvest of viruses was collected. To do that, the media covering the cells was collected, centrifuged to pellet down cell debris, filtrated and aliquoted. 10mL of fresh media were added to the cells and the second harvest of viruses was collected the following day following the same steps.

For viral titration, HEK293T cells were seeded (100.000 cells per well in p12 plates) and infected with different volumes of the recovered supernatant (500uL, 100uL, 50uL, 20uL, 10uL, 2uL) for the first and the second harvest separately. Cells were incubated for 72 hours and then dissociated using trypsin and interrogated using flow cytometry (FACSDiva Version 6.1.3). The percentage of green cells was calculated per each condition. Results with the different infections volumes from the same harvests were plotted together. The linear phase was identified and a regression line was applied in order to extrapolate the amount of volume necessary for infecting 100% of the cells. Assuming that every viral particle infects one single cell and taking into account the total amount of cells seeded the amount of viral particles per microliter was estimated (IFUs/µL).

3.13 Western-Blot

3.13.1 Protein extraction and quantification

Cell lysates were obtained from cell pellets using RIPA lysis buffer supplemented with protease and phosphatase inhibitors. Cell pellets were resuspended in the lysis buffer and incubated in ice for 30 min applying vortex every 5-10 minutes. Samples were centrifuged at 1200g for 10min at 4ºC and supernatants were collected and stored at -80ºC. To quantify the total protein content of the lysates Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific, DE, USA) was used according to manufacturer's instructions.

3.13.2 SDS-PAGE and Membrane Transference

For protein electrophoresis between 20-50 µg of total protein were loaded into in-house made polyacrylamide gels. Gels were made using 30% acrylamide/bis-acrylamide, 37.5:1 reagent (Bio-Rad, California, USA). Previously, loading buffer was added to the samples followed by an incubation step at 95ºC for 5 minutes to promote the denaturalization of the proteins. Along with loaded samples, Precision Plus Protein™ Kaleidoscope™ Prestained Protein Standards (Bio-Rad, California, USA) was used. After SDS-PAGE, the content was transferred from the gels into nitrocellulose membranes (Amersham Protran Premium 0.45 NC, GE Healthcare Life Sciences).

3.13.3 Antibody staining and development

Membranes were blocked with 5% Skim Milk diluted in TBS-Tween 0.1% or in 5% BSA (Bovine Serum Albumin, Sigma-Aldrich) diluted in TBS-Tween 0.1% depending on the specifications of the primary antibody. Membranes were incubated in the same solution at 4ºC o/n with the primary antibody (diluted according to the specifications of the antibody) in rotation in 50mL Falcon tubes. Three washes of 5 minutes in TBS-Tween 0.1% were performed followed by an incubation with the secondary antibody (conjugated with HRP) diluted 1:2000 in 5% Skim Milk TBS-Tween 0.1% for 1 hour at room temperature in rotation in 50mL Falcon tubes. Three washes of 5 minutes in TBS-Tween 0.1% were performed and membranes were developed using Novex™ ECL Chemiluminescent Substrate Reagent Kit (ThermoFisher Scientific, DE, USA).

4. INVESTIGATING MUTATIONS IN THE MIRNA BINDING SITES OF FOLLICULAR LYMPHOMA

4.1 Background

Follicular lymphoma (FL) is the most common form of low-grade B-cell lymphoma accounting for approximately 20% of all diagnosed lymphomas [5]. Due to its slow progressive nature FL is often considered an incurable malignancy, with a median life expectancy of approximately 12 years [48]. However, up to 45% of FL patients undergo high-grade transformation into a much more aggressive form of lymphoma with a poorer outcome, generally termed as transformed FL (tFL) [49, 89]. Many efforts have been made to understand the molecular mechanisms leading this process. For example, somatic alterations involved in transformation of FL have been widely studied. Alterations in genes involved in cell-cycle progression and DNA damage responses (e,g, *CDKN2A/B, TP53* and *MYC*) [91], as well as mutations the affect the NF-κB signalling pathway such as *MYD88* and *TNFAIP3* have also been attributed to transformation [82]. In addition, mutations in *BCL2* that have been shown increased at transformation and correlate with a higher risk of mortality [96]. However, despite the high frequency of this event, the transformation process is only poorly understood at the molecular level, and as yet no predictive biomarkers exist for patients at risk from this phenomenon.

Therefore, the goal of this project was to investigate whether mutations in miRNA binding sites are associated with the transformation process as well as determine if they could be useful as biomarkers to predict this phenomenon. It has been postulated that miRNAs play a role in the transformation process. Some high-throughput studies have shown differing miRNA profiles associated with FL clinicopathological features [327-329], including transformation [236]. The latest study identified six miRNAs (*miR-223, 217, 222, 221* and *let-7i* and *7b*) that could predict transformation in FL patients suggesting that miRNA expression signatures have potential as prognostic markers and may be used to identify FL patients at risk of high-grade transformation. Also, an independent study identified *miR-31* and *miR-17-5p* levels to change during FL transformation [330]. However, to date no one has investigated the role of mutations associated with miRNAs in FL transformation.

4. Investigating mutations in the miRNA binding sites of follicular lymphoma

Recurrent mutations in *miR-142* in diffuse large B-cell lymphomas and follicular lymphomas, and editing of the *miR-376* cluster have been reported, providing evidence for microRNA editing in lymphomagenesis [331]. Also, a somatic mutation in *E2F1* gene within a *miR-136-5p* binding site was demonstrated to impair gene regulation leading to an increase in the gene activity of colorectal cancer [332]. Similarly, a mutation in *CFTR* gene showed to increase the affinity for *miR-509-3p in-vitro* reducing the expression of the CFTR protein [333]. Several additional studies have demonstrated that nucleotide variants in miRNA sequences and miRNA binding sites are associated with susceptibility to certain cancer types such as breast cancer [334-336], gastric cancer [337], hepatocellular carcinoma [338] and also NHL [339]. In NHL they have also been associated with survival [340, 341]. Computational studies have also been performed to predict disruption of miRNA binding sites by somatic mutations in cancer genomes [342]. A database called SomamiR has been created to collect somatic mutations that may create or disrupt miRNA target sites in cancer as well as germline and somatic mutations that have been experimentally shown to impact miRNA function and have been associated with cancer [343].

To test our hypothesis, we used whole genome sequencing (WGS) data from longitudinally obtained samples from 6 FL patients that underwent transformation. This data was produced by the group of Jude Fitzgibbon in the Haematological Centre of Barts Cancer Institute in London, and data related to somatic coding mutations was published in Nature Genetics [82]. We were kindly allowed to use this data for this study, in addition to using some of their FL samples for this study. Using this data, we explored the presence of somatic mutations in computationally predicted miRNA binding sites and evaluated the prevalence of the identified mutations in an extended cohort of 55 patients using targeted panel sequencing. Identified recurrent variants were selected as candidates and *in-vitro* functional experiments were performed in order confirm these bioinformatic predictions and also to experimentally assess the effect of the mutations in the miRNA binding. Finally, this mechanism was evaluated in lymphoid cell models and FL patients.

4.2 Methodology

4.2.1 Patient samples

A total of 214 tumour samples from 181 patients were used in this study. From the 181 patients, 174 were initially diagnosed with FL and from those, 35 underwent a histological transformation to DLBCL (tFL). The remaining seven cases were diagnosed as *de novo* DLBCL. tFL cases (31 out of 35) were selected according to the availability of the pre-transformed (antecent FL) along with tFL paired material. In addition, 139 FL cases were selected for this study based on the absence of reported transformation in the clinical databases of the hospital of origin. For 57 cases, the tumor sample included in the study was the biopsy at relapse. This material was obtained from Barts Cancer Institute Biobank (143 patients), Cruces University Hospital (5 patients) and Donostia University Hospital (33 patients). Detailed information of each sample is available in Table 1-Annex.

For the inclusion of samples in this study from local Hospitals written consent was obtained from patients and approved by local ethics committees. These samples were collected from the Basque Tissue Biobank. For the samples from the UK, written consent was obtained for collection and use of specimens for research purposes with ethical approval obtained from the Institutional Review Board.

4.2.2 Whole Genome Sequencing (WGS)

WGS was carried out at Barts as described in this publication [82]

4.2.2.1 Sample characteristics

WGS data obtained from six patients with FL who underwent histological transformation to DLBCL (tFL) was used for this study (Figure 4.1). These six patients were selected based on the availability of sequential tumor lymph node biopsies (a sample with FL and a subsequent sample with histological transformation (tFL)), and matched germline material consisting of remission bone marrows or peripheral blood specimens. The clonality between the samples from the same patient was confirmed analysing the BCL2-IGH breakpoint. All biopsies were previously histologically reviewed to confirm diagnosis and tumor content before DNA extraction from fresh frozen tumor

samples [82]. In total, 14 samples from these six patients were included in this study as indicated in Figure 4.1. This group represents the discovery cohort.
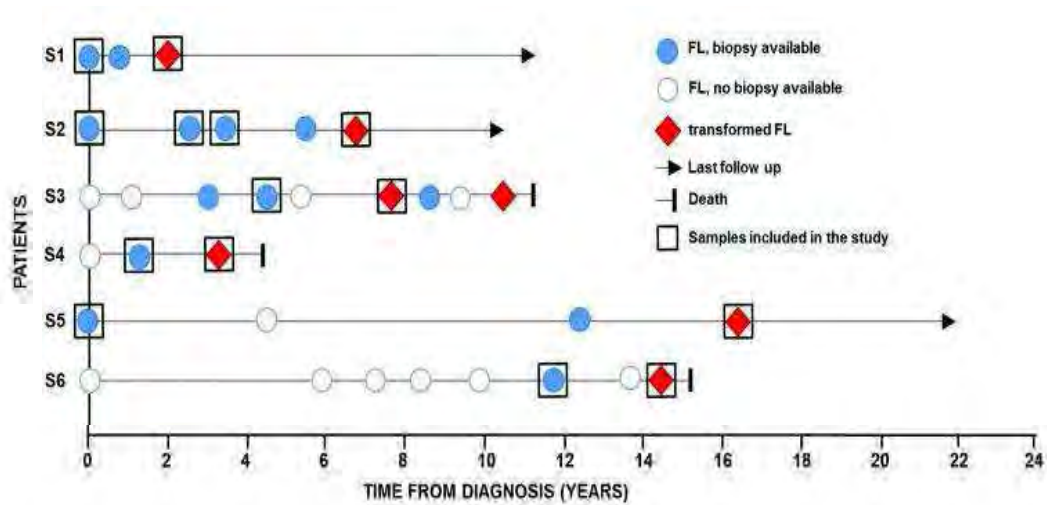
Figure 4.1: Biopsy information with disease event timelines for the 6 WGS cases (discovery cohort). *Modified from Okosun, et al. 2014 [82]*

4.2.2.2 Whole genome sequencing performance

5µg of genomic DNA from each sample was fragmented to average 300 bp insert-size libraries and sequenced using the HiSeq 2000 platform with paired-end reads of 100 bp, according to manufacturer's protocols (Illumina, San Diego, CA). Sequencing was performed with a mean coverage of 37X with 96% of the targeted bases covered at >10-fold coverage [82].

4.2.2.3 WGS pre-processing and variant calling

Raw sequencing data were aligned to the reference human genome (hg19) using the Illumina ELANDv2 aligner to generate BAM files. PCR duplicates were removed and poorly aligned and anomalous reads were re-aligned using CASAVA (v1.8). Somatic single nucleotide variants (SNVs) and indels were identified using Strelka algorithm due to its sensitivity of mutation detection with varying tumor purities. Those variant candidates were subsequently filtered: only the somatic calls originating from homozygous reference alleles in the normal samples were selected and those SNVs and indels with somatic quality score Q < 15 and < 30, respectively, were discarded [82].

4.2.3 Computational prediction of mutations present in miRNA binding sites using a bespoke script based on the TargetScan and miRanda miRNA target prediction algorithms.

For the prediction of mutations present in bioinformatically predicted miRNA binding sites, only the identified somatic mutations present in genes that encode for proteins were considered (n= 2056). The computational script was developed and run by Dr. Jose Afonso Guerra-Assunção of the Sanger Centre in Cambridge. In brief, the mature sequences of all human microRNAs were retrieved from miRBase [344]. The sequences were then clustered and collapsed into discreet families according to if they shared the same seed region or not. Genomic coordinates for human 3'UTR's were obtained from Ensembl bioMart [345] and a matching repeat masked whole genome FASTA file was retrieved from the Ensembl FTP servers. A custom set of scripts was then employed to produce FASTA files containing UTRs without mutations, along with the mutations identified by WGS in the patients.

The TargetScan algorithm was employed to predict target sites in the generated sequence files [262], and the results depending on the mutations being present or absent were compared. Only mutations located among the 75% most confident target sites for each miRNA were considered, excluding the 25% of with the lowest scores. The miRanda target prediction algorithm was then employed on these targets to try to assess the thermodynamic stability and the effects of the mutations on potential microRNA binding to the target [266]. This analysis was restricted to targets previously identified by TargetScan.

4.2.4 Pathways enrichment analysis: Ingenuity

Ontogeny analysis of mutations identified in predicted miRNA-binding sites was carried out using Ingenuity Pathway Analysis tool [346].

4.2.5 Validation of the candidate mutations with Sanger Sequencing and analysis of the validated mutations in a preliminary extended cohort

Candidate mutations were analysed in the original discovery cohort by Sanger Sequencing (see protocol in section 3.4) using specific PCR primers listed in Table 4.1. Validated mutations (14/43) were then measured in an extended cohort of patients named as preliminary extended cohort. This cohort consisted of 49 cases including the 6 cases from the discovery cohort: 27 samples from 14 FL patients who underwent transformation, 28 non-transformed FL (ntFL) cases and 7 de novo DLBCL cases. See Table 1-Annex for details of the patients in each cohort.

Gene	Amplicon Genomic Coordinates	Fwd primer (5'-->3')	Rev primer (5'--> 3')
IFT81	no designable primers	-	-
ATP10A	chr15:25939883-25940117	CCCATCTCGTGGTGTAGCAT	GGCCTACGCTCTCGAGAAA
ATP10A	chr15:25958683-25959172	CATGCAATTCTCCAATGGGCAT	ATCGCCAGCAACGGCTA
BCL2	chr18:60793672-60794161	ACTGAATGAATCTCATGGGTTTAACCAA	TGGAATAACTCTGTGGCATTATTGCAT
EZH2	chr7:148508535-148508948	GCCCAGCACAATCCAGTTACTA	TGCCCAATTACTGCCTTAGAACA
GDI2	chr10:5807406-5807888	GCTTTCTCCAAATGACTCCCTATGT	CCTGCTTTTGTAATGAAAATGGAGAGA
MEF2B	chr19:19257430-19257884	GGTTCTGCAGGCCACTGTAG	TCTCCCTTCCGACCAGCA
MEF2B	chr19:19259889-19260386	GCCAGGTTGTGACATGAGAA	TCCCTTCTAGGGACAGCAGA
PCSK5	chr9:78975019-78975537	GCAGGGTCTTACAGATTTCCTTGT	GCACAAGCCTTTAGTAAACTGAAGA
SLC29A2	chr11:66130554-66131039	CCACTCTGAACCCTCTGGTCA	TCCTGTTCCCGGGTGTTCTA
RETSAT	chr2:85569518-85570203	CAGCAAAGTATTCATCTTACCAGTCCA	GGGCAATGGAATCACTGCT
ITGB1BP1	chr2:9546041-9546910	CAGCATGAATCCTGGTCTCCA	ACTGCAATGCTGAAATGTTATGCAAATA
MYO5b	chr18:47349334-47350033	CCTTTCCATGAAGCATGTCA	TGGATTCACCCCTGAATATGA
BCL2	chr18:60793295-60793785	GACTCCATTAAAATGATTTTGGCAGGAT	TGTCCCTTTGACCTTGTTTCTTGAA
PCSK5	chr9:78808241-78808644	CTGCAAAACATGTACATTTCAAGGCT	AAGCAAGTCAATGCAAAAGGGAAAA
BMP7	chr20:55745350-55745743	GCTCTTTCCCCTCCCACTAGA	GGTCATTGGCTGGGAAGTCTC
MYO5B	chr18:47350150-47350594	CTTCAGCAACAACTCATGTGTACATT	ACTTGGAGGCTAGAAGAAGCATCT
MYO5B	chr18:47352489-47353001	GGACATTCTCTAGTTCTATGCAAAGATC	CCCAATGTGAGCAAGAAGGAAGTATAT
PEX19	chr1:160247446-160247683	GCTAGGCTAGCAGAAGTGAGAATC	GGAAGAGCCCAGATCTGAATGTAA

Table 4.1: Primers for PCR amplification and Sanger Sequencing for validation of mutations identified in the discovery cohort.

4.2.6 Ampliseq (Ion Torrent) NGS panel design and performance

To explore the frequency of all the identified miRNA-binding site mutations (i.e n=544), we designed a custom NGS panel using Ion AmpliSeq targeted sequencing technology. The Ion Ampliseq Designer v3.4 tool was used to design the panel using coordinates obtained from the original WGS data. The panel consisted of a single pool of 482 separate amplicons (125-175bp in length), covering a total of 57kb, and included 503 of the 544 (92%) of the originally identified mutations.

A total of 98 samples from 55 patients were analysed using this sequencing panel. This group represented the validation cohort and consisted of 64 samples from 34 FL patients who underwent transformation (including the 6 cases from the discovery cohort), and 21 samples of non-transformed FL patients. In addition, five pools from 34 healthy controls together with 8 lymphoma cell lines (DLBCL and FL) were sequenced. DNA from healthy controls was sequenced in order to identify sequencing artefacts and genetic polymorphisms not associated with the disease. The reason of using pools was to maximize the detection of SNPs in the normal population keeping a detectable variant allele frequency (VAF) number. For more detailed information see Table 1-Annex. In total, seventy-five samples were sequenced on an Ion Torrent Personal Genome Machine (PGM) in the genomics facility of Biodonostia. This data was generated during eleven different sequencing runs using 500 flows with paired-end reads of 150 bp. Fifty-eight samples were run on 316v2 chips (4-8 samples per chip) and eighteen samples on two 318v2 chips (9 samples per chip). The remaining twenty-three samples were sequenced externally by Genetracer Biotech (Santander, Spain) on an Ion Proton machine.
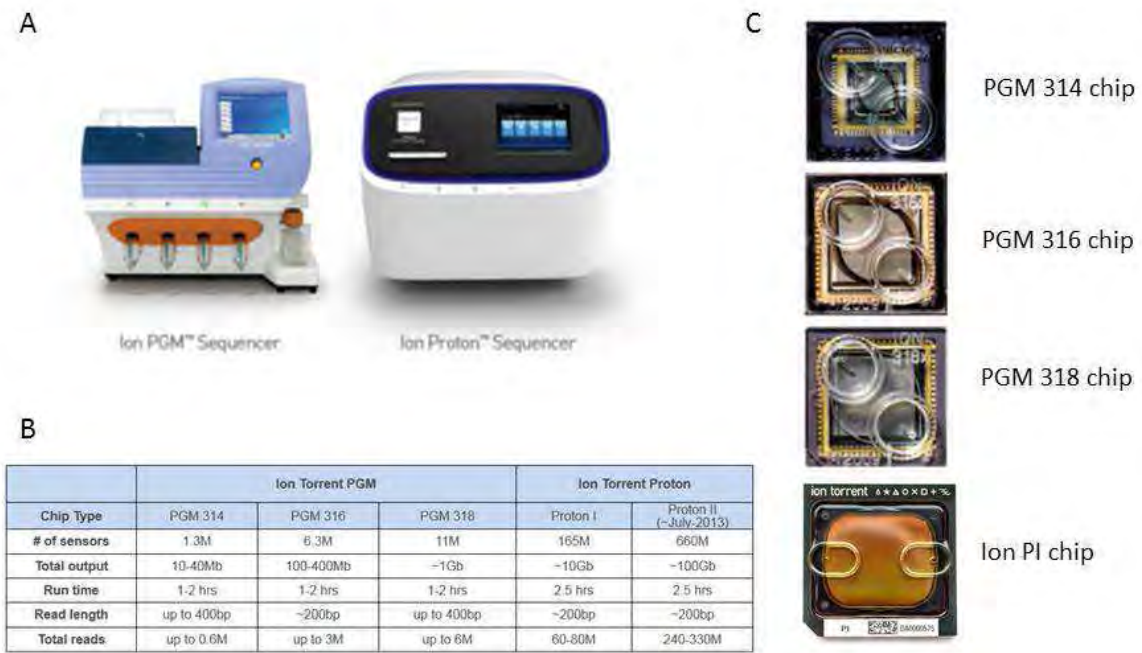
Figure 4.2: *Ion Torrent platforms (A), specifications of Ion Torrent different platforms and chips (B), illustration of Ion available sequencing chips (C).*

Signal processing and base calling were performed within the Torrent Server using default parameters. After de-multiplexing, reads were aligned against the hg19 version of the human genome, and variants were called using the Ion Reporter v5.2 online tool using somatic low stringency parameters from the Torrent Variant Caller 4.4 plugin.

4.2.7 Functional *in-vitro* studies - Luciferase Assays

Luciferase assays were performed on seven candidate genes: *BCL2, EZH2, MEF2B, SPOCK3, ARMC10, METTL15* and *SENP1*. We analysed all the variants that we identified in these seven genes. For *BCL2* gene, three different loci where analysed. For EZH2, an additional mutation (*hg19 chr7:g.[148508727T>G]*) identified in SU-DHL-4 cell line and commonly identified in FL patients [80] was included. So, in total, nine miRNA-binding sites were analysed (see table 4.2). The cDNA sequences were cloned into psiCHECK-2 luciferase vector to carry out the assays. Firstly, wild type sequences were analysed. In those validated as *bona fide* target sites the assay was performed together with all the mutated sequences for each locus.

Gene	(*) Strand / Location	miRNA	miRNA-binding site	Variants/ mutations (**)
ARMC10	+ / 3'UTR	miR-222-5p	chr7:102739177-102739198	*hg19 chr7:g.[102739179A>G]* (1) **Mut** *hg19 chr7:g.[102739188C>T]* (2) **Var**
METTL15	+ / 3'UTR	miR-4313	chr11:28353429-28353448	*hg19 chr11:g.[28353434G>A]* **Mut**
EZH2	- / Exon	miR-144-3p	chr7:148508722-148508742	*hg19 chr7:g.[148508727T>A]* (1) **Mut** *hg19 chr7:g.[148508727T>G]* (2) **Mut** *hg19 chr7:g.[148508728A>T]* (3) **Mut**
MEF2B	- / Exon	miR-1265	chr19:19260038-19260055	*hg19 chr19:g.[19260045T >A]* (1) **Mut** *hg19 chr19:g.[19260055T>C]* (2) **Var**
BCL2(1)	- / 3'UTR	miR-27a-5p	chr18: 60793903-60793925	*hg19 chr18:g.[60793906G>A]* (1) **Mut** *hg19 chr18:g.[60793921T>C]* (2) **Var**
BCL2(2)	- / 3'UTR	miR-138-5p	chr18:60793644-60793665	*hg19 chr18:g.[60793653T>C]* (1) **Mut** *hg19 chr18:g.[60793655T>C]* (2) **Var**
BCL2(3)	- / 3'UTR	miR-5008-5p	chr18:60793436-60793458	*hg19 chr18:g.[60793447G>A]* **Var**
SENP1	- / 3'UTR	miR-196a-3p	chr12:48438308-48438329	*hg19 chr12:g.[48438315A>G]* **Var**
SPOCK3	- / 3'UTR	miR-192-5p	chr4:167655214-167655235	*hg19 chr4:g.[167655217A>T]* (1) **Mut** *hg19 chr4:g.[167655218C>T]* (2) **Var** *hg19 chr4:g.[167655235A>G]* (3) **Var**

Table 4.2: Candidate mutations/variants for luciferase assays performance. *All the variants identified with targeted re-sequencing located in the seven candidate genes were analysed using luciferase assays. These genes are: BCL2, EZH2, MEF2B, SPOCK3, ARMC10, METTL15 and SENP1. (*) Strand is referred to the strand where the gene is located. Locations means in what genomic location the variant is present in the gene. (**) "**Mut**" refers the confirmed somatic mutations identified in FL tumours and not in the germ line material in the discovery cohort. "**Var**" refers to variants detected in the validation cohort that cannot be identified with certainty as somatic mutations as germ line material from these patients was lacking.*

4.2.7.1 Cloning

Primers were designed to clone the miRNA-binding sites along with a flanking region of approximately 300bp (Table 4.3). The primers also contained the PmeI and NotI binding sites for cloning into the psiCHECK-2 luciferase vector (Promega) (Figure 4.3-A). For the PCR amplification of these regions, DNA from lymphoma cell lines or patients was used as a template (see detailed protocol of PCR amplification in section 3.2). When the template DNA containing the mutations/variants was limited or difficult to amplify, we used synthetic produced DNA for cloning instead. This synthetic DNA contained the

restriction sites for PmeI in 5' and NotI in 3' (GeneArt, ThermoFisher) (Figure 4.3-B). Lyophilized synthetic DNA fragments were reconstituted in RNAse/DNAse free water to get a final concentration of 10ng/µL and half of the volume was used for the cloning (see section 3.10 for the detailed cloning protocol). For more detailed information about the luciferase constructs see Table 4.3.
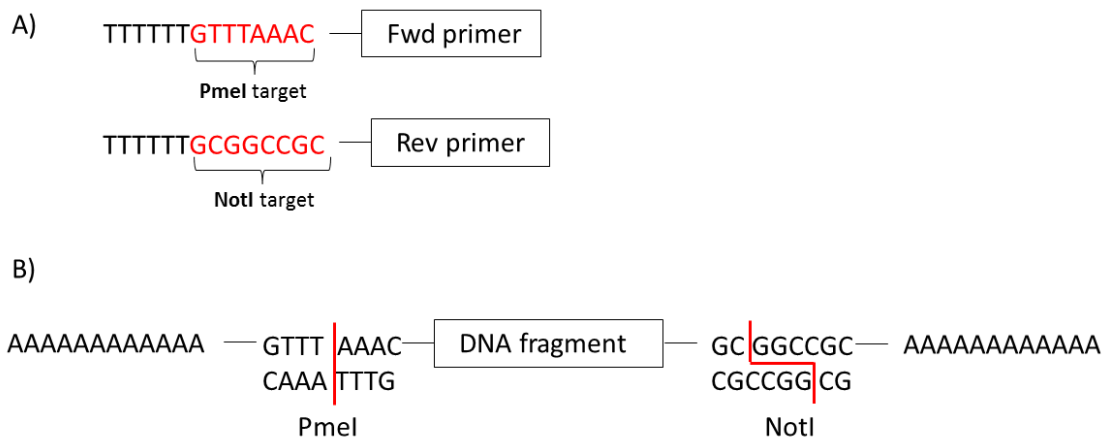
Figure 4.3: *Schematic example of the designed primers (A) and of the designed synthetic DNA (B) both containing restriction sites for PmeI and NotI.*

Construct	Mutation /variant position	Strand	Template	Fwd primer (5'--> 3')	Rev primer (5'--> 3')	Location
ARMC10 wt	-	+	RIVA cell line	TTTTTT**GTTTAAAC**GGGATTCTCCCAGCTGCTAA	TTTTTT**GCGGCCGC**TGCCACTTCATGTAGGGAAG	3'UTR
ARMC10 mut1	chr7:102739188	+	RIVA cell line	TTTTTT**GTTTAAAC**GGGATTCTCCCAGCTGCTAA	TTTTTT**GCGGCCGC**TGCCACTTCATGTAGGGAAG	3'UTR
ARMC10 mut2	chr7:102739179	+	RIVA cell line	TTTTTT**GTTTAAAC**GGGATTCTCCCAGCTGCTAA	TTTTTT**GCGGCCGC**TGCCACTTCATGTAGGGAAG	3'UTR
SPOCK3 wt	-	-	FL18 cell line	TTTTTT**GTTTAAAC**GACCAGCAACTTAGGGAGCA	TTTTTT**GCGGCCGC**TTTTCTTCATTCCAGCAGCA	3'UTR
SPOCK3 mut2	chr4:167655218	-	FL18 cell line	TTTTTT**GTTTAAAC**GACCAGCAACTTAGGGAGCA	TTTTTT**GCGGCCGC**TTTTCTTCATTCCAGCAGCA	3'UTR
SPOCK3 mut3	chr4:167655235	-	Su-DHL-4 cell line	TTTTTT**GTTTAAAC**GACCAGCAACTTAGGGAGCA	TTTTTT**GCGGCCGC**TTTTCTTCATTCCAGCAGCA	3'UTR
SPOCK3 mut1	chr4:167655217	-	Synthetic DNA	-	-	3'UTR
BCL2(1) wt	-	-	RIVA cell line	TTTTTT**GTTTAAAC**TTTGTGAGCAAAGGTGATCG	TTTTTT**GCGGCCGC**TTCTGTTGACGTGGAAATGC	3'UTR
BCL2(1) mut1	chr18:60793906	-	Synthetic DNA	-	-	3'UTR
BCL2(1) mut2	chr18:60793921	-	Synthetic DNA	-	-	3'UTR
BCL2(2) wt	-	-	RIVA cell line	TTTTTT**GTTTAAAC**TTCCGCATTTAATTCATGGT	TTTTTT**GCGGCCGC**AAGGACAGCCATGAGAAAGC	3'UTR
BCL2(3) wt	-	-	RIVA cell line	TTTTTT**GTTTAAAC**TCAGGGTCTTCCTGAAATGC	TTTTTT**GCGGCCGC**GCAGCACAGGATTGGATATT	3'UTR
BCL2(3) mut	chr18: 60793447	-	Synthetic DNA	-	-	3'UTR
MEF2B wt	-	-	Su-DHL-4 (cDNA)	TTTTTT**GTTTAAAC**CAACCGCCTCTTCCAGTATG	TTTTTT**GCGGCCGC**AGCCTCCGAAACTTCTCTCC	Exon
MEF2B mut1	chr19:19260045	-	Su-DHL-4 (cDNA)	TTTTTT**GTTTAAAC**CAACCGCCTCTTCCAGTATG	TTTTTT**GCGGCCGC**AGCCTCCGAAACTTCTCTCC	Exon
MEF2B mut2	chr19:19260055	-	Synthetic DNA	-	-	Exon
EZH2 wt	-	-	SU-DHL-10 (cDNA)	TTTTTT**GTTTAAAC**GCTGACCATTGGGACAGTAAA	TTTTTT**GCGGCCGC**TGGATTTACCGAATGATTTGC	Exon
EZH2 mut1	chr7:148508727	-	SU-DHL-10 (cDNA)	TTTTTT**GTTTAAAC**GCTGACCATTGGGACAGTAAA	TTTTTT**GCGGCCGC**TGGATTTACCGAATGATTTGC	Exon
EZH2 mut2	chr7:148508727	-	Su-DHL-4 (cDNA)	TTTTTT**GTTTAAAC**GCTGACCATTGGGACAGTAAA	TTTTTT**GCGGCCGC**TGGATTTACCGAATGATTTGC	Exon
EZH2 mut3	chr7:148508728	-	FL18 (cDNA)	TTTTTT**GTTTAAAC**GCTGACCATTGGGACAGTAAA	TTTTTT**GCGGCCGC**TGGATTTACCGAATGATTTGC	Exon
METTL15 wt	-	+	Su-DHL-1 cell line	TTTTTT**GTTTAAAC**AACAAATTGAGGAGATTGATTCAT	TTTTTT**GCGGCCGC**TTCACTTGGATTCACCAACC	3'UTR
METTL15 mut	chr11:28353434	+	Su-DHL-1 cell line	TTTTTT**GTTTAAAC**AACAAATTGAGGAGATTGATTCAT	TTTTTT**GCGGCCGC**TTCACTTGGATTCACCAACC	3'UTR
SENP1 wt	-	-	THD13T0034	TTTTTT**GTTTAAAC**GACAGGTCCCAAAGTTCCAA	TTTTTTG**CGGCCGC**CCAAACTAGATCTGCATTCATCC	3'UTR
SENP1 mut	chr12:48438315	-	THD13T0034	TTTTTT**GTTTAAAC**GACAGGTCCCAAAGTTCCAA	TTTTTTG**CGGCCGC**CCAAACTAGATCTGCATTCATCC	3'UTR

Table 4.3 Genomic location of the mutations/variants in the putative miRNA-binding sites: *the tables also shows the cases where synthetic DNA was used for the cloning. For the rest of the cases, the sequences of the primers used for the cloning into psiCHECK-2 luciferase vector are shown together with the template used. For the cloning of regions located in exons, cDNA was used as template to mimic as much as possible the original target sequence (mRNA). Highlighted in bold letters are PmeI restriction site (Fwd primer) and NotI restriction site (Rev primer).*

After cloning and PCR screening, 6-10 colonies with the correct insert size were sequenced using Sanger Sequencing (see section 3.4). Once the correct clones were confirmed (wild type and/or mutated), bacteria from the original colonies were grown in 5mL of LB broth with 100µg/mL Ampicillin at 37ºC 180 rpm o/n and plasmids were extracted (QIAGEN Miniprep Kit). After plasmid extractions, DNA concentration and quality were measured using nanodrop NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific, DE, USA) and the insert of these plasmids were re-analysed using bi-directional Sanger Sequencing. The resulting plasmid DNA was stored at -20ºC until further use.

4.2.7.2 Transfection and luciferase activity measurement

For luciferase assays HEK293 cells were seeded in 12-well plates, 24 hours before transfection. After this time cells were transfected with 200ng/well of psiCHECK-2 vector along with 50nM of the miRNA mimic or a scramble miRNA sequence used as a negative control (miRIDIAN miRNA mimics, Dharmacon). DharmaFECT Duo (Dharmacon) was used as the transfection reagent (1,2µL/well), and transfections were performed in biological triplicate. 48 hours after transfection, the cells were washed with PBS and lysed using lysis buffer from the Dual-Luciferase Reporter Assay (Promega) kit. Cell lysates were collected and stored at -80ºC (Figure 4.4).

Firstly, 20uL of each lysate were added to 100uL of Luciferase Assay Reagent II (LAR II) to measure the luciferase activity of Firefly luciferase. After quantifying the luminescence, 100uL of Stop&Glo reagent are added. Luminescence was measured using a PHERAstar microplate reader (BMG LABTECH). The detected luminiscence from Renilla luciferase was normalized using the luminiscence from Firefly luciferase for each well. The normalized luciferase signal of each condition was taken and the luciferase activity was calculated taking as reference the normalized luciferase signal of the scramble condition for each different vector type. For the stadistical analysis, the Mann–Whitney U test was applied.
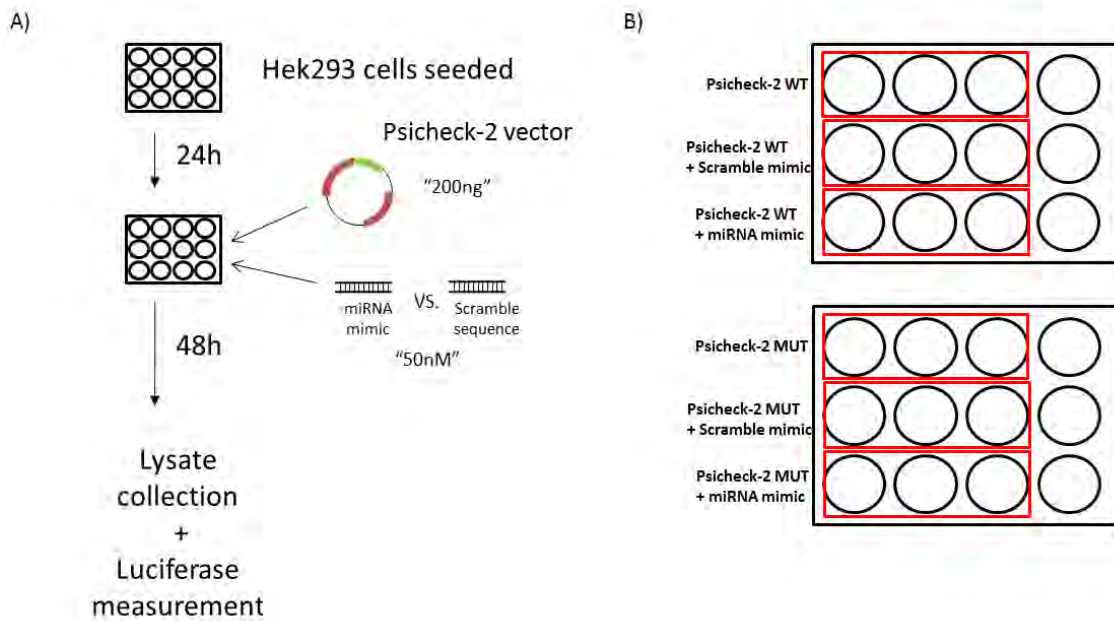
Figure 4.4: Luciferase assay protocol. A) Schematic representation of the transfection and luciferase measurement protocol for luciferase assays. B) Representation of an assay performed in biological triplicate in 12-well plates for both, wild type and mutated conditions.

4.2.8 Analysis of the EZH2 and BCL2 mutation prevalence in an extended cohort using Custom TaqMan SNP Genotyping Assays:

For analysing the prevalence of these specific variants/mutations in *BCL2* and *EZH2* genes, Custom TaqMan® SNP Genotyping Assays (ThermoFisher) were performed in a cohort of 117 FL patients. For EZH2, an additional mutation (*hg19 chr7:g.[148508728A>G)]* usually found in FL patients and affecting the same locus was included. To analyse that, 5 Taqman SNP genotyping assays were performed with probes designed in order to detect the mutations present in *BCL2* and *EZH2* genes using quantitative real time PCR:

4. Investigating mutations in the miRNA binding sites of follicular lymphoma

Assay	Gene	Genomic variant
1	BCL2	hg19 chr18:g.[60793447G>A]
2	EZH2	hg19 chr7:g.[148508727T>A]
3	EZH2	hg19 chr7:g.[148508727T>G]
4	EZH2	hg19 chr7:g.[148508728A>T]
5	EZH2	hg19 chr7:g.[148508728A>G]

Table 4.4: Taqman assays used for sample genotyping. Each assay contains probes for the specific detection of the wild type and the mutated sequence. Both alleles (wild type and mutated) can be differentiated within the same amplification reaction as they carry different fluorophores (VIC and FAM respectively).

The tumoral DNA from these patients was obtained from our collaborators in Barts Cancer Institute. These samples consist of non-paired samples of DNA from FL patients at diagnosis and also DNA from FL patients at relapse (Table 1-Annex.). For the qRT-PCR reactions, Taqman Genotyping Master Mix (Applied Biosystems) was used together with the genotyping primers with the probes. 10ng of template DNA was used for each reaction in a final volume of 10μL in a 96 well plate.

Genotyping Master Mix 2X	5μL
Taqman genotyping primers 40X	0,25μL
DNA Template (5ng/μL)	2μL
H$_2$O	2,75μL
	10μL total volume

The PCR programme used had a pre-incubation step of 10 minutes at 95ºC followed by 45 cycles of 10 seconds at 95ºC and 1 minute at 60ºC.

To estimate the limit of detection of the technique the Taqman assays number 1, 2 and 4 listed in table 4.4 were used. To do that, DNA from mutated samples with a known VAF percentage was used. This DNA was sequentially diluted to simulate mutation at different and decreasing VAF percentages. A non-mutated DNA was used as a negative control. As template material for the qRT-PCR reaction 5ng versus 10ng of DNA were used in parallel. Similarly, to estimate the VAFs of the identified mutations standard curves were used. Mutated DNA with known VAF percentages were run together with

sequential dilutions of these DNAs and a non-mutated DNA was used as a negative control. Resulting decreasing curves were plotted and used as reference.

The platform used for estimating the limit of detection of the technique was a LightCycler® Real-Time PCR System (Roche Life Science) whereas for the rest of the assays a StepOnePlus™ Real-Time PCR System (Applied Biosystems, California, USA) was used. In the last platform the number of PCR cycles was increased to 50 in order to better observe the amplification curves.

4.2.9. Functional validation of this mechanism in lymphoid cell models

WSU-DLCL2 and HT DLBCL cell lines were transfected with 1µg of a *miR-144* mimic in parallel with a scramble miRNA sequence used as a negative control (miRIDIAN miRNA mimics, Dharmacon). WSU-DLCL2 cell line carries the mutation *hg19 chr7:g.[148508727T>A]* in *EZH2* whereas HT cell line has a wild type *EZH2*. For the transfection, a Nucleofector™ 2b Device was used (program G-016) with the Cell Line Nucleofector® Kit V (Lonza, Visp, Switzerland). Half of the cells were recovered after 48 hours since the transfection and total RNA was extracted as indicated in section 3.1.3. *MiR-144* and *EZH2* expression levels were measured using qRT-PCR (see sections 3.5 and 3.6). 72 hours post-infection the rest of the transfected cells were harvested, washed with PBS and lysed. Protein lysates (30µg) were run on a polyacrylamide gel and transferred to a nitrocellulose membrane for Western-blot analysis (see section 3.13). Membranes were blocked with 5% Skim Milk diluted in TBS-Tween 0.1%. A monoclonal antibody for the identification of EZH2 protein was used (*Ezh2* (D2C9) XP ® Rabbit mAb Antibody, Cell Signaling Technology, Leiden, Netherlands). This antibody was incubated in 5% BSA diluted in TBS-Tween 0.1% according to the specifications of the manufacturer. Tubulin staining (Purified anti-Tubulin β 3 (TUBB3) Antibody, Clone TUJ1, BioLegend, San Diego, CA) was used as a loading control (dil 1:1000 in 5% Skim Milk diluted in TBS-Tween 0.1%). goat anti-rabbit IgG-HRP and goat anti-mouse IgG-HRP (Santa Cruz Biotechnology, Texas, USA) diluted 1:2000 in in 5% skimmed milk TBS-Tween 0.1% were used respectively as secondary antibodies.

4.2.10 *In-silico* analysis of *miR-144* and *EZH2* expression levels in FL patients

Public microarray data accessible via the Gene Expression Omnibus (GEO) database repository was used in order to perform *in-silico* expression analysis. For the analysis of *miR-144* expression levels, data from the study entitled "microRNA signature in b-cell lymphomas and control human B cells from fresh tonsils" was used (GEO accession number: GSE29493) which contains 147 B-cell lymphoma samples (29 DLBCL samples and 23 FL samples) together with purified human B cell population controls (n=7). The platform used in this dataset was the Agilent-016436 Human miRNA Microarray 1.0.

For the analysis of *EZH2* expression, data from a study of gene expression profiling in Follicular Lymphomas was used (GEO accession number: GSE55267). The aim of this study was to determine differential gene-expression between the *EZH2* mutated vs unmutated subgroups in FL. For that purpose, they used Affymetrix Human Genome U133 Plus 2.0 Arrays for analysing the gene expression profile in a total of 63 FL samples.

4.3 Results

4.3.1 WGS data analysis showed that mutations in putative miRNA binding sites are frequent and occur preferentially in genes associated with haematological malignancies.

WGS data was generated by the group of Jude Fitzgibbon from the Centre for Haemato-Oncology in Barts Cancer Institute, and further details of the sequencing data can be found in the following publication [82]. In brief, approximately 10,000 somatic variants were identified per tumor, with 150-300 located in protein coding genes. In all of the six cases a total of 2056 different somatic variants were identified in 1605 genes. Nearly half of these mutations were in coding regions (47%), with nearly the same number of mutations in the 3'UTRs (38%), and less so in the 5'UTRs (15%).
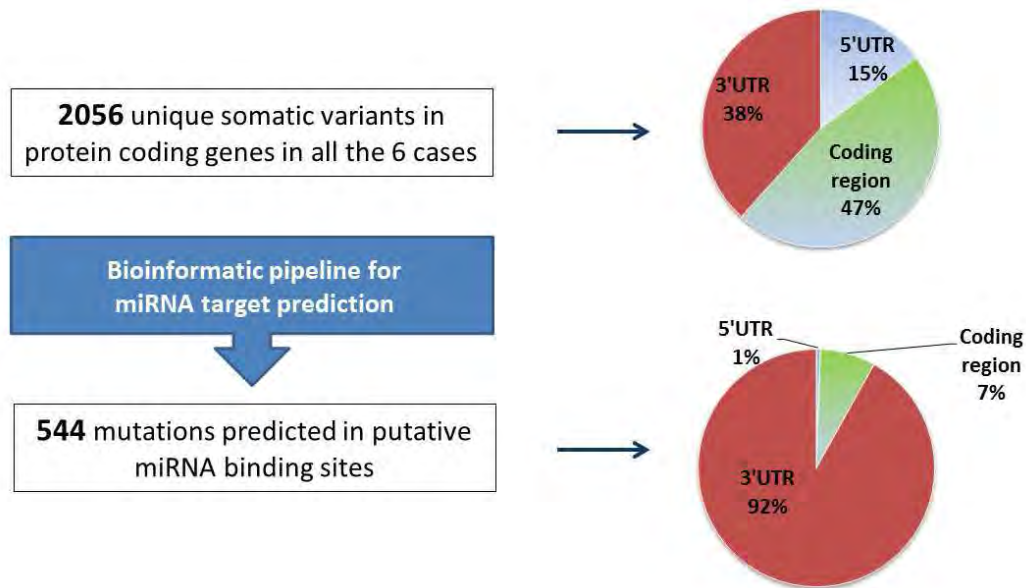
Figure 4.5: Comparative between the distribution in genomic location of the total identified mutations in protein coding genes and the distribution of mutations only in miRNA-binding sites.

These data were interrogated using a bespoke bioinformatic pipeline for miRNA target prediction. From the initial 2056 somatic variants identified, we determined that a total of 544 mutations (located in 490 genes) were present in predicted miRNA binding sites (Table 2-Annex). The vast majority of these (92%) were located in 3'UTRs (Figure 4.5). From all the identified somatic mutations in the 3'UTRs, 68% arose in predicted miRNA-binding sites (Figure 4.6). Only 1.6% of genes had mutations both in coding regions and 3'UTRs.
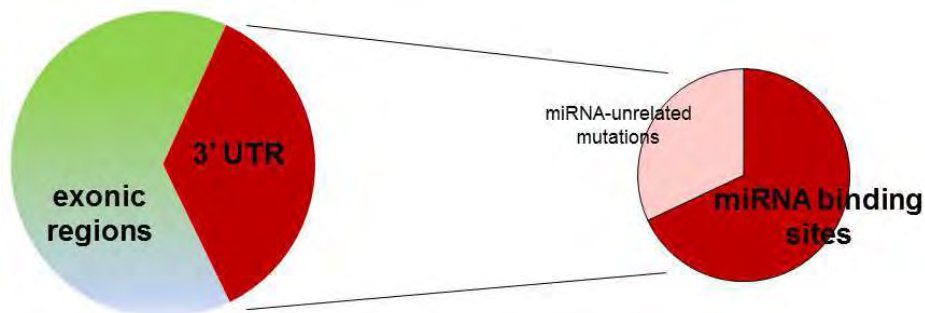
Figure 4.6: Distribution of frequency of mutations in protein coding genes in 6 WGS cases (n = 2056 mutations).

The number of mutations identified in each sample is represented in Figure 4.7. As can be seen 5 out of 6 patients had a higher number of mutations in the transformed state.

4. Investigating mutations in the miRNA binding sites of follicular lymphoma

Only 3 mutations were present in more than one patient, those in *EZH2, MEF2B and POM121* genes.
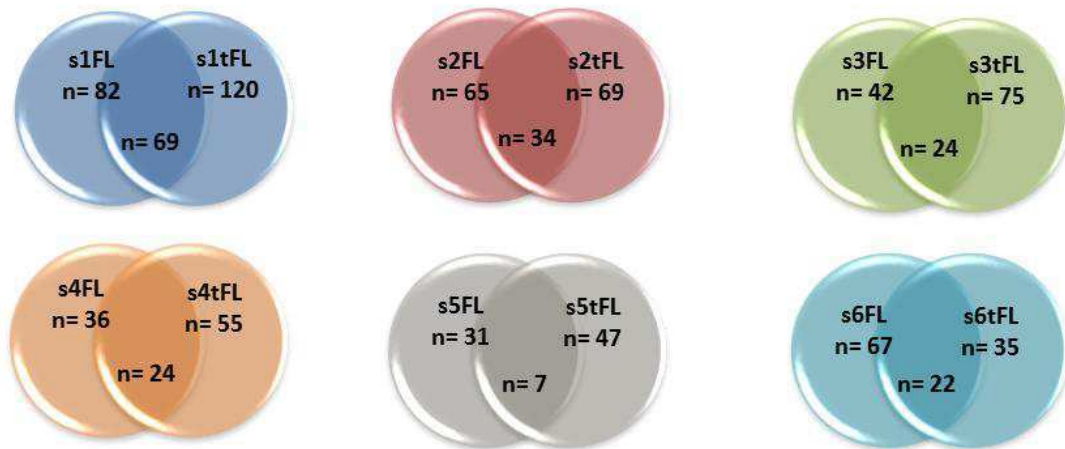
Figure 4.7: Number of mutations in predicted microRNA binding sites identified in each of the 6 patients using WGS.

We carried out ontology analysis on the 490 genes with mutations in putative miRNA binding sites. This analysis showed a significant enrichment for genes associated with haematological malignancies ($P=2.18 \times 10^{-4}$) (Figure 4.8). This data suggested that mutations in miRNA binding sites are not random and they might play a regulatory role in FL pathogenesis. Therefore, we decided to study this in more depth. To do that, we selected some candidate genes to study the prevalence of the identified mutations in a larger cohort of patients.
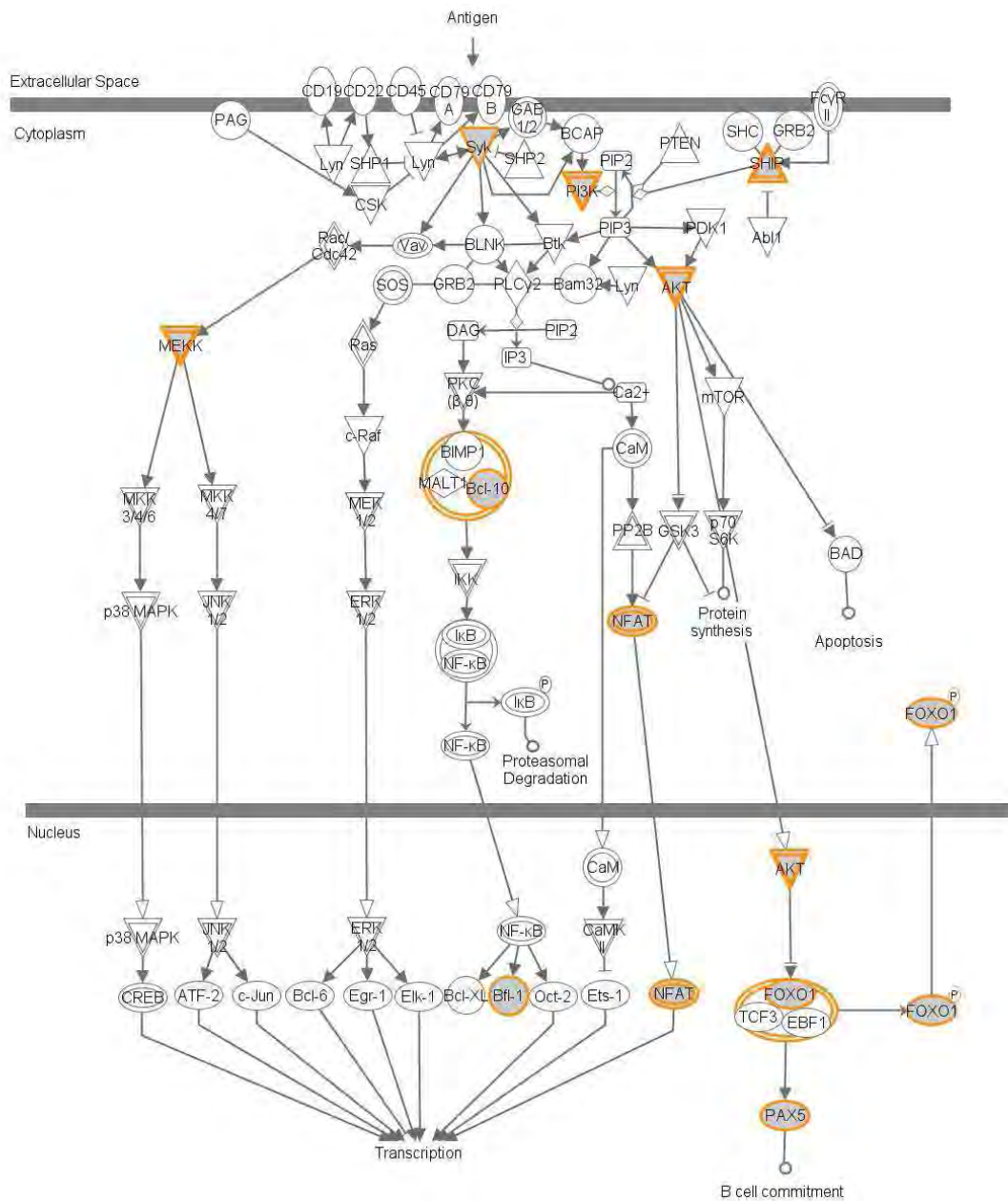
Figure 4.8: Identified genes with mutations in predicted miRNA binding sites involved in B-cell signalling. *. The figure represents the B-cell signalling pathways and the identified mutated genes are highlighted in the figure.*

4.3.2 Sanger Sequencing validation led us validate one third of the identified mutations using WGS. Searching of the validated mutations in a preliminary extended cohort showed recurrent mutations in *EZH2* and MEF2B genes
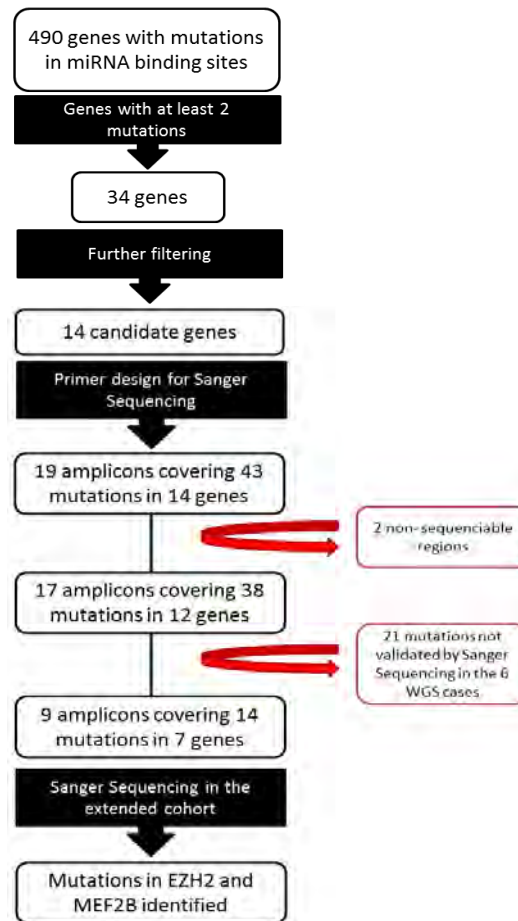
*Figure 4.9: **Workflow diagram for candidate gene selection.** Sanger Sequencing validation of candidate mutations and analysis of their prevalence in a preliminary extended cohort.*

From the 490 genes we focused on those containing more than one mutation in putative miRNA-binding sites; a total of 34 genes. We selected 14 of these genes for further analysis (43 mutations) (Table 4.5), according tothe following criteria:

1) Genes related with FL physiopathology or with other lymphoid malignancies

2) Genes with multiple mutations

3) Mutations affecting multiple patients

4) Mutations present only in the transformed FL, so they might be associated with the transformation phenomenon.

Genes	Number of mutations	Patients affected (n=6)	Pre-transformation (FL)	Post-transformation (tFL)
MYO5B	9	3	0	3
BCL2	5	1	1	1
ATP10A	3	3	2	3
IKZF3	3	1	1	0
BMP7	2	1	1	1
EZH2	2	2	2	2
MEF2B	2	2	1	2
PCSK5	2	2	0	2
IFT81	4	2	2	1
GDI2	3	1	0	1
ITGB1BP1	2	1	0	1
PEX19	2	1	1	0
RETSAT	2	2	2	0
SLC29A2	2	2	2	1

Table 4.5: List of the 14 genes with more than one mutation prioritized for validation

The frequency of the candidate mutations was analysed using Sanger Sequencing in a validation cohort of 49 patients. We were unable to design primers covering the mutations in *IFT81* gene (see loci in table 4.6), probably as these mutations were located within a large repetitive sequence region. The amplicons covering the *PCSK5* gene (see table 4.6) gave ambiguous results, again probably as a result of the mutation being close to a highly repetitive sequence region. As a consequence, these two amplicons were discarded and the mutations removed from further analysis. The remaining 37 mutations were analysed for the validation of the WGS data. Fourteen of the 37 mutations were validated in this manner (encoded within seven genes) (Table 4.6, Figure 1-Annex).

4. Investigating mutations in the miRNA binding sites of follicular lymphoma

Genes	Mutation position	Coordinates Amplicon	Amplicon lenght (bp)	Validation cohort (6 cases)
IFT81	chr12:110606884 chr12:110606885 chr12:110606888 chr12:110606889	no designable primers	-	**UNSEQUENCIABLE**
PCSK5	chr9:78808412	chr9:78808241-78808644	403	**UNSEQUENCIABLE**
BCL2	chr18:60793447 chr18:60793448 chr18:60793641 chr18:60793653	chr18:60793295-60793785	490	**NOT VALIDATED**
BMP7	chr20:55745511 chr20:55745512	chr20:55745350-55745743	393	**NOT VALIDATED**
MYO5B	chr18:47350291 chr18:47350304 chr18:47350307 chr18:47350444 chr18:47350494 chr18:47350459	chr18:47350150-47350594	444	**NOT VALIDATED**
MYO5B	chr18:47352742 chr18:47352754	chr18:47352489-47353001	323	**NOT VALIDATED**
MYO5B	chr18:47349754	chr18:47349334-47350033	700	**NOT VALIDATED**
PEX19	chr1:160247610 chr1:160247627	chr1:160247446-160247683	237	**NOT VALIDATED**
RETSAT	chr2:85569666 chr2:85570186	chr2:85569518-85570203	685	**NOT VALIDATED**
ITGB1BP1	chr2:9546301 chr2:9546324	chr2:9546041-9546910	869	**NOT VALIDATED**
ATP10A	chr15:25958929 chr15:25958933	chr15:25958683-25959172	489	**VALIDATED**
ATP10A	chr15:25940037	chr15:25939883-25940117	234	**VALIDATED**
BCL2	chr18:60793906	chr18:60793672-60794161	489	**VALIDATED**
GDI2	chr10:5807657 chr10:5807705 chr10:5807722	chr10:5807406-5807888	717	**VALIDATED**
PCSK5	chr9: 78975278	chr9:78975019-78975537	518	**VALIDATED**
SLC29A2	chr11: 66130901 chr11: 66130796	chr11:66130554-66131039	485	**VALIDATED**
EZH2	chr7: 148508727 chr7: 148508728 chr7: 148508764	chr7:148508535-148508948	413	**VALIDATED**
MEF2B	chr19: 19257600	chr19:19257430-19257884	454	**VALIDATED**
MEF2B	chr19:19260045	chr19:19259889-19260386	498	**VALIDATED**

Table 4.6: *Sanger Sequencing validation results of the mutations identified using WGS in 6 FL-tFL patients*

The frequency of these 14 mutations were then analysed in an extended validation cohort of 49 patients. Two of these seven genes were found to be recurrently mutated. These mutations were located in exonic regions of *EZH2* and *MEF2B* genes (Table 4.7).

Gene	Genomic variant	Location	FL-tFL cases (n=14, 27 samples)	ntFL (n=28)	DLBCL (n=7)	TOTAL (n=49)
ATP10A	*hg19 chr15:g.[25958933G>A]*	Exonic (p.S744S)	1	-	-	1
	hg19 chr15:g.[25958929G>A]	Exonic (p.R746H)	1	-	-	1
ATP10A	*hg19 chr15:g.[25940037G>A]*	Exonic (p.S1006N)	1	-	-	1
BCL2	*hg19 chr18:g.[60793906G>A]*	3' UTR	1	-	-	1
GDI2	*hg19 chr10:g.[5807657A>C]*	3' UTR	1	-	-	1
	hg19 chr10:g.[5807705A>G]	3' UTR	1	-	-	1
	hg19 chr10:g.[5807722A>C]	3' UTR	1	-	-	1
PCSK5	*hg19 chr9:g.[78975278A>T]*	3' UTR	1	-	-	1
SLC29A2	*hg19 chr11:g.[66130901G>A]*	3' UTR	1	-	-	1
	hg19 chr11:g.[66130796G>C]	3' UTR	1	-	-	1
EZH2	*hg19 chr7:g.[148508727T>A]*	Exonic (p.Y646F)	2	2	2	6
	hg19 chr7:g.[148508727T>C (*)]	Exonic (p.Y646C)	-	1	1	2
	hg19 chr7:g.[148508728A>G] (*)	Exonic (p.Y646H)	-	1	-	1
	hg19 chr7:g.[148508764T>C]	Exonic (p.K634E)	1	-	-	1
MEF2B	*hg19 chr19:g.[19257600C>G]*	Exonic (p.S209X)	1	-	-	1
MEF2B	*hg19 chr19:g.[19260045T>A]*	Exonic (p.D83V)	2	-	1	3

Table 4.7: *Sanger Sequencing prevalence analysis in a preliminary extended cohort (including the 6 WGS cases) of the identified and validated mutations in the discovery cohort.*
(*) Not present in WGS cases but located in the predicted binding site

4.3.3 Using a targeted re-sequencing approach we observed that the recurrently mutated genes identified are highly enriched for GC-like B-cell lymphoma genes

In order to expand the number of identified mutations we could screen simulatenously in a larger number of cases we designed a bespoke panel covering the 544 mutations for deep targeted re-sequencing. This panel was initially tested on the validation cohort (n = 55) of paired FL samples (before and after transformation) (n = 34) and non-transformed FL (ntFL) samples (n =21).

We achieved an average of 1,225,872 reads per sample and a mean depth of coverage of 2270X (Figure 4.10). Data related to the sequencing performance for each of the sequencing runs is shown in Table 4.8.
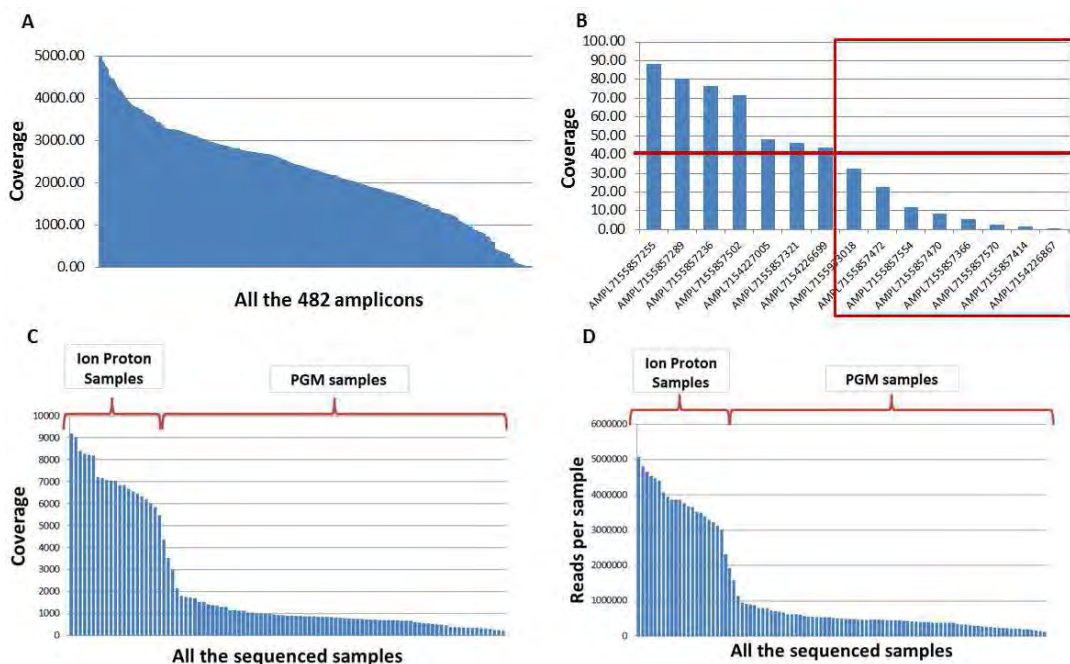
Figure 4.10: *Technical results for the sequencing of all the 98 samples using the custom design Ampliseq panel in PGM and Proton Ion Torrent platforms.*

A) Coverage of each of the 482 amplicons in all the sequencing runs.

B) Amplicons with a coverage lower than 40X and excluded for further analysis.

C) Coverage achieved in each of the sequenced samples. A mean coverage of 876X (206X-3022X) was achieved for the samples sequenced with PGM and 6874X (3536X-9198X) for the samples sequenced with Ion Proton.

D) Number of reads per sample achieved in every sequenced samples. An average of 467,901 (109,639- 1,581,329) was achieved for the samples sequenced with PGM and 3,730,470 (1,921,778-5,079,331) for the samples sequenced with Ion Proton

Run	Chip	Number of samples	Loading (%)	Policlonal reads (%)	Usable reads (%)	Total reads	Aligned bases (%)
PGM-1	316v2	4	74	24	74	3,448,493	99
PGM-2	316v2	7	79	62	33	1,655,780	98
PGM-3	316v2	8	80	46	48	2,471,031	98
PGM-4	316v2	8	80	41	55	2,764,683	98
PGM-5	316v2	7	87	46	52	2,842,486	98
PGM-6	316v2	5	69	31	66	2,890,211	98
PGM-7	316v2	6	78	19	78	3,771,469	98
PGM-8	316v2	6	78	41	58	2,843,343	98
PGM-9	316v2	6	87	45	52	2,882,236	97
PGM-10	318v2	9	94	34	61	6,460,820	99
PGM-11	318v2	9	69	38	53	4,097,967	97
Proton-1	P1v3	23	91	27	67	88,977,093	95

Table 4.8: *Table shows data related to the sequencing performance of the panel for each of the sequencing runs using PGM and Proton Ion Torrent platforms. This data includes loading efficiency, polyclonal reads, total number of reads achieved for each chip and so on.*

Using this approach, on the discovery cohort, 80% of the analysed mutations (401/503) were validated (Figure 4.11). However, the mutation in *POM121* gene, identified in more than one of the six patients from the discovery cohort (sample S3FL and S4tFL), was not validated using this approach. This suggests that this mutation was probably a sequencing atefact, as the WGS had a much low coverage (mean coverage 37X) compared to the targeted deep sequencing (mean coverage 2270X). From the validated mutations, 16 were found to be recurrently mutated in the validation cohort. In addition, 72 new variants were identified into the analysed predicted miRNA target sites (in surrounding positions to the originally identified mutation). Forty-eight of these new variants were present in multiple cases.
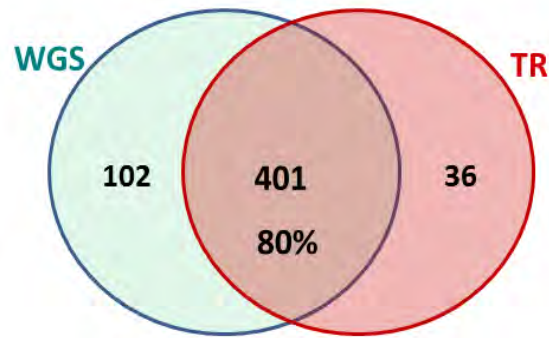
FIGURE 4.11: Comparison of variants detected by WGS and TR approaches

To this list of 88 candidate variants (16 validated and recurrent mutations together with the 72 new identified variants) we applied two filtering steps. We first removed variants also present in the healthy control samples (n=35). Secondly, variants present in homopolymer or repetitive regions were removed (n=23). This QC filtering led us to follow-up 30 variants in 28 genes (Figure 4.13, Table 4.9).

Gene enrichment analysis of the 28 remaining genes demonstrated that these genes were highly enriched for GC-like B-cell lymphoma genes (P=4.39x10^{-5}). The 30 variants were analysed to see if they were differentially present between tFL and antecedent FL paired samples, and between FL cases that transformed with FL cases that did not transform at least 5 years follow-up. Results led us to identify 4 variants only present in transformed samples and 12 variants present exclusively in patients that underwent a transformation (Figure 4.12). However, the number of patients carrying these variants is low (1-2 cases per variant) (Figure 4.12). Therefore, a statistical analysis to confirm an association with the transformation process was not able to be performed.

Gene	miRNA	miRNA-target site	Genomic variant	Cases
SYNJ2	*hsa-mir-2355*	chr6:158519568-158519589	*hg19 chr6:g[158519586_158519587del]*	17
EZH2	**hsa-mir-144**	**chr7:148508722-148508742**	**hg19 chr7:g.[148508727T>A]**	**10**
EZH2	*hsa-mir-144*	chr7:148508722-148508742	*hg19 chr7:g.[148508728A>T]*	6
METTL15	**hsa-mir-4313**	**chr11:28353429-28353448**	**hg19 chr11:g.[28353434G>A]**	**13**
MEF2B	**hsa-mir-1265**	**chr19:19260038-19260055**	**hg19 chr19:g.[19260045T >A]**	**4**
MEF2B	*hsa-mir-1265*	chr19:19260038-19260055	*hg19 chr19:g.[19260055T >C]*	1
SENP1	*hsa-mir-196a*	chr12:48438308-48438329	*hg19 chr12:g.[48438315A>G]*	1
ATP2B4	*hsa-mir-2682*	chr1:203709114-203709135	*hg19 chr1:g.[203709129A>T]*	1
HSDL2	*hsa-mir-1277*	chr9:115232898-115232919	*hg19 chr9:g.[115232916 A>G]*	2
C2orf44	*hsa-mir-302a*	chr2:24253440-24253462	*hg19 chr2:g.[24253452A>G]*	1
N6AMT1	*hsa-mir-1277*	chr21:30244896-30244918	*hg19 chr21:g.[30244899T>G]*	1
HIST1H2AC	*hsa-mir-3176*	chr6:26123667-26123685	*hg19 chr6:g.[26123679TC>AG]*	1
PRSS37	*hsa-mir-1299*	chr7:141537912-141537933	*hg19 chr7:g.[141537916G>T]*	1
SYS1-DBNDD2	*hsa-mir-1306*	chr20:44004534-44004555	*hg19 chr20:g.[44004537T>C]*	1
BCL2	*hsa-mir-5008*	chr18:60793436-60793458	*hg19 chr18:g.[60793447G>A]*	1
TUBB	**hsa-mir-1302**	**chr6:30692752-30692772**	**hg19 chr6:g.[30692754_30692755insTT]**	**9**
ARMC10	**hsa-mir-222**	**chr7:102739177-102739198**	**hg19 chr7:g.[102739179A>G]**	**21**
POU1F1	*hsa-mir-1306*	chr3:87308560-87308581	*hg19 chr3:g.[87308573del]*	4
MYO5B	*hsa-mir-216b*	chr18:47352725-47352746	*hg19 chr18:g.[47352732del]*	10
THNSL1	*hsa-mir-4653*	chr10:25314522-25314543	*hg19 chr10:g.[25314537del]*	2
EVL	*hsa-mir-2467*	chr14100604624-100604645	*hg19 chr14:g.[100604644C >T]*	3
PCDH7	**hsa-mir-329**	**chr4:30732964-30732985**	**hg19 chr4:g.[30732984_30732985del]**	4
NCMAP	*hsa-mir-4781*	chr1:24933785-24933805	*hg19 chr1:g.[24933795G>A]*	1
SERBP1	*hsa-mir-150*	chr1:67874042-67874063	*hg19 chr1:g.[67874057G>A]*	1
SH3BGRL	*hsa-mir-4662a*	chrX:80553647-80553668	*hg19 chrX:g.[80553647C>T]*	1
ZNF449	*hsa-mir-1305*	chrX:134481762-134481783	*hg19 chrX:g.[134481764T>G]*	1
RC3H1	**hsa-mir-548an**	**chr1:173901933-173901952**	**hg19 chr1:g.[173901941_173901943del]**	1
GOLGA6L2	*hsa-mir-1236*	chr15:23686235-23686255	*hg19 chr15:g.[23686253T >C]*	1
APOL2	*hsa-mir-128*	chr11:36623174-36623196	*hg19 chr11:g.[2317496C >G]*	1
RAB3D	*hsa-mir-1289*	chr19:11432999-11433021	*hg19 chr19:g.[11433019del]*	1

Table 4.9: *Recurrent variants found in miRNA-binding sites using TR after QC filtering. A total of 30 variants in 28 genes were identified. The mutations originally identified using WGS are highlighted in bold letters.*
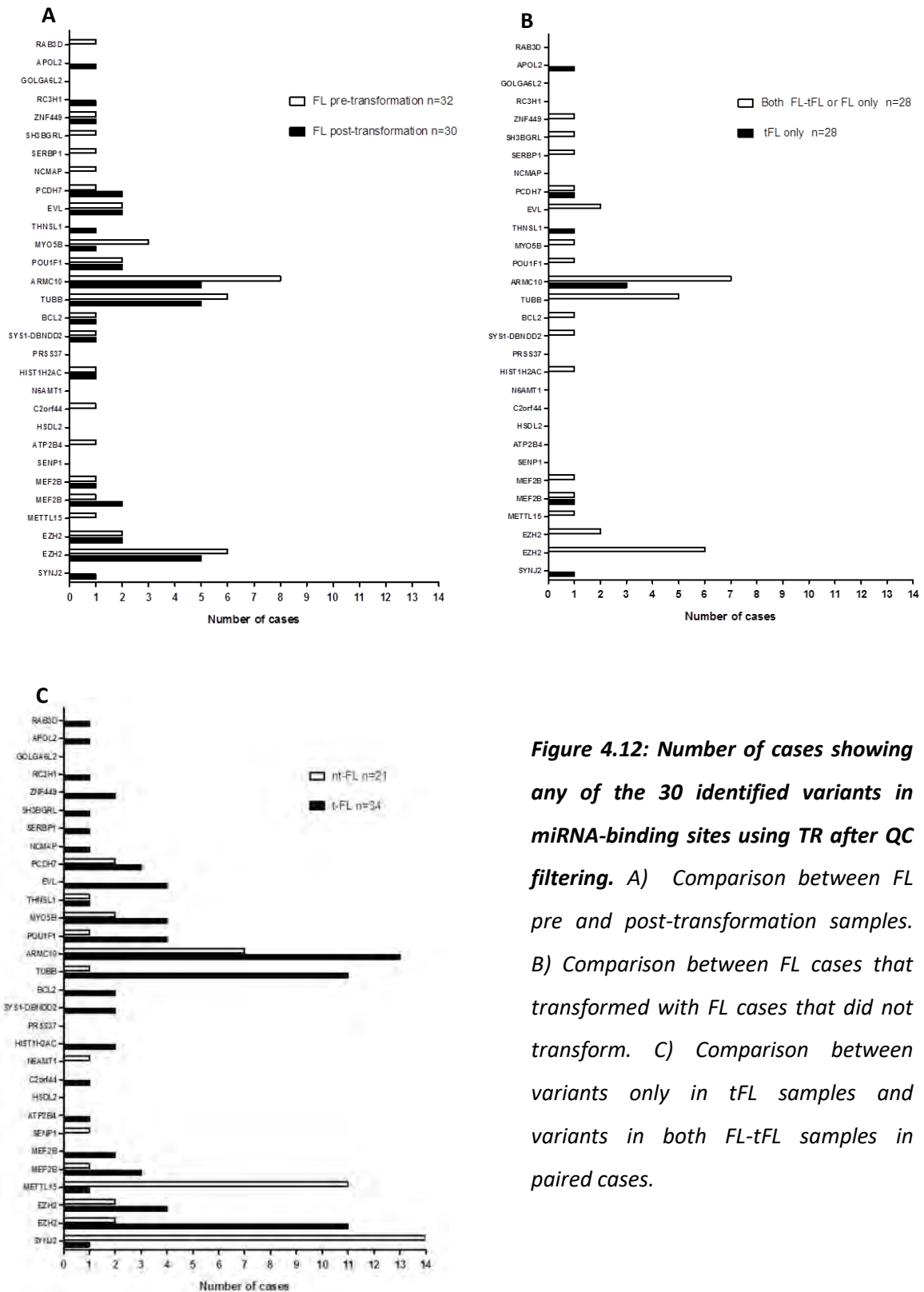
Figure 4.12: Number of cases showing any of the 30 identified variants in miRNA-binding sites using TR after QC filtering. A) Comparison between FL pre and post-transformation samples. B) Comparison between FL cases that transformed with FL cases that did not transform. C) Comparison between variants only in tFL samples and variants in both FL-tFL samples in paired cases.

From the 30 variants, 12 were present in multiple patient samples and four of these were previously reported in dbSNP [347] to be common SNPs (with a minor allele frequency of at least 1%), (*THNSL1, POU1F1, MYO5B,* and *EVL* genes) and so were removed from further analysis. Interestingly, six of the remaining eight variants were originally identified as somatic mutations in the discovery cohort (Figure 4.13).

We Sanger sequenced the eight variants along with an additional three mutations/variants that although detected only in a single case the gene had previously been associated with haematological malignancies. These are BCL2 [96], SPOCK3 [348] and SENP1 [349]. Variants in the genes *TUBB*, *PCDH7* and *SYNJ2* genes were not validated and therefore were excluded from further functional testing (table 4.10).
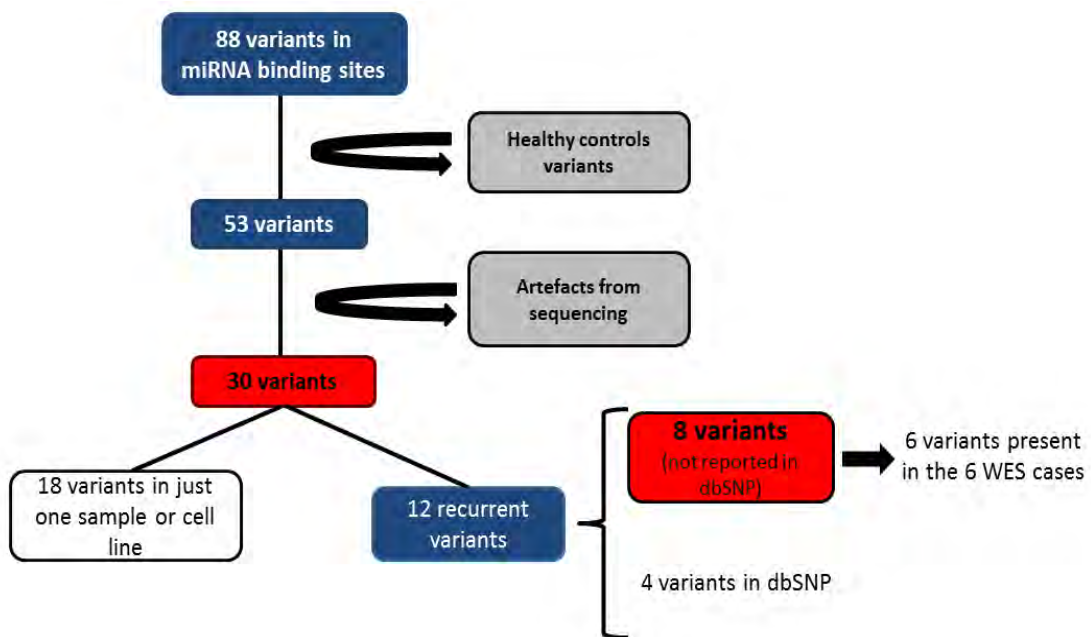
Figure 4.13: *TR variant filtering workflow*

Gene	miRNA	Mutation/Variant (*)	Location	Cases	Frequency (%)
ARMC10	*hsa-mir-222*	*hg19 chr7:g.[102739179A>G]* **Mut**	3'UTR	21	33.3
METTL15	*hsa-mir-4313*	*hg19 chr11:g.[28353434G>A]* **Mut**	3'UTR	13	20.6
EZH2	*hsa-mir-144*	*hg19 chr7:g.[148508727T>A]* **Mut**	Exonic	9	14.3
EZH2	*hsa-mir-144*	*hg19 chr7:g.[148508728A>T]* **Mut**	Exonic	6	9.5
MEF2B	*hsa-mir-1265*	*hg19 chr19:g.[19260045T >A]* **Mut**	Exonic	4	6.3
BCL2	*hsa-mir-5008*	*hg19 chr18:g.[60793447G>A]* **Var**	3'UTR	1	1.6
SENP1	*hsa-mir-196a*	*hg19 chr12:g.[48438315A>G]* **Var**	3'UTR	1	1.6
SPOCK3	*hsa-mir-192*	*hg19 chr4:g.[167655217A>T]* **Mut**	3'UTR	1	1.6
TUBB	*hsa-mir-1302*	*hg19 chr6:g.[30692755_30692756del]*	3'UTR	9	14.3
PCDH7	*hsa-mir-329*	*hg19 chr4:g.[30732984_30732985del]*	Intronic	4	6.3
SYNJ2	*hsa-mir-2355*	*hg19 chr6:g.[158519587del]*	3'UTR	17	27.0

Table 4.10: Initial candidate mutations/variants for functional testing. *Variants excluded for further testing are in grey cells. (*) "**Mut**" refers the confirmed somatic mutations identified in FL tumours and not in the germ line material in the discovery cohort. "**Var**" refers to variants detected in the validation cohort that cannot be identified with certainty as somatic mutations as germ line material from these patients was lacking.*

4.3.5 Functional in-vitro assays demonstrated that the mutations in miRNA binding sites can interfere in the microRNA activity

The luciferase reporter gene system was used to test whether the identified variants affected the binding capabilities of the cognate miRNAs. To do that, we cloned the miRNA-binding region corresponding to the wild type and separately the region containing the variant sequence into a luciferase reporter plasmid. The identified variants in a total of nine miRNA target sites were analysed in this way. Detailed information for each gene, target site and mutation/variant is available in Table 4.2.

The first thing to do was to see whether the predicted miRNA binding site was functional or not by looking at the wild type sequences transfected along with their cognate miRNA. These results showed that two of the prediceted miRNA binding sites (*ARMC10*; chr7:102739177-102739198 and *BCL2*; chr18: 60793644-60793665) did not result in a reduction in luciferase activity when transfected with their respective

miRNAs, *miR-222-5p* and *miR-138-5p* (Figure 4.14). All the other seven predicted binding sites were experimentally validated as the luciferase activity was significantly reduced compared with the control condition when analysing the non-mutated or wild type target sites (Figure 4.15). The effect of the mutations and variants were then analysed in these seven loci.
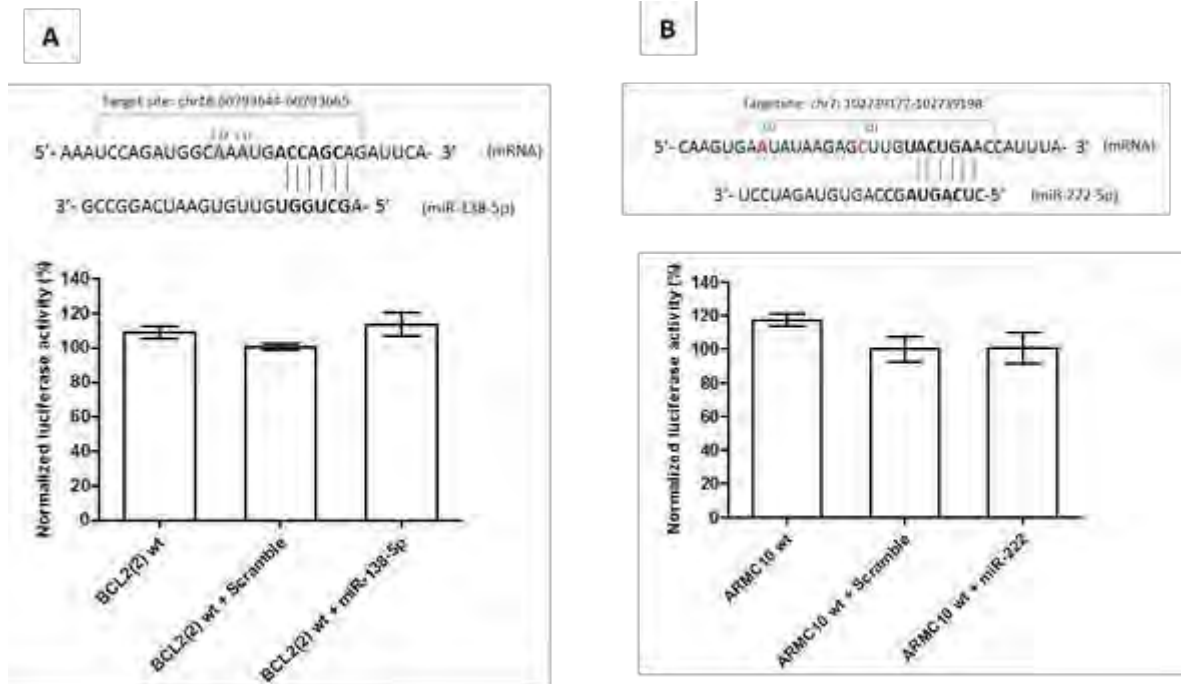
Figure 4.14: Luciferase assay results: Luciferase assays for the predicted miRNA binding sites in BCL2 (A) and ARMC10 (B) didn´t experimentally validate the computational miRNA target predictions. Results don't show any effect in the reduction of the luciferase activity under the miR-138-5p and miR-222-5p targeting respectively.

As can be seen from Figure 4.15 variants present in *MEF2B* [hg19 chr19:g.[19260045T >A] (1), hg19 chr19:g.[19260055T >C] (2)], *BCL2* [*hg19 chr18:g.[60793906G>A]* (1), *hg19 chr18:g.[60793921T >C](2)], METTL15 (hg19 chr11:g.[28353434G >A]), SENP1 (hg19 chr12:g.[48438315A >G])* and *SPOCK3 [hg19 chr4:g.[167655217A >T] (1), hg19 chr4:g.[167655218C >T]* (2)] did not to have a significant effect on the binding of their cognate miRNAs, *miR-1265, miR-27a-5p, miR-4313, miR-196a-3p* and *miR-192-5p* respectively. Interestingly, the variant *hg19 chr4:g.[167655235A >G]* of *SPOCK3* appeared to enhance *miR-192-5p* binding (Figure 4.15).

4. Investigating mutations in the miRNA binding sites of follicular lymphoma
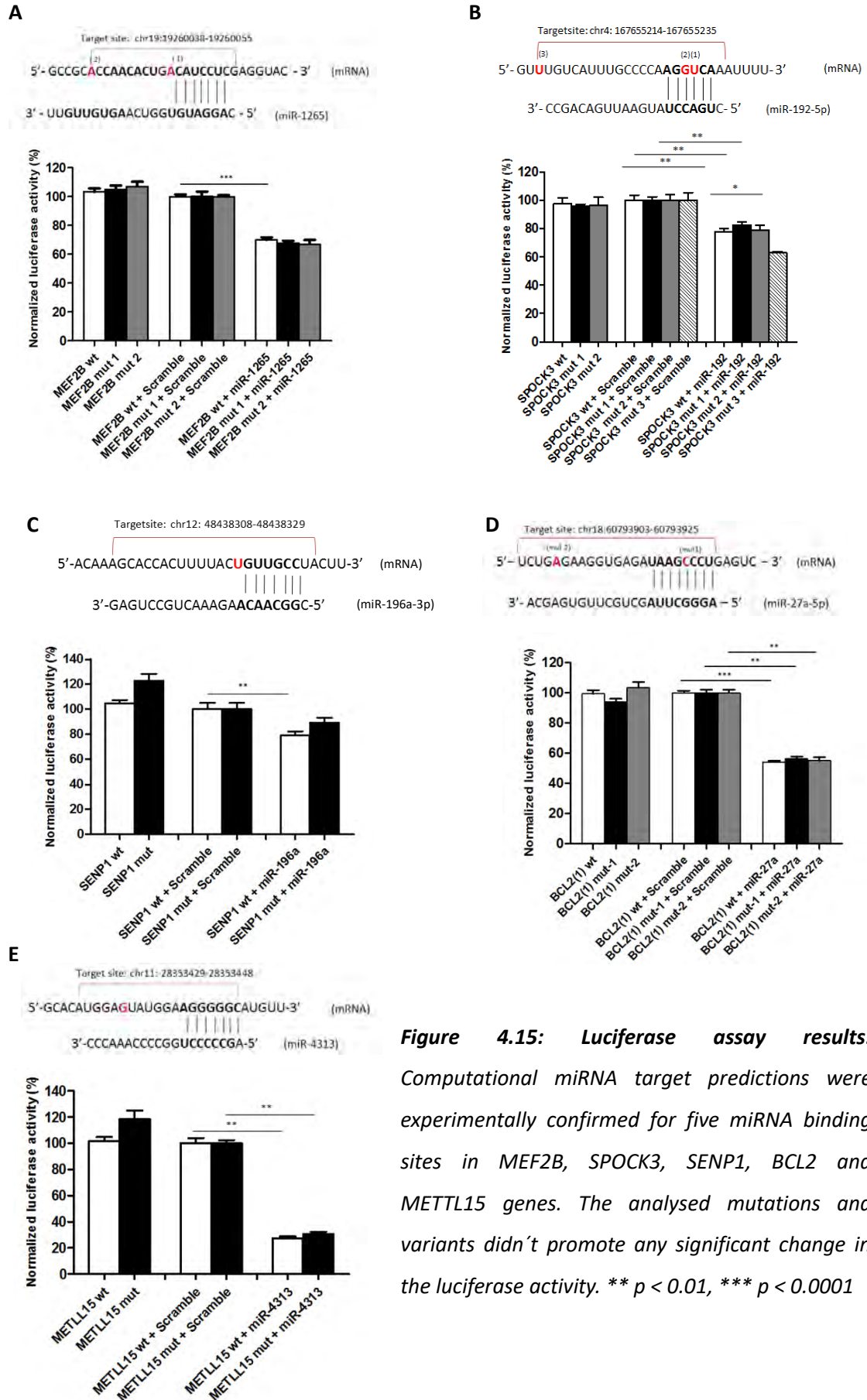
Figure 4.15: Luciferase assay results: Computational miRNA target predictions were experimentally confirmed for five miRNA binding sites in MEF2B, SPOCK3, SENP1, BCL2 and METTL15 genes. The analysed mutations and variants didn´t promote any significant change in the luciferase activity. ** p < 0.01, *** p < 0.0001

In contrast, the mutations of *EZH2* [*hg19 chr7:g.[148508727T>A]* (1), *hg19 chr7:g.[148508727T>G]* (2) and *hg19 chr7:g.[148508728A>T]* (3)] and that of *BCL2* (*hg19 chr18:g.[60793447G>A])* resulted in a significant increase in the luciferase activity compared to the wild type sequence (Figure 4.16).
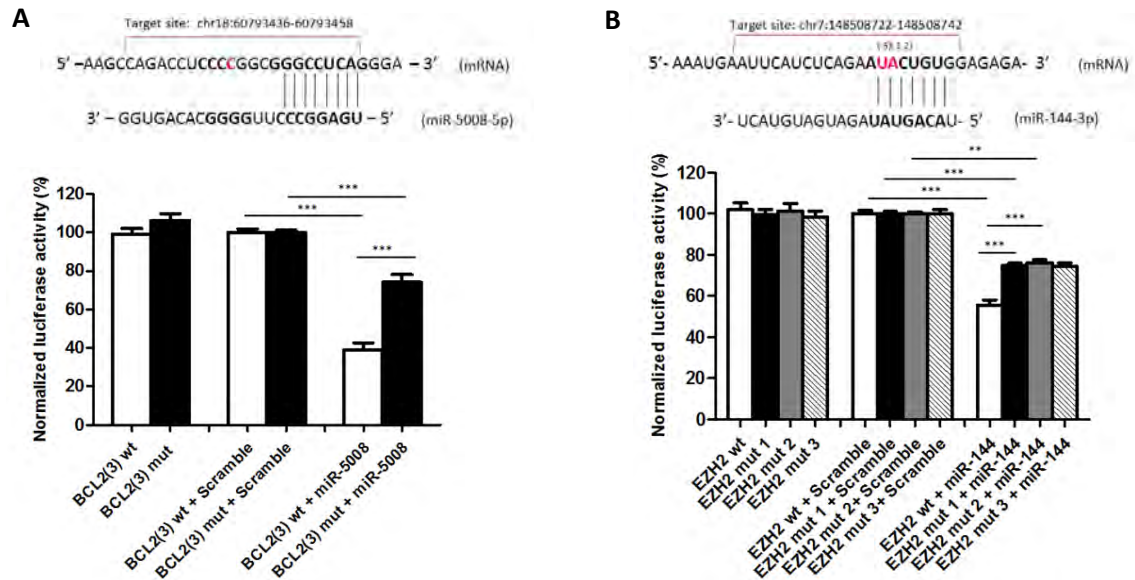
Figure 4.16: Luciferase assay results: *Computational miRNA target predictions were experimentally confirmed for two loci in BCL2 (A) and EZH2 (B) genes. The analysed mutations in EZH2 and the variant in BCL2 did show to promote a significant difference in the luciferase activity. *** p < 0.0001*

4.3.6 Estimating the prevalence of *EZH2* mutations and the *BCL2* variant in FL patients

For analysing the prevalence of the tested variant/mutations with luciferase assays in *BCL2* (*hg19 chr18:g.[60793447G>A])* and *EZH2* (*hg19 chr7:g.[148508727T>A]*, *hg19 chr7:g.[148508727T>G]*, *hg19 chr7:g.[148508728A>T])* among FL patients, together with an additional mutation in EZH2 (*hg19 chr7:g.[148508728A>G]* commonly identified in FL patients (Bodor), we designed and tested Custom Taqman SNP Genotyping Assays on 117 FL patients. A total of 5 Taqman SNP genotyping assays were used, one for the detection of each mutation. Firstly, we estimated the limit of detection for three of the five assays. Using 10 ng of input DNA we determined the assay was able to detect VAF down to 5% (Figure 4.17).

4. Investigating mutations in the miRNA binding sites of follicular lymphoma
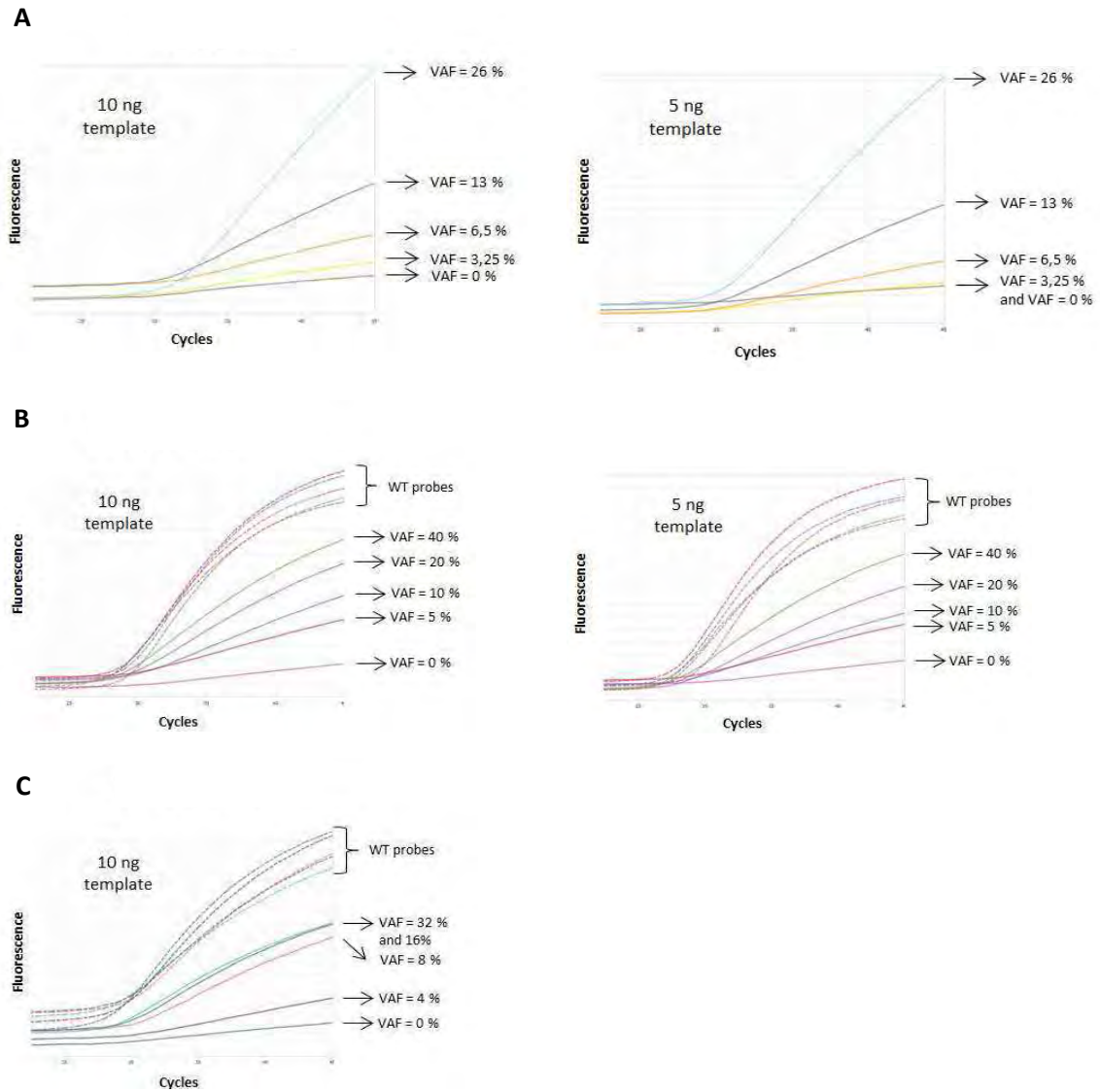
Figure 4.17: Limit of detection estimation for TaqMan SNP Genotyping Assays: *Results for three of the five assays (two targeting EZH2 and one BCL2 gene) demonstrated that this genotyping technique is highly sensitive allowing identifying mutations with a very low VAF (up to 5%) using only 10ng of template DNA. (A) Results for assay 2 targeting hg19 chr7:g.[148508727T>A] mutation in EZH2; 10ng vs 5ng of template. (B) Results for assay 4 targeting hg19 chr7:g.[148508728A>T] in EZH2; 10ng vs 5ng of template. (C) Results for assay 1 targeting hg19 chr18:g.[60793447G>A] in BCL2; 10ng of template.*

Next, we extended the analysis to the 117 FL cases. Unfortunately, the Taqman assay for the detection of *hg19 chr7:g.[148508727T>G]* mutation in *EZH2* gene didn't give any signal so it was excluded from further analysis. qRT-PCR results show a total of 5 samples carrying *hg19 chr7:g.[148508727T>A]* mutation in *EZH2*, with a VAF estimation of around 20-25% for four of the samples and 10% for one sample (Figure 4.18-A).

Similarly, 8 patients were positive for *hg19 chr7:g.[148508728A>T*] mutation in *EZH2* with a VAF estimation of around 20-40% for 6 samples and around 10% for 2 samples (Figure 4.18-C). Also one patient show one sample positively mutated for *hg19 chr7:g.[148508728A>G]* in *EZH2* and one patient carrying the variant *hg19 chr18:g.[60793447G>A]* in *BCL2* (Figures 4.18-D ,B).
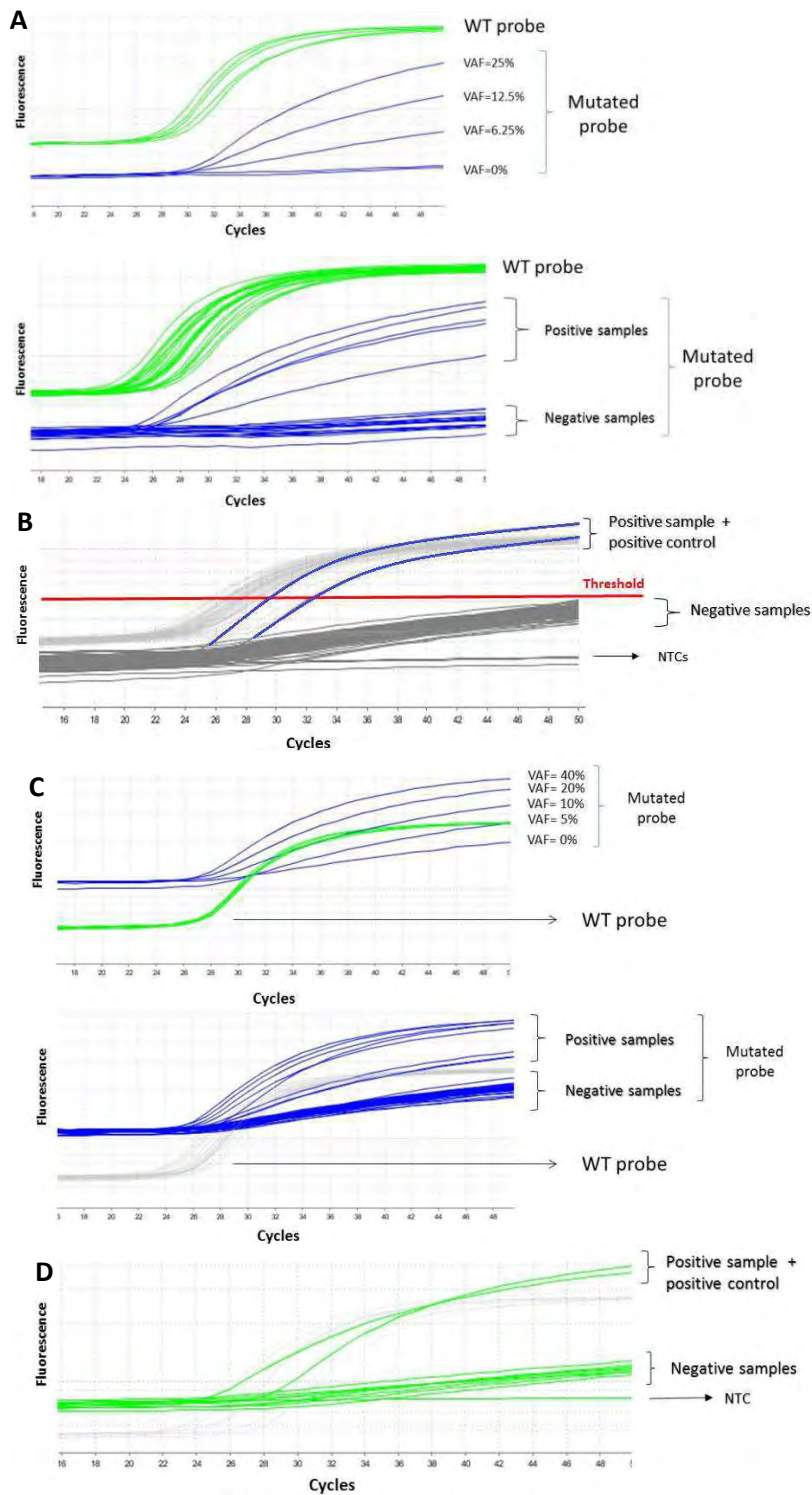
Figure 4.18: TaqMan SNP Genotyping results for mutations/variant in EZH2 and BCL2. *A)*
Genotyping results for hg19 chr7:g.[148508727T>A] EZH2 mutation together with a reference
curve for VAF estimation. B) Genotyping results for hg19 chr18:g.[60793447G>A] BCL2 variant.
C) Genotyping results for hg19 chr7:g.[148508728A>T] EZH2 mutation together with a
reference curve for VAF estimation. D) Genotyping results for hg19 chr7:g.[148508728A>G]

EZH2 mutation. "NTC" indicates non-template-control. "WT" indicates wild type. "VAF" indicates variant allele frequency.

Combining the Taqman Genotyping system and NGS approaches on the 171 FL patients gives us that the mutations in *EZH2* were found to be present in 14% of FL patients whereas the variant in *BCL2* was found in 1.17% of the FL patients (Table 4.11).

Gene	Genomic mutation/variant	Aminoacid change	Mutated samples n= 171	Frequency among FL patients (%)
BCL2	hg19 chr18:g.[60793447G>A]	-	2	1.17
EZH2	hg19 chr7:g.[148508727T>A]	Y641F	11	6.4
EZH2	hg19 chr7:g.[148508727T>G]	Y641S	0	0
EZH2	hg19 chr7:g.[148508728A>T]	Y641N	12	7
EZH2	hg19 chr7:g.[148508728A>G]	Y641H	1	0.58
EZH2	TOTAL EZH2 mutations	-	24	14

Table 4.11: Prevalence of BCL2 and EZH2 variant/mutations. *The table shows an estimation of the frequency of the functionally validated mutations in EZH2 and variant in BCL2 among FL patients (n=117) determined using high-throughput and high sensitivity sequencing approaches.*

4.3.7 Functional validation of this mechanism in lymphoid cell models

In order to test whether this mechanism based on the disruption of the miRNA binding observed in luciferase assays, is also functional in lymphoid cells the effect of transitory transfections of *miR-144* in *EZH2* expression were analysed in a mutated (Y641F) and a wild type *EZH2* DLBCL-derived cell models. When transfected the expression levels of *miR-144* were checked using RT-qPCR to ensure a correct modulation of this miRNA and the effect of this modulation was measured by analysing the expression levels of *EZH2* both at mRNA level (RT-qPCR) and protein level (Western-blot). Results showed that, even though not statistically significant, *EZH2* was down-regulated when *miR-144* was over-expressed (Figure 4.19). In concordance with the results observed in the luciferase assays, the downregulation was stronger in the cell line with wild type *EZH2* (HT cell line) comparing to the cell line with mutated *EZH2* (WSU-DLCL2). This suggests

that the mutation might be interfering with the correct binding of the miRNA to the target site (Figure 4.19).
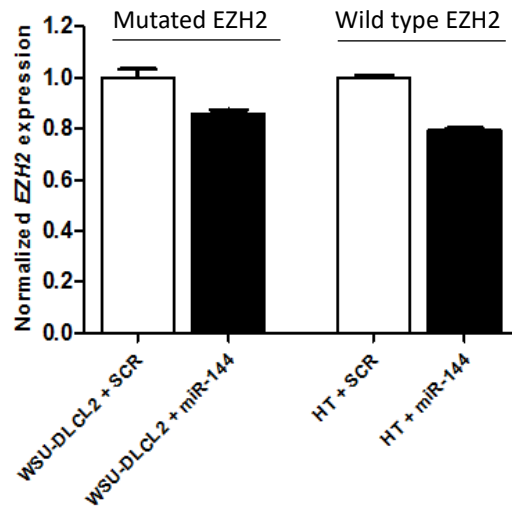
Figure 4.19: EZH2 expression under transitory transfection of miR-144 in lymphoma cell lines. *EZH2 expression levels were measured using RT-qPCR in two different DLBCL cell lines. TATA-binding protein (TBP) expression was used for normalization. The figure shows a downregulation of EZH2 in both cell lines upon miR-144 upregulation. However, the wild type EZH2 is less expressed than the mutated EZH2. SCR refers to transfected with a scramble miRNA sequence.*

Consistent with the qRT-PCR results, *EZH2* protein levels showed no difference in the cell line containing the mutated *EZH2* (WSU-DLCL2), but did express less protein when transfected with *miR-144* in the *EZH2* wild type (HT cell line) (Figure 4.20).
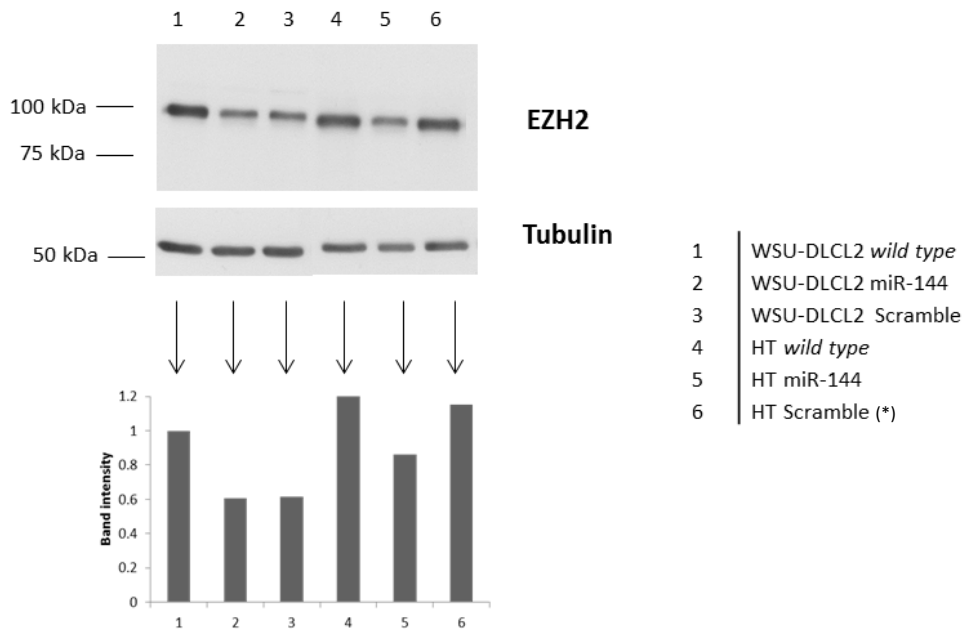
Figure 4.20: EZH2 protein expression analysis under transitory transfection of miR-144 in lymphoma cell lines. *EZH2 protein levels were measured in cell lysates by a western-blot analysis in two different DLBCL cell lines. Results show less expression of EZH2 in the EZH2 wild type cell line (HT cell line) comparing to the control conditions (wild type and scramble). WSU-DLCL2 cell line (mutated EZH2 cell line) shows the same EZH2 levels in the transfected conditions, both with the miR-144 and with the scramble control condition.*

(*) Note: HT scramble sample was infected with a lentivirus expressing a scramble sequence instead of transfected with a scramble mimic

4.3.8 *In-silico* analysis of *miR-144* expression and *EZH2* expression

To see how relevant this mechanism was in FL patients we looked at levels of *miR-144* and *EZH2* in microarray datasets contained within the GEO database [350]. Levels of *miR-144* were significantly downregulated in FL (n=23) and DLBCL (n=29) patients compared to normal B lymphocytes (n=7) although no a difference in miRNA expression between FL and DLBCL cases was detectable (Figure 4.21-A).

4. Investigating mutations in the miRNA binding sites of follicular lymphoma
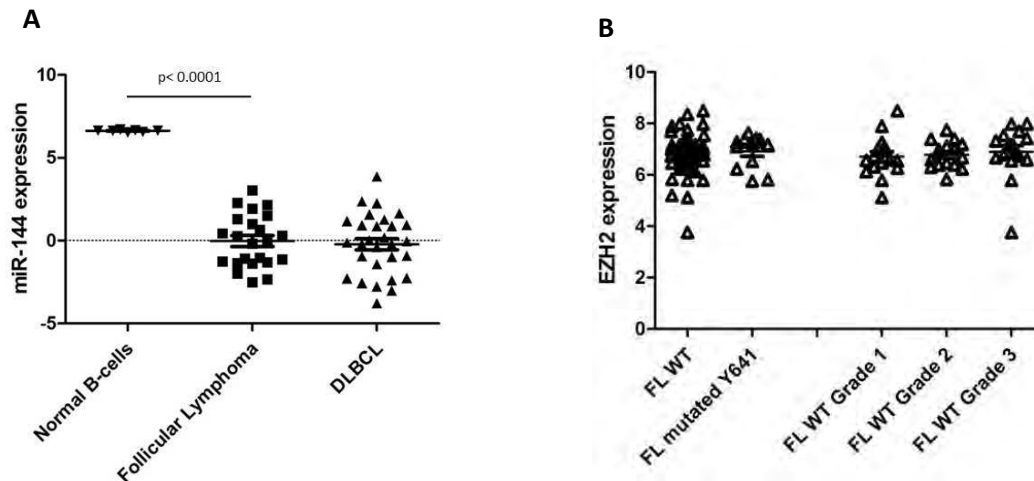
Figure 4.21: *In-silico analysis of miR-144 (A) and EZH2 (B) expression levels in patients using microarray data available in GEO database. Figure A shows the comparison of miR-144 expression levels between normal B-cells, FL patients and DLBCL patients. MiR-144 expression is variable in FL and DLBCL patients and significantly downregulated comparing to normal B lymphocytes. Figure B shows gene expression differences of EZH2 between EZH2 wild type and mutated FL cases together with EZH2 expression in FL cases at different grades.*

Next, we analysed the expression levels of *EZH2*, in 12 FL samples carrying the the Y641 mutation versus samples with a wild type *EZH2* (n=51), and between different clinical stages of FL. There was no significant difference in the levels of *EZH2*, although those cases with mutated Y641 do appear to have higher expression levels of *EZH2* (Figure 4.21-B).

Unfortunately, we couldn't perform an *in-silico* analysis for *miR-5008* and *BCL2* as there was no accessible data for *miR-5008* expression, or expression levels of *BCL2* linked to the *BCL2* genotype in the published datasets.

4.4 Discussion

Mutations in miRNA binding sites have previously been proposed as a pathogenic mechanism in cancer. For example, mutations in *E2F1* and *CFTR* genes showed to alter the affinity for *miR-136-5p* and *miR-509-3p* respectively leading to a deregulation of the targeted gene [332, 333]. However, there have not been similar studies carried out on NHL. In this work we explored WGS data from six longitudinally obtained samples from FL biopsies to identify and explore the role of mutations in miRNA binding sites in FL. Our results indicated that of the total 788 mutations identified in 3'UTRs, 68% of them were within bioinformatically predicted miRNA-binding sites. Interestingly, mutations in the miRNA gene sequence themselves were not identified. The possible reason could be that, miRNA coding genes are normally very conserved sequences [262]. Genes with mutations in miRNA-binding sites were highly enriched for genes involved in hematological malignancies, suggesting that mutations are not random, and giving weight to their functional importance in lymphomagenesis. Morever, 5 of the 6 cases had more miRNA binding site mutations in the transformed FL compared to the antecedeent FL. This could suggest that these mutations are acquired during the transformation process and may play a role in this process.

Sanger sequencing of 34 mutations resulted into validation of only approximately one third. This apparently poor validation is probably because of the low sensitivity of Sanger sequencing compared to NGS. In Sanger Sequencing, mutations with a VAF lower than 15-20% are not detectable as they become indistinguishable from the background noise [351]. We found that from the positively validated mutations, those located in coding-region of the *EZH2* and *MEF2B* genes were present in multiple cases (10/49 and 3/49 respectively). These mutations are missense mutations that have previously been described as gain-of function-mutations [80, 352].

Using deep targeted re-sequencing instead of Sanger sequencing, we validated 80% of the original mutations identified by WGS, even though the platforms use very different chemistries (Figure 4.22). As part of our QC filtering strategy we included the removal of variants near homopolymer and repetitive regions. The reason for this is that a substantial fraction of indel errors are caused by homopolymers with the base-calling

accuracy decreasing with homopolymer length. This is especially true for homopolymers of 7-8 nucleotides or more [353, 354]. After QC filtering, we found that the identified mutated genes (n=28) were highly enriched for GC-like B-cell lymphoma genes, strongly suggesting that these variants may have a biological significance in the disease.
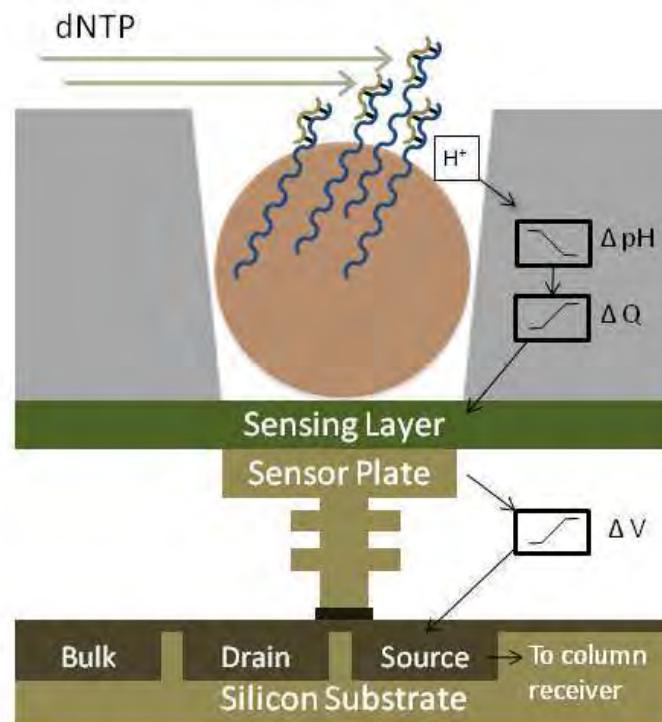
Figure 4.22 Ion Chip well architecture. *Clonally amplified libraries attached to the beads are loaded into Ion Chips. Each bead will be deposited into a well on the chip. The chip is flooded with one of the four nucleotides sequentially. Whenever a nucleotide is incorporated into the single stranded DNA molecule a hydrogen ion is released. This release of ions will trigger a change in the pH of the solution that can be detected by the sensing layer in each well. Voltage changes are translated into digital information for base calling (JM Rothberg et al. Nature. 2011) [355].*

For the luciferase assays, the genes with variants present in multiple patients that were not reported as common SNPs were selected (7/28 genes) as well as three additional genes related with haematological malignancies with mutations in only one case. The genes with variants that were not validated with Sanger seqeucing were excluded (n=3). Therefore, luciferase assays were performed in a total of nine loci in seven

candidate genes. This technique allows the *in-vitro* testing of the binding capacity of a miRNA to a particular target site. Interestingly, we found that the effect of the variant on the miRNA-binding was not associated with its position within the binding site. For example, variants located in the predicted seed region of the binding site, such as *hg19 chr4:g.[167655217A>T]* and *hg19 chr4:g.[167655218C>T]* in *SPOCK3* and *hg19 chr18:g.[60793906G>A]* in *BCL2*, had no affect on miRNA binding, whereas the variant *hg19 chr18:g.[60793447G>A])* present in *BCL2*, but outside of the predicted seed region, did strongly affect miRNA binding. These observations suggest that the mechanism of miRNA binding remains to be completely understood and underlies the difficulties in miRNA-target gene prediction (the subject of the next chapter). Indeed, widespread non-canonical microRNA targeting has been experimentally identified [274, 356, 357]. Importantly, luciferase assays allowed us to functionally validate this mechanism for two key genes in B cell lymphomagenesis, *EZH2* and *BCL2*.
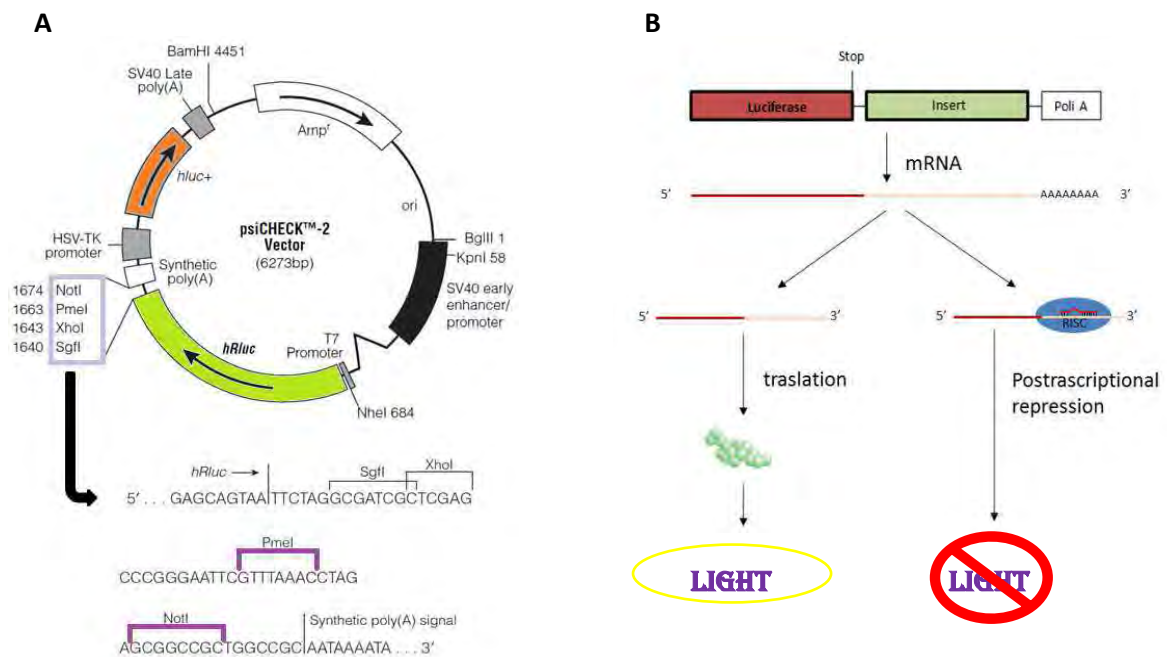
Figure 4.23: Luciferase assay. *A) psiCHECK-2 vector map showing the sequence of the multiple cloning regions after the stop codon of Renilla luciferase gene. Firefly luciferase is used as a control for the transfection. B) Schematic representation of luciferase assay. Under the presence of a miRNA targeting the 3'UTR cloned sequence a reduction in the emitted amount of light is detected.*

4. Investigating mutations in the miRNA binding sites of follicular lymphoma

EZH2 is a well characterized histone methyltranferase which is upregulated in B cells during the GC reaction promoting cell proliferation [79]. Activating somatic mutations in the catalytic SET domain of EZH2 have been identified in 25% of follicular lymphoma patients [80]. The vast majority of them affect Y641 codon [125], which is the same codon as the mutations we tested in this chapter. We therefore hypothesised that these mutations, apart from altering the protein function as previously described, might also alter their own expression by impairing the miRNA targeting activity of *miR-144*. Indeed, it has been reported that these *EZH2* gain-of-function mutations associated with H3K27me3 hyperactivation also increase the stability of the protein [358]. *EZH2* has been already reported as a direct target of *miR-144* in bladder cancer, osteosarcoma and astrocytoma [359-361]. Some of these studies show an inverse correlation between *miR-144* and *EZH2* expression in samples of patients, with *miR-144* usually down-regulated and *EZH2* up-regulated [360, 361], as well as a strong correlation with aggressiveness of the associated tumour [360].

We carried out functional assays to investigate the *EZH2* Y641F/*miR-144* axis in lymphoid cell models. When *miR-144* was over-expressed in DLBCL-derived cell lines, *EZH2* was down-regulated of at mRNA and protein levels. This was not the case for cell lines carrying the Y641F mutation. We would have liked to follow up these results in FL patients. However, tumor RNA samples were not available for the vast majority of the cases. Therefore, we decided to perform an *in-silico* analysis to look at *EZH2* expression and genotype as well as *miR-144* expression levels in FL patients. Results indicated that *miR-144* is downregulated in FL and DLBCL patients but this expression is very variable suggesting that this mechanism might be effective for patients with higher *miR-144* levels. Moreover, a significant role for *miR-144* in B-cell lymphoma has already been demonstrated through its targeting of *BCL6* [362].

Nevertheless, it has to be taken into consideration that mutations in *EZH2* are located at the end of the coding region so the accessibility to the miRNA silencing machinery might be lower than if it was located in the 3'UTR. In fact, the luciferase assay proved that this sequence might be targeted by *miR-144* in physiological conditions but in a hypothetical short (~300pb) 3'UTR so this should be considered as well. Also, the fact that the assay was performed in HEK293 cell line, which is a cell model widely accepted

to do this type of assays, but it the binding itself was not proven in a lymphoid cellular context.

BCL2 is an anti-apoptotic protein that participates in the GC B-cell selection [32]. The t(14;18) translocation leads to constitutive expression of BCL2 protein as it comes under control of the IGH promoter region. This event is considered the hallmark of FL as it is present in 85% of FL patients [63]. However, some 10-15% of FL cases lack the t(14;18) translocation, but curiously also display high expression of BCL2 protein by some unknown mechanism [363]. We hypothesised that changes in the miRNA binding site could provide an explanation. Unfortunately *in-silico* microarray data was not available for this miRNA (*miR-5008*), was there data available regarding the genotype of FL patients for this chromsomal position, or indeed whether patients are t(14;18) translocated. However, the chromosome gain of 1q arm, where *miR-5008* is encoded, has been described for 20% of FL [95]. It could also be envisaged that such as mechanism might only be relevant for FL patients lacking t(14;18). However, as only 10-15% of cases are negative for this translocation, it makes it difficult to find sufficient material to prove this hypothesis [63]. Among our cohort 181 cases, only two cases contained the variant but we had only DNA material available and were not able to measure the expression levels of *miR-5008* and *BCL2* for these patients. We also tried to find clinical data from these samples to find out if they were negative for t(14;18) but unfortuanatley were not able to. For FL patients with t(14;18) the major breakpoint region (MBR), this proposed mechanism is unlikely to be very important as this breakpoint occurs in the 3'UTR upstream of the variant location so the part of the 3'UTR where the variant is located is not translocated together with *BCL2* gene [364]. That said, the MBR occurs only in 50% of the patients with the t(14;18) translocation. The remaining 50% of patients exhibit minor downstream breakpoints [365], which means that this variant could be of more relevance for those patients. It would be interesting to test this hypothesis further.

Similarly, the accessibility of the binding site identified in *BCL2* gene should be also considered. This binding site is located very centered in a long 3'UTR region so secondary structures of the mRNA might occur decreasing its accessibility to miR-5008. Also, germline material from the validation cohort was lacking so, we cannot truly

confirm that this variant is indeed a somatic mutation. However, as the VAFs for this variant were 32% and 18% in FL and tFL samples respectively, it is likely to be a somatic mutation as, if it was a germline polymorphism we would expect VAFs of 50% or 100%.

In summary, in this chapter of the thesis we demonstrated that activating mutations in *EZH2* as well as a variant in the 3'UTR of *BCL2* identified in FL patients impair the binding of *miR-144* and *miR-5008* respectively promoting a deregulation of these genes in the performed *in-vitro* assays. This is a potentially novel mechanism in FL that requires further investigation to assess its relevance to FL patients.

5. ELUCIDATING THE MIRNA-TARGETOME OF DLBCL

5.1 Background

Aberrant expression of microRNAs (miRNAs) is a widespread phenomenon in cancer [186, 203-205]. However, the functional significance of such deregulation is poorly understood as the target genes (the targetome) of miRNAs are notoriously difficult to predict computationally, and moreover differ according to the cellular context [283, 286]. An alternative approach is to directly sample the targetome *in situ* using cross-linking immunoprecipitation (CLIP) techniques coupled with high-throughput sequencing. This technology has evolved through the development of several variations with arguably the most promising being Photoactivatable-Ribonucleoside-Enhanced CLIP (PAR-CLIP) which has far better signal-to-noise ratio than other CLIP-based technologies [292]. PAR-CLIP improves the crosslinking strategy by incorporating photoactivatable ribonucleosides into RNA synthesis thereby introducing characteristic mutations (T-to-C) indicating the positions where a crosslink between the incorporated ribonucleoside and a protein occurred [296]. In the original PAR-CLIP protocol HEK293 exogenously expressing a tagged protein was used for immunoprecipitation [296, 366, 367]. Subsequently this technique has been used in varous cell models [299, 368] although to date there are only two studies have used this approach in lymphoid cells to study the target genes of miR-17~92 in human B-cells inmortalized by Epstein-Barr virus [368, 369] and to study the viral targetome of Kaposi's sarcoma-associated herpesvirus and Epstein-Barr virus in primary effusion lymphoma cell lines [368, 369].

As our original aim was to look at the targetome of primary lymphoma cells which would be difficult to isolate in large numbers, and exogenously express tagged proteins would be near impossible to produce in these cells, our objectives were to 1) reduce the number of cells needed for PAR-CLIP and 2) use only endogenous produced Ago2 protein for IP. Furthermore, in order to reduce the signal to noise ratio further in our modified protocol we aimed to systematically perturb levels of key miRNAs (starting with *miR-155*) by comparing the targetome of cells with over-expressed miRNAs with those where the miRNA was inhibited, in addition with wild type cells. We started this project by looking at the targetome of *miR-155* in DLBCL. The oncogenic function of *miR-155* in B-cell lymphoma is well known [243-246]. This miRNA plays a key role in normal hematopoiesis and lymphomagenesis (see section 1 for more details).

Furthermore, the expression of *miR-155* differs between the molecular subtypes of DLBCL with higher expression in ABC-type DLBCL than GCB-DLBCL [370, 371]. Therefore, for these experiments we used both GCB-DLBCL cell line (SU-DHL-10) and ABC-DLBCL cell line (RIVA) models.

5.2 Materials and methods

5.2.1. Analysis of puromycin sensibility of DLBCL cell lines

The intrinsic puromycin sensitivity of SU-DHL-10 (GCB-DLBCL cell line) and RIVA (ABC-DLBCL cell line) was measured in order to ensure a correct puromycin selection of the lentiviral-infected cells to stably modulate *miR-155* levels. To do that, cell lines were seeded in 96-well plates at a concentration of 80,000 cells/mL (8000 cells/well) and were exposed to increasing concentrations of puromycin in the culture media (0-20 ng/µL). Cell viability was analysed after 3 and 6 days using MTT colorimetric assays (Sigma-Aldrich, Missouri, USA). Absorbance was measured in a Multiskan Ascent plate reader (Thermo Fisher Scientific, DE, USA).

5.2.2 Lentiviral infection: MOI optimization and monitorization of *miR-155* expression

DLBCL cell lines were transfected with lentiviral vectors to get three different conditions from each cell line. These vectors encode: the *pre-miR-155* sequence, a scramble pre-miRNA sequence (control), and a short hairpin RNA that, after processing, produces an anti-sense *miR-155* sequence in order to inhibit *miR-155* expression. Plasmid vectors encoding for the pre-miRNA constructs were purchased from GeneCopoeia (Rockville, Maryland, USA) (CmiR001-MR03 and HmiR0358MR0310), and the plasmid vector encoding for the anti-sense *miR-155,* a miRZip knockdown vector, was purchased from System Biosciences (Palo Alto, California, USA) (MZIP-155-PA-1). All the vectors also encode a gene for puromycin resistance and a green fluorescent protein (GFP) gene reporter (Figure 5.1).
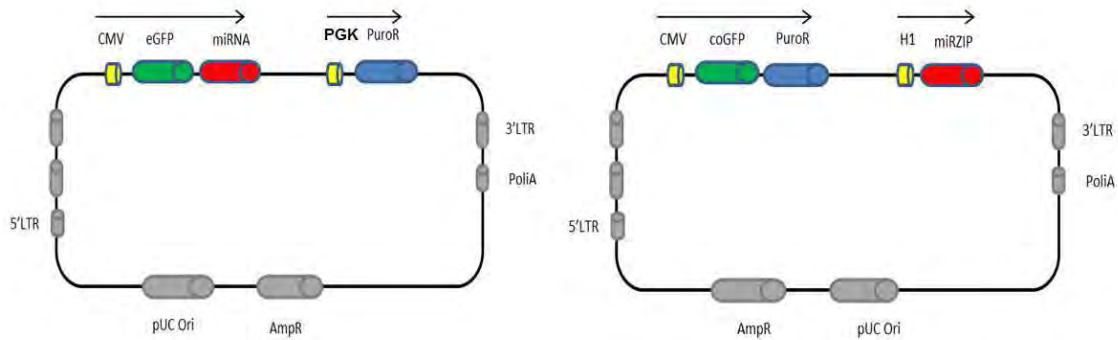
Figure 5.1: Schematic representation of the plasmid vectors for lentiviral production. *A) Backbone for the lentivectors encoding the pre-miR-155 sequence (over-miR-155 condition) and a scramble pre-miRNA sequence (scramble or control condition). B) Backbone for the lentivector encoding an anti-sense sequence for miR-155 (anti-miR-155 condition).*

Viral particules were produced as indicated in section 3.12. For the viral infections different MOIs were tested in order to optimize the infection process. A total of 100.000 cells were infected for each condition. Cells were resuspended in media containing the viral particles and were seeded in 12-well plates. 24 hours after the infection, the media was replaced for fresh media. After 72 hours post-infection, puromycin was added to the culture at a concentration of 1 ng/μL to select the infected cells. The puromycing was kept in the culture media for 6 days. Cells were infected at four different MOIs: 15, 30, 45, and 60 using 1.5, 3, 4.5 and 6 x 10^6 viral particles respectively for the 100.000 infected cells per condition.

In order to evaluate the efficiency of the infection for each MOI condition the GFP expression was analysed one week after the infection using flow cytometry (FACSDiva version 6.1.3). The GFP-positive cells were classified according to expression as either low, medium or high. Dead cells were stained using Topro3 (ThermoFisher Scientific, DE, USA). Then, SU-DHL-10 infected cells at MOI 15 and MOI 60 were stably selected with puromycin to analyse the expression levels of *miR-155* by qRT-PCR as described in sections 3.5.1 and 3.6.1 using the snoRNA RNU6B as house-keeping gene. For subsequent experiments, DLBCL cell lines were infected at a MOI of 15.

5.2.3 Analysis of Ago2 expression

Cell lysates from DLBCL cell lines (SU-DHL-10 and RIVA) as well as the HEK293 FLAG-HA-Ago2 cell line were obtained as described in section 3.13.1. For the induction of Ago2 in HEK293 FLAG-HA-Ago2 cell line, cells were cultured with 1µg/mL Doxycycline as indicated in section 3.11.1.

To measure Ago2 expression we carried out Western-blot analysis as described in section 3.13. Total protein loading was visualised by Ponceau S staining of the nitrocellulose membrane (Sigma-Aldrich, Missouri, USA). Membranes were blocked with 5% skimmed Milk in TBS-Tween 0.1% and probed with either an anti-Ago2 monoclonal antibody (Clone 11A9, Merck Millipore, Darmstadt, Germany) or anti-FLAG monoclonal antibody (Clone M2, Sigma-Aldrich, Missouri, USA) both diluted 1:2000 in in 5% skimmed milk TBS-Tween 0.1%. As secondary antibodies goat anti-rat IgG-HRP and goat anti-mouse IgG-HRP (Santa Cruz Biotechnology, Texas, USA) diluted 1:2000 in in 5% skimmed milk TBS-Tween 0.1% were used respectively.

Three washes of 5 minutes in TBS-Tween 0.1% were performed followed by an incubation with the secondary antibody (conjugated with HRP) diluted 1:2000 in 5% Skim Milk TBS-Tween 0.1% for 1 hour at room temperature in rotation in 50mL Falcon tubes.

5.2.4 4-SU toxicity assays on DLBCL cell lines

4-ThioUridine (4-SU) was purchased from Santa Cruz Biotechnology (Heidelberg, Germany) (ref. sc-204628A), and dissolved in DMSO.

SU-DHL-10 and RIVA cells were seeded in a 96-well plate at a concentration of 500.000 cells/mL (50.000 cells/well) and were exposed to increasing concentrations of 4-SU (0, 100, 200, 400 µM) in the culture media for 18 hours in biological triplicates. Toxicity was assessed by counting live cells using Trypan Blue staining.

5.2.5 Analysis of the incorporation of 4-SU into the nascent RNAs

DLCBL cell lines were cultured for 18 hours in the presence of 100µM 4-SU. Cells were pelleted and RNA extracted using TRIzol reagent (Invitrogen). Biotinylation of 5µg of

RNA was perfomed using EZ-Link™ Iodoacetyl-LC-Biotin (Cat number: 21333, ThermoFisher Scientific, Waltham, Massachusetts, USA) according to the manufactures instructions. Biotinylated RNA was then purified using RNAeasy MinElute Cleanup kit (Qiagen). 250ng of the biotinylated RNA (5uL vol) were spotted onto a Amersham Hybond-N$^+$ blotting membrane (GE Healthcare). A 5' biotinylated DNA oligo (80 nt) was used as a positive control. The membrane was incubated for 30 minutes in PBS 10% SDS and 15 minutes more in the same solution containing HRP-Conjugated Streptavidin (dil 1:10,000) (cat number: N100, ThermoFisher). The membrane was washed in PBS with decreasing amounts of SDS (10% - 0.1%) and developed with Novex ECL (HRP Chemiluminescent Substrate Reagent kit, Cat no. WP20005, Invitrogen).

5.2.6 Immunoprecipitation control and Western-blot analysis

Cell pellets (40 million cells per condition) were resuspended in 3 volumes of NP40 lysis buffer and incubated on ice for 30 minutes. Cell lysates were centrifuged to remove undissolved material, followed by filtration with a 5 μm pore syringe filter. The lysate was then treated with RNAse T1 (Fermentas, Waltham, Massachusetts, USA, EN0541) for 10 minutes at 22ºC. Half of the lysate was incubated with Anti-AGO2-coated magnetic beads (Clone 11A9, purchased from Ascenion, Munich, Germany), and the other half with rat IgG2a (ThermoFisher Scientific, DE, USA, 02-9688), an isotype control for the immunoprecipitation (IP). Samples were incubated on a rotating wheel for 18h at 4ºC. The following day, the IP fraction not bound to the beads was stored at - 80ºC. The beads were washed using a Dynal MPC-S magnetic particle concentrator and resuspended in 50uL sample buffer for the SDS-PAGE which contains β-mercaptoethanol for protein denaturalization and incubated 5 minutes at 95ºC. Half of the volume (25uL) was loaded onto the gel.

As controls 30 μg of cell lysates from the total lysate fraction and the IP negative fractions (or unbound fractions) were also denatured and loaded onto 8% polyacrylamide gels. Western-blot was performed as indicated in section 3.13. For Ago2 staining the monoclonal (Clone 11A9, Merck Millipore, Darmstadt, Germany) anti-Ago2 antibody was used.

5.2.7 Gene expression profiling of DLBCL cell lines with the *miR-155* modulation using microarrays

DLBCL cell lines (SU-DHL-10 and RIVA) were infected in biological triplicates with the lentiviral vectors for the modulation of *miR-155* expression [encoding *miR-155* sequence (over-miR-155), a scramble sequence (scramble) and an inhibitor of *miR-155* (anti-miR-155)] and total RNA from these cell lines was extracted (see section 3.1.3). A total of 500 ng of RNA treated with DNase I (Invitrogen Ref: 18068-015) were used for the sample preparation using GeneChip™ WT PLUS Reagent Kit (Applied Biosystems™, California, USA) in accordance with the manufacturer's instructions. Samples were hybridised to Clariom™ D microarrays (Applied Biosystems™, California, USA) using a GeneChip® Fluidics Station 450 (Affymetrix, CA, USA). Data analysis was performed in R. The generated CEL files were pre-processed and RMA-normalized using oligo package [372]. Transcripts annotated by Aceview were removed from the analysis. Also, the 40% of probe sets with the smallest interquartile range. The fold-change of the identified genes was measured between over-miR-155 and scramble condition. The top 20 genes showing downregulation in the over-miR-155 condition comparing to the scramble condition with the largest fold-change were prioritized.

5.2.8 PAR-CLIP modified protocol

Stably-transduced cells were cultured for 18h in the presence of 100 µM 4-SU. The original protocol by Spitzer *et al.* [366] was modified to avoid radioactive labelling and improve the sensitivity of the technique. In brief, lymphoid cells were harvested, pelleted and resuspended in cold PBS before being UV irradiated (365 nm light, 150 mJoules/cm²) on ice. Cross-linked cells were centrifuged, and pellets snap-frozen and stored at -80ºC. Frozen cell pellets were resuspended in 3 volumes of NP40 lysis buffer, on ice for 30 minutes. Cell lysates were centrifuged and then filtered through a 5 µm syringe filter.

Lysates were treated with RNAse T1 (Fermentas, Waltham, Massachusetts, USA, EN0541) for 10 minutes at 22ºC. Anti-AGO2-coated magnetic beads were then added to the digested cell lysate (Clone 11A9, purchased from Ascenion, Munich, Germany,

MABE253), and incubated on a rotating wheel for 18h at 4ºC. The following day, beads were retained using a Dynal MPC-S magnetic particle concentrator, resuspended in washing buffer, and then subjected to an additional RNAse T1 digestion (10U/μl for 10 minutes at 22ºC), followed by several washes (following the original protocol) until finally resuspensed in NEB3 buffer. The resultant RNA was then dephosphorylated using Calf Intestinal alkaline Phosphatase (CIP) (New England Biolabs, Ipswich, Massachusetts, USA) at a concentration of 0.5 U/μl for 1 hour at 37ºC, followed by end-repairing with T4 Polynucleotide Kinase (NEB, M0201) and 100 μM ATP for 35 minutes at 37ºC.

AGO2-bound beads were denatured with β-mercaptoethanol for 5 minutes at 95ºC and proteins were run through a 4-12% Bis-Tris PAGE alongside a protein standard (Precision Plus Protein Kaleidoscope, Bio-Rad). Gel sections corresponding to size 75-100KDa were excised from each lane. These pieces werethen minced to ~1 mm^3 pieces, and finally transferred to 300K Pall Nanosep Omega membrane columns (Sigma-Aldrich, St. Louis, Missouri, USA, Z722170). The RNA-AGO2 complexes were then eluted from the gel pieces through centrifugation using 0,1% SDS. Eluants were then transferred onto 10K Pall Nanosep Omega membrane columns (Sigma-Aldrich, Z722065), and recovered in 100 μl nuclease-free water. RNA-protein complexes were dislodged using proteinase K. RNA fragments were then concentrated using the mercury RNA Isolation Kit Biofluids (Exiqon, Vedbaek, Denmark, 300112), and eluted in 2x20 μl nuclease-free water (Figure 5.2).
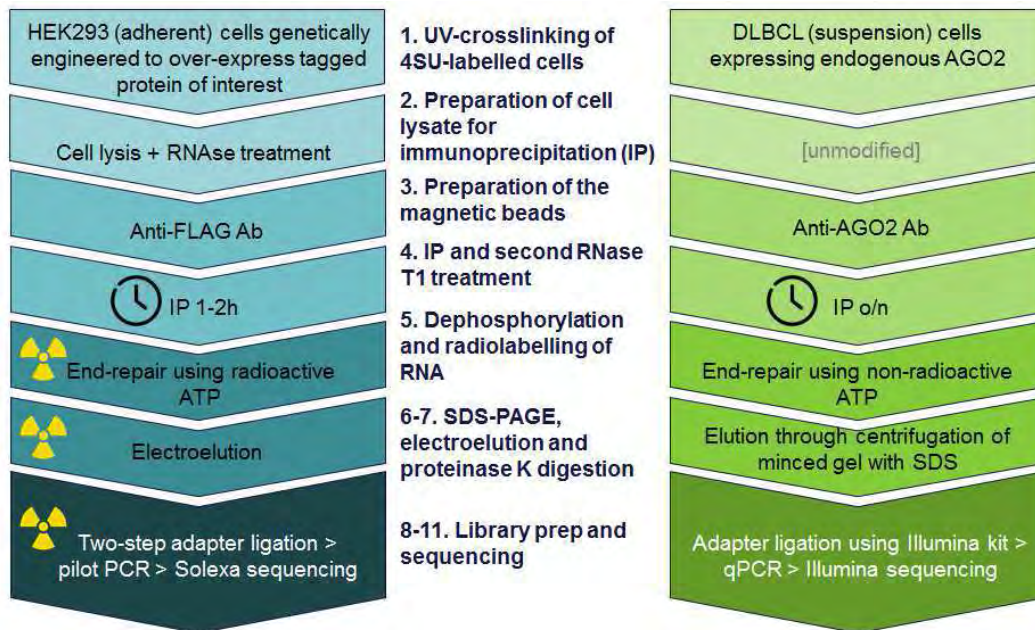
Figure 5.2: Outline of PAR-CLIP protocol. *PAR-CLIP steps are listed 1-11 following the outline of the original protocol [306]. The blue flowchart on the left highlights the main differences with the protocol used in the present work, which is summarised on the green flowchart on the right.*

5.2.9 Library building and sequencing

The quantity and quality of the RNAs were evaluated using a 2100 Bioanalyzer (Agilent, Santa Clara, California, USA). Libraries for sequencing were prepared using *TruSeq Small RNA Sample Prep Kit* following the *TruSeq® Small RNA Sample Preparation Guide* (Illumina, San Diego, California, USA). Briefly, ligation of the 3′ adapter was conducted by incubating total RNA of each sample with the adapter for 2 minutes at 70ºC. Then the 5′-adapter was ligated using a truncated T4-RNA ligase 2 (New England Biolabs, M0351S) and incubated at 28ºC for 1 hour. The total volume of the ligation product was used in a double reverse transcription reaction with SuperScript II reverse transcriptase (Life Technologies, 18064-014) in a thermocycler for 1 hour at 50ºC. Next, enrichment of the cDNA was performed using PCR cycling: 30 sec at 98ºC; 17 cycles of 10 sec at 98ºC, 30 sec at 60ºC and 15 sec at 72ºC; a final elongation of 10min at 72ºC. PCR products were resolved on 8% Novex TBE PAGE gels (Life Technologies, EC6265BOX), and fragments between 145 and 350 bp cut from the gel. Small RNAs were extracted from polyacrylamide gel using an adapted protocol for MinElute gel extraction kit (Qiagen, 28604), in which gel slices were dissolved in diffusion buffer

(0.5M ammonium acetate; 10mM magnesium acetate; 1mM EDTA, pH8.0; 0.1% SDS) overnight at room temperature plus 3 hours and 30 min at 50ºC. A second round of gel-purification was performed in order to completely eliminate empty-adapters. The libraries were then visualized on an Agilent 2100 Bioanalyzer using the Agilent High Sensitivity DNA kit (Agilent, G2938-90320) and quantified using Qubit® 2.0 Fluorometer (ThermoFisher Scientific). The libraries were pooled, and hybridised onto three lanes, at 9 pM (each library was expected to be sequenced at 1.782 pM per lane). High-throughput sequencing was carried out in a HiScanSQ platform (Illumina) to generate single-ended 50 bp reads.

5.2.10 Sequencing data analysis

Raw data were clipped, and reads shorter than 20 nt were removed. All fastq files from the same sample were then merged, and reads were collapsed to remove duplicates. Sequence reads were then aligned to the human genome (hg19/GRCh37) using Bowtie1 (-v 2 –m 10 –best –strata). Sam files containing aligned reads were converted to sorted indexed bam files using samtools. Subsequently, clusters (overlapping reads aligning in the same locus) were identified using three different software's specialized in CLIP data analysis: wavClusteR [373], BMix [374], dCLIP [375]. In wavCluster and BMix analysis a minimum coverage of two was applied. In the analysis using dCLIP a minimum number of tags were established as five both for the first and second condition. The mutational profile of the aligned reads was obtained using wavClusteR tool. Candidate genes were identified using Annovar software tool [376]. For the visualization of the clusters the Integrative Genomics Viewer (IGV) tool was used [377].

5.2.11 Pathway enrichment analysis

A KEGG pathway enrichment analysis was made over the identified PAR-CLIP candidate genes using cluster Profiler package [378].

5.3 Results

5.3.1 DLBCL cell lines are highly sensitive to puromycin

Two DLBCL cell lines, ABC-type, RIVA, and GC-type, SU-DHL-10, were used to elucidate the targetome of *miR-155* using the PAR-CLIP assay. We used lentiviral vectors to systematically modulate the levels of *miR-155* in these cells in order to improve the reliability of the data obtained. Previous to the infection, the intrinsic puromycin sensitivity of these cell lines was measured in order to ensure a subsequent correct puromycin selection of the infected cells.

Results of MTT assays showed that these cell lines are highly sensitive to puromycin treatments. Even with very low doses (0.3 ng/µL), more than 90% of the cells seem not to be metabolically active. After 6 days of exposure to the drug there are not alive cells even at the lowest concentration (Figure 5.3). Therefore, we established a standard treatment of 1ng/µL of puromycin for 6 days.
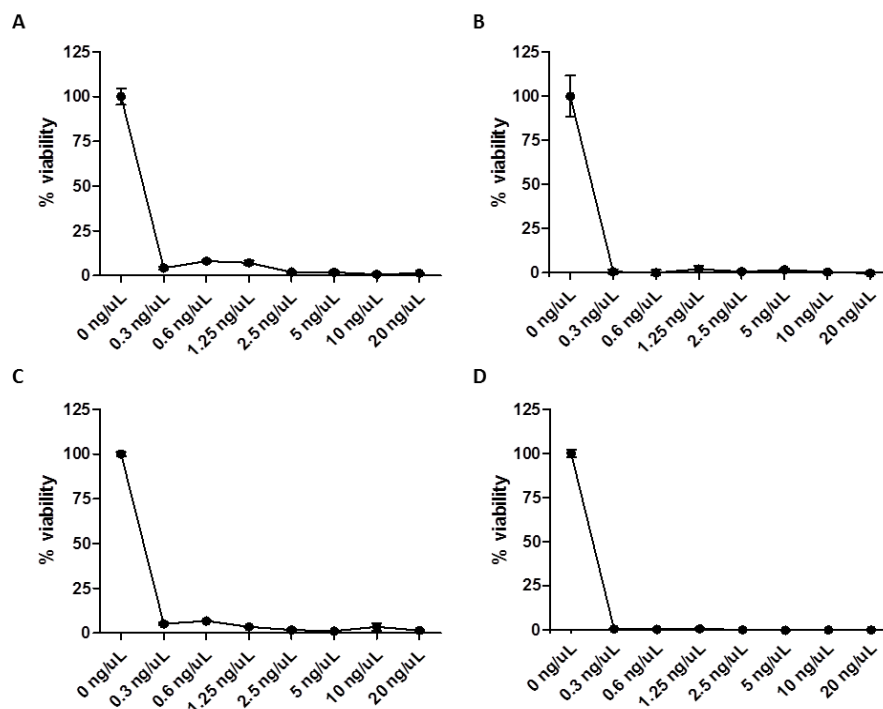
Figure 5.3: Puromycin sensitivity results for DLBCL cell lines. *Figure shows puromycin killing curves for SU-DHL-10 with 3 days (A) and 6 days (B) of exposition and RIVA cell lines with 3 days (C) and 6 days (D) of exposition.*

5.3.2 Lentiviral infection with a MOI 15 promotes an optimum infection of DLBCL cell lines

We calculated the optimum multiplicity of infection (MOI) by titration for the DLBCL cell lines. Cells were infected under 4 different MOI conditions (i.e. MOI 15, 30, 45 and 60) with three lentiviral vectors in order to upregulate the expression of *miR-155* (Over-miR155), inhibit *miR-155* expression (Anti-miR-155), and with a negative control for the infection (Scramble sequence). The efficacy of the infection was evaluated by analysing the reporter gene GFP with flow cytometry. The green intensity was divided into low, medium and high intensity fluorescence in order to see whether a higher green intensity could correspond with a higher number of the vector copies per cell and therefore stronger modulation of the *miR-155*. We observed that increasing the MOI didn't have a significant impact on the number of GFP-positive cells (Figure 5.4). This suggests that with a MOI of 15 the infection is already optimal and that increasing the MOI would not be beneficial. To confirm this, we analysed the expression levels of *miR-155* in SU-DHL-10 cells infected with MOI 15 or MOI 60. As shown in Figure 5.5 we didn't notice a significant difference in the level of *miR-155* modulation between these two conditions.

Cell line	Lentivirus	MOI	Green and Alive Cells (%)				Dead cells (%)
			Low	Medium	High	Total green cells	
RIVA	Scramble	MOI 15	7	15	4	34.9	16.6
		MOI 30	7.2	13.6	3.2	33.9	9.9
		MOI 45	7.5	13.9	4.4	35	16.2
		MOI 60	7.4	15.6	5.2	38.5	10.1
	OVER-miR155	MOI 15	10.8	22.2	13.1	61.7	11.1
		MOI 30	10.5	23.3	9.4	57.9	34.4
		MOI 45	9.4	23	18.8	67.2	13.8
		MOI 60	12.4	22.9	9.6	61.3	8.7
	ANTI-miR155	MOI 15	42.6	16.2	2.1	78.1	10.5
		MOI 30	40.4	20.2	3.9	82.4	9.9
		MOI 45	41.7	19.2	3.7	82.3	12.2
		MOI 60	31.2	28.5	10.3	88.8	15.1
SU-DHL-10	SCR	MOI 15	4.9	1.3	0.5	13.3	6.3
		MOI 30	4.4	1.1	0.3	11.7	4
		MOI 45	4.3	1	0.3	11.9	4.2
		MOI 60	4.5	1.1	0.4	12.9	4.7
	OVER-155	MOI 15	12.6	3.9	3	30.9	4.4
		MOI 30	12.6	3.5	2.7	30.3	3.6
		MOI 45	12.6	2.9	2.8	29.7	4.2
		MOI 60	10.9	3.1	2.4	26.9	3.9
	ANTI-155	MOI 15	17.5	6.2	1.8	47.4	3.5
		MOI 30	19.2	6.7	2	50.6	4.3
		MOI 45	18.7	6.6	1.8	49.6	3.7
		MOI 60	17.4	6.2	1.7	47.5	4.4

Figure 5.4: Results of GFP expression as an indicator of the lentiviral infection yield. (A) Example of the stratification of GFP+ cells into three categories: low expression, medium expression, high expression. (B) Flow cytometry data of the amount of cells in each category per condition together with the amount of total green cells and dead cells.
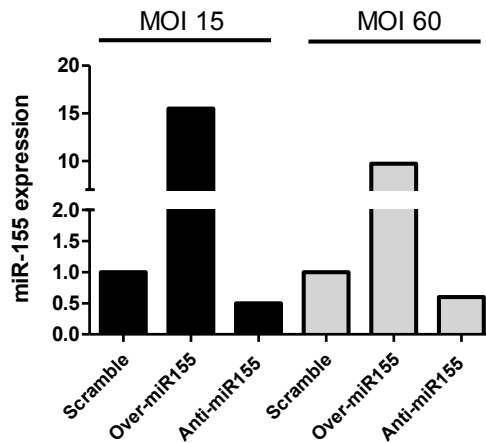
Figure 5.5: RT-qPCR data. *MiR-155 expression in SU-DHL-10 cell line infected with the lentiviral vectors with two different MOIs.*

5.3.3 *MiR-155* expression monitoring showed a stable upregulation of *miR-155* in the infected DLBCL cell models

DLBCL cell lines were stably transfected in biological triplicate with lentiviral vectors (MOI 15) to modulate *miR-155* expression. These vectors encode *miR-155* sequence (Over-miR-155), a scrambled control sequence (Scramble), or an inhibitor of *miR-155* (Anti-miR-155). Cell were stably selected with puromycin and maintained in culture. *miR-155* expression was measured in these cells and we observed that up-regulation of *miR-155* was maintained in culture up to 4 weeks post-infection. In contrast, however, inhibition of *miR-155* using the miRZip knockdown vector was not maintained for more than 2 weeks (Figure 5.6). Therefore, for subsequent PAR-CLIP experiments the anti-miR-155 condition was discarded and only the scramble condition was used as a control.
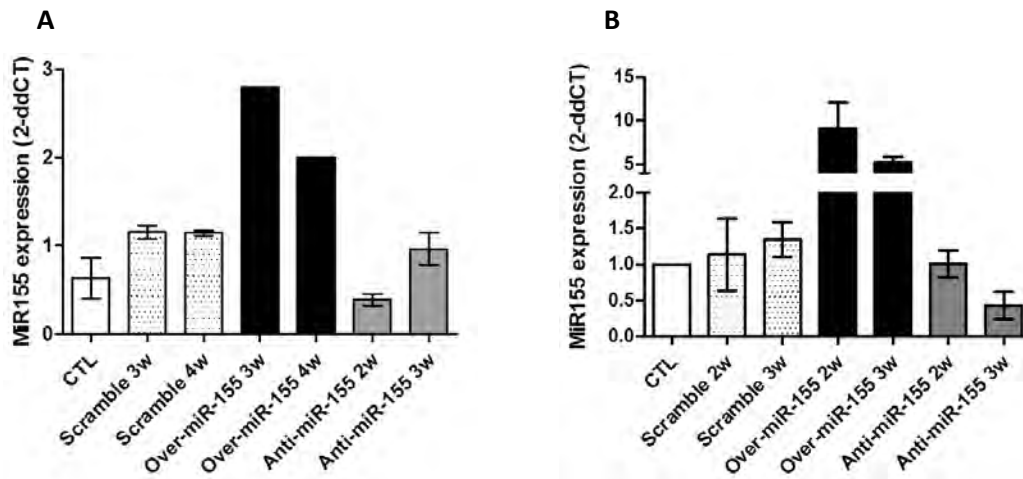
Figure 5.6: RT-qPCR results for miR-155 expression. *Figure shows miR-155 expression at different time points after infection, from two weeks (2w) until 4 weeks (4w) in RIVA (A) and SU-DHL-10 (B) cell lines. Results showed a stable upregulation of miR-155 up to 3-4 weeks post-infection but a variable effect of the miR-155 downregulation.*

6.3.4 DLBCL cell lines express high levels of endogenous Ago2

A western-blot analysis was made in order to compare the expression levels of endogenous Ago2 protein in the DLBCL cell lines we were using. Results showed that Ago2 is highly expressed in DLBCL cell lines. In HeK293 FLAG-HA-Ago2 induced with doxycycline an additional slightly higher band can be observed corresponding to the exogenous FLAG-HA-tagged Ago2 protein. RIVA cell line seems to strongly express Ago2. This expression is similar to the total Ago2 expression in the induced HeK293 FLAG-HA-Ago2 cells. However, Ago2 expression in SU-DHL-10 is lower, approximately half the expression in RIVA (Figure 5.7).
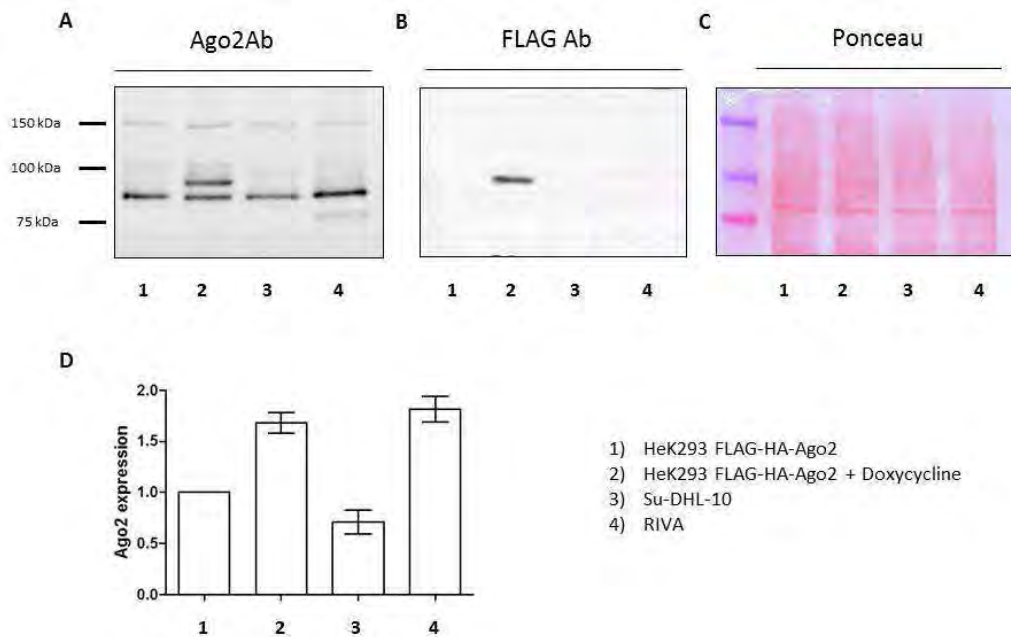
Figure 5.7: Ago2 expression in DLBCL cell lines and HeK293 FLAG-HA-Ago2: *Ago2 expression was analysed with Western-blot in DLBCL cell lines compared to HeK293 FLAG-HA-Ago2 cell line with and without inducing FLAG-HA-tagged Ago2 protein expression. Figure shows results of a western-blot for Ago2 (A), a western-blot for FLAG tag (B), a Ponceau staining of the membrane verifying the same amount of protein loaded in each lane (C) and the quantification of the relative Ago2 expression analysed by western-blot in these cell lines from three different experiments (D).*

6.3.5 4-thiouridine (4-SU) is not toxic to DLBCL cell lines

In order to ensure that the 4-SU was not toxic to the DLBCL cell models we tested increasing doses of 4-SU and evaluated their viability and proliferation. We observed a decrease in cellular proliferation when applying higher doses of 4-SU. Toxicity was not detected at concentrations below 200µM. At 200µM cells do not seem to be replicative and, at concentrations above 200µM, we could detect a certain grade of toxicity, since the number of alive cells was lower than the amount of cells seeded (Figure 5.8). Based on these data we performed all further experiments on B-cells at 100 µM 4-SU. We observed that the cells remained replicative at that concentration. Of note, 100µM 4-SU was the concentration used for previous PAR-CLIP-based studies [366].
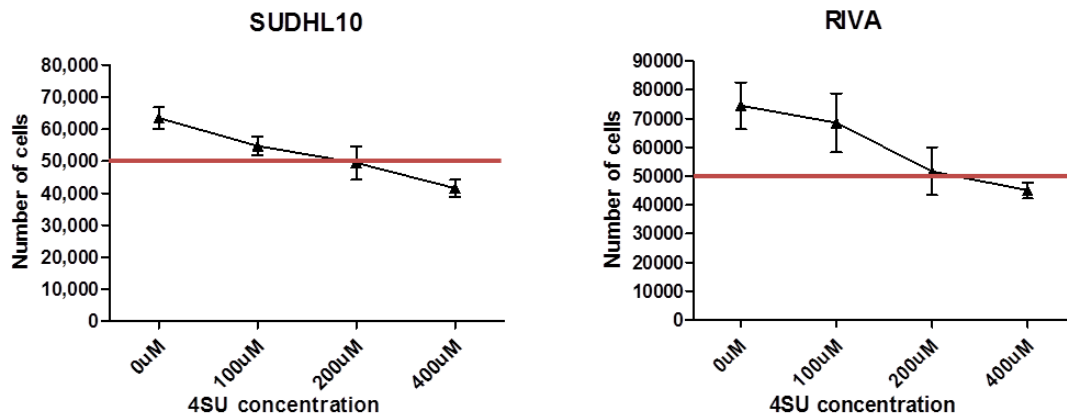
Figure 5.8: 4-SU toxicity assays on DLBCL cell lines. *SU-DHL-10 and RIVA cells were exposed to increasing concentrations of 4-SU for 18 hours. The red line shows the amount of the initially seeded cells (50.000 cells/well). Data shows that, at 200μM cells seem to stop being replicative and for higher concentrations they seem to experiment some toxicity as the number of alive cells falls below 50.000 cells/well.*

5.3.6 DLBCL-derived cells effectively incorporate 4-SU into the nascent RNAs at a similar rate as HEK293 cell line

We wanted to determine the number of DLBCL cells required to successfully perform a PAR-CLIP assay, bearing in mind that the reference cell line HEK293, may not have the same quantities of RNA or ability to incorporate 4-SU as the DLBCL cell lines.

Our data show that we were able to isolate 15 µg of total RNA per million of HEK293 cells whereas in DLBCL cell lines the amount of isolated RNA was 5 µg of total RNA per million cells (Figure 5.9). However, as for the PAR-CLIP assay, the Ago2-bound RNA is that synthetized during the 4-SU incorporation we also looked at the RNA synthesis rate for these cells. Results show that all the cells are metabolically active during the 4-SU exposure and that, it was incorporated to the nascent RNA molecules. The RNA synthesis rate for DLBCL cells is 1,2-1,5 less than for HEK293 cell line (Figure 5.10). So, for the same amount of Ago2 protein, the amount of RNA capable of being cross-linked to Ago2 in HEK293 is approximately 4 times higher than that of DLBCL cells.

5. Elucidating the miRNA-targetome of DLBCL
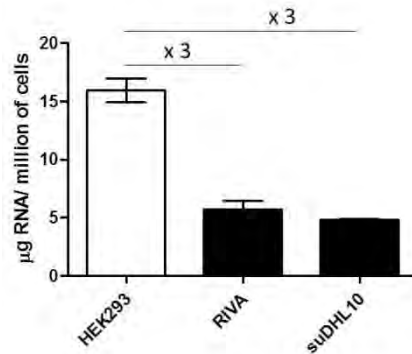
Figure 5.9: Amount of RNA in HEK293 and DLBCL cell lines.
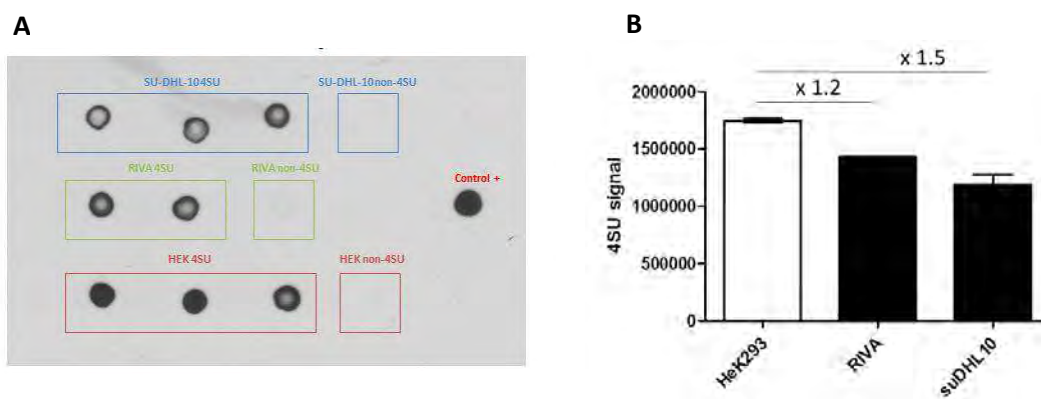
Figure 5.10: Dot-blot results. *Figure A shows the Dot-blot assay for 4-SU incorporation into nascent RNA in HEK293 and DLBCL cell lines and figure B the quantification of the signal.*

5.3.7 Ago2 immunoprecipitation (IP) analysis by Western-blot verified a correct enrichment of Ago2 protein after the IP step in PAR-CLIP protocol.

An immunoprecipitation (IP) of Ago2 protein from SU-DHL-10 cells over-expressing *miR-155* was made following the PAR-CLIP protocol in order to verify the enrichment of Ago2 protein in the immunoprecipitated fraction. The presence of Ago2 was analysed using Western-blot in the total lysate along with the immunoprecipitated fraction and the unbound fraction. Results showed a clear enrichment of the band corresponding to Ago2 in the IP fraction. Also, the band corresponding to Ago2 in the negative IP fraction was clearly less intense than the band from the total lysate and the band from the unbound fraction of the IP with the isotype control (Figure 5.11). The isotype control is useful to differentiate non-specific background signal from specific antibody signal as

146

they have no relevant specificity to a target antigen. The positive fraction of the IP with the isotype control gave a very intense background signal due to non-specific interactions. It should be noted that although the anti-Ago2 antibody also appears to recognize a protein of 150kDa (Figure 5.11, lane 4), for the PAR-CLIP assay we isolate only proteins between 75 and 100 kDa in size (see in PAR-CLIP modified protocol, section 5.2.7).

1. HEK293 FLAG-HA-Ago2 + Doxycycline
2. Total lysate suDHL10 Over-miR-155
3. Unbound IP fraction Anti-Ago2 antibody suDHL10 Over-miR-155
4. IP positive fraction Anti-Ago2 antibody suDHL10 Over-miR-155
5. Unbound IP fraction Anti- IgG2A antibody suDHL10 Over-miR-155 (Control)
6. IP positive fraction Anti-IgG2A antibody suDHL10 Over-miR-155 (Control)
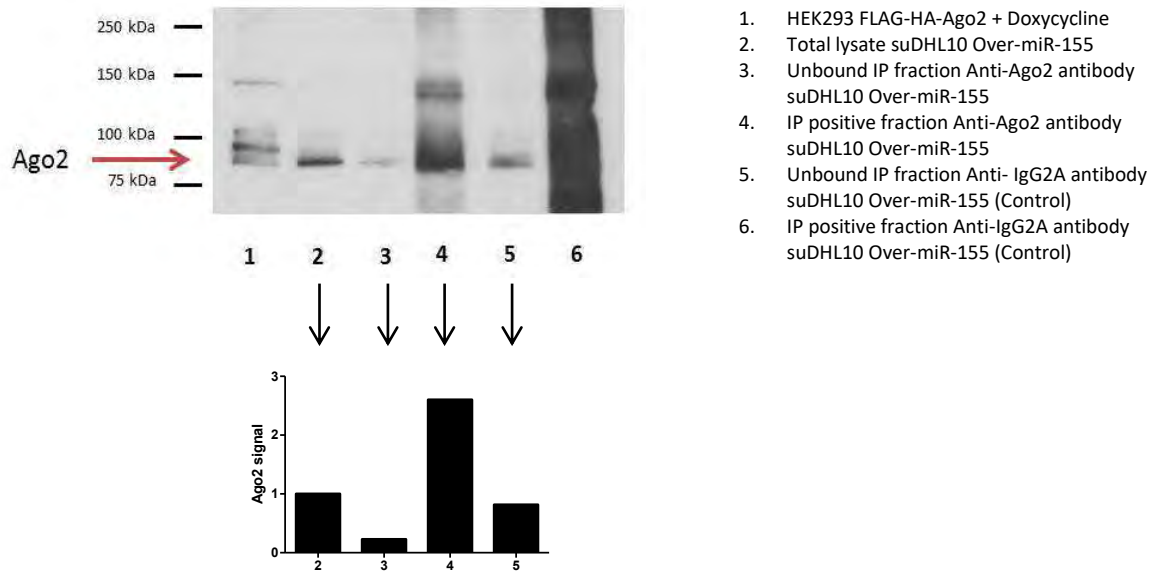
Figure 5.11: Western-blot analysis of the IP experiment against Ago2 protein. Results showed a correct IP of Ago2 protein as it can be notice when comparing Ago2 in the total lysate with the negative and the positive fractions of the IP (conditions 2, 3 and 4 respectively). Conditions 5 and 6 correspond to the negative control of the IP (negative and positive fraction). Condition number 1 is the total lysate of HEK293 FLAG-HA-Ago2 cell line treated with doxycycline, so a higher band corresponding to the FLAG-HA-tagged Ago2 protein can be also observed.

5.3.8 Gene expression profiling of DLBCL cell lines with the *miR-155* modulation using microarrays

A transcriptome analysis of the DLBCL cell models employed for the PAR-CLIP assay was performed in order to combine these data with the targetome analysis data. For each cell line, the 20 most down-regulated genes (over-miR-155 vs. Scramble). The gene expression microarray data showed a total of three and six genes significantly differentially expressed in RIVA and SU-DHL-10 cell lines respectively (Table 5.1).

5. Elucidating the miRNA-targetome of DLBCL

Gene	Cell line	log FC	p value	Adjusted p value
OR2B6	RIVA	-2.433	7.52E-07	0.010
OR2W6P	RIVA	-2.413	9.95E-08	0.003
HIST1H1T	RIVA	-1.632	6.75E-06	0.074
RP1-86C11.7	RIVA	-1.599	3.26E-05	0.194
FCGR2B	RIVA	-1.366	1.99E-04	0.379
RNU2-62P	RIVA	-1.263	1.09E-04	0.315
HIST1H2BO	RIVA	-1.093	1.11E-04	0.315
FCGR2C	RIVA	-1.016	5.13E-04	0.542
HIST1H2APS3	RIVA	-1.014	2.22E-04	0.379
RP5-951N9.1	RIVA	-0.992	2.39E-03	0.684
RP11-616L12.4	RIVA	-0.953	1.35E-03	0.635
BCAS1	RIVA	-0.950	7.32E-04	0.545
IL3RA	RIVA	-0.916	1.76E-03	0.647
TCONS_00027962	RIVA	-0.903	1.75E-03	0.647
RP11-818F20.5	RIVA	-0.895	3.83E-03	0.699
OR2B2	RIVA	-0.888	7.60E-04	0.553
SNORA51	RIVA	-0.877	1.58E-03	0.635
S100A5	RIVA	-0.865	1.02E-03	0.598
LPCAT1	RIVA	-0.863	2.05E-02	0.715
LINC01628	RIVA	-0.853	1.30E-03	0.635
G2E3	SU-DHL-10	-1.049	3.67E-06	0.012
FAM155A	SU-DHL-10	-0.970	2.86E-05	0.052
PHACTR2	SU-DHL-10	-0.942	2.66E-05	0.051
C9orf72	SU-DHL-10	-0.936	3.88E-05	0.062
MIR548I2	SU-DHL-10	-0.934	2.26E-05	0.048
OSTN	SU-DHL-10	-0.875	6.55E-05	0.087
LINC01344	SU-DHL-10	-0.860	9.77E-05	0.114
ANKRD50	SU-DHL-10	-0.838	2.31E-04	0.183
AF129408.17	SU-DHL-10	-0.825	1.94E-04	0.170
RP3-468B3.4	SU-DHL-10	-0.808	3.38E-04	0.202
RP4-635A23.6	SU-DHL-10	-0.796	3.11E-04	0.198
RP5-855D21.2	SU-DHL-10	-0.789	4.51E-04	0.221
LY86-AS1	SU-DHL-10	-0.778	3.48E-04	0.203
RNU6-865P	SU-DHL-10	-0.777	4.52E-04	0.221
RNU7-45P	SU-DHL-10	-0.754	5.36E-04	0.244
OR5M10	SU-DHL-10	-0.748	7.25E-04	0.273
CTD-2124B8.2	SU-DHL-10	-0.746	5.91E-04	0.256
MIR5690	SU-DHL-10	-0.741	1.31E-03	0.353
TCONS_l2_00023407	SU-DHL-10	-0.729	1.09E-03	0.326
RP11-493L12.7	SU-DHL-10	-0.725	1.08E-03	0.326

***Table 5.1: Differentially expressed genes according to microarray data under miR-155 modulation in DLBCL cell lines.** The table shows the 20 genes downregulated upon*

148

upregulation of miR-155 showing the largest fold-change values together with the p value and the adjusted p value for each gene. The significantly differentially expressed genes are highlighted in bold letters.

5.3.9 PAR-CLIP sequencing results

5.3.8.1 Sequencing results

In general terms, we obtained more reads (>20 nt) in the RIVA cell samples than SU-DHL-10 cells. Collapsing the data to remove replicated reads demonstrated that less than 10% of the reads were unique. An average of 21 % (range 18-23%) of these reads were aligned to the human genome (hg19/GRCh37). The processed PAR-CLIP reads showed to cover an average of 0.22% of the genome (range 0.22-0.27%) (Table 5.2).

Samples	Generated reads (> 20 nt)	Reads after duplicate removal	Aligned reads after deduplication	Breadth of coverage (bp) and % of the genome covered
RIVA Over-miR-155	46282662	1649822 (3.56%)	389876 (23%)	6177892 (0.2%)
RIVA Scramble	56963089	1812782 (3.18%)	377963 (20%)	6136782 (0.2%)
SU-DHL-10 Over-miR-155	27199700	2396222 (8.8%)	440186 (18%)	6985702 (0.22%)
SU-DHL-10 Scramble	34179327	2274925 (6.65%)	538491 (23%)	8481297 (0.27%)

Table 5.2: Number of PAR-CLIP sequencing reads per condition. The table shows the number of reads as a result of the PAR-CLIP sequencing for each of the conditions. The first column shows the number of generated reads after removing those shorter than 20 nucleotides. The second column shows the number of reads after removing the PCR duplicates in each condition and the third column the number of these unique reads successfully aligning against the human genome (hg19/GRCh37). The percentages are in relationship to the data in the previous column. In the last column the total length covered (bp) with the processed data as well as the percentage of the genome covered is shown for each sample.

5.3.8.2 Cluster calling

As only reads with T-to-C changes (resulting from 4-SU incorporation) are indicative of properly being associated with the cross-linked Ago2, we used specialised algorithms, wavClusteR [373] and BMix [374] that only consider reads with that specific change. In parallel, dCLIP tool was employed which is not specific for PAR-CLIP data [375]. Each software for CLIP data analysis identified clusters, which are overlapping reads aligning in the same locus, from the aligned reads using its own parameters. wavCluster and BMix are pipelines specific for PAR-CLIP data analysis and it considers PAR-CLIP mutations (each programme uses its own parameter for filtering out T-to-C mutations not specific from PAR-CLIP) for the cluster calling [373, 374]. In contrast, dCLIP tool can be applied for the analysis of different CLIP data and allows differential analysis between two conditions. This tool is unique in finding differences in binding strengths rather than just the binary event of whether a binding site is shared or not [375]. The total number of clusters identified for each programme is shown in table 5.3. The genomic location of the identified clusters is represented in Figure 5.13.

As can be seen from figure 5.12, all of the samples were enriched for the T-to-C change when compared to all the other possible variants found in the sequencing data.
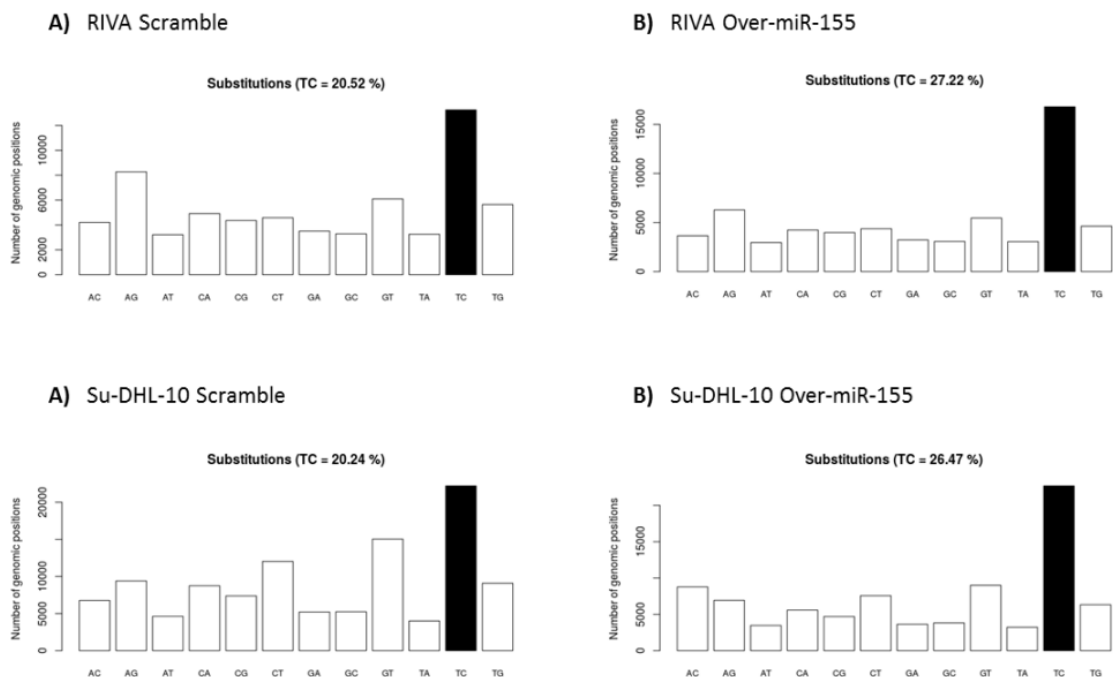
Figure 5.12: Number of genomic positions showing any of the possible conversions.

Sample	Identified Clusters
RIVA Over-miR-155 (wavClusteR)	8772
RIVA Scramble (wavClusteR)	6970
RIVA Over-miR-155 (BMix)	986
RIVA Scramble (BMix)	291
RIVA (dCLIP)	3144
SU-DHL-10 Over-miR-155 (wavClusteR)	12499
SU-DHL-10 Scramble (wavClusteR)	12899
SU-DHL-10 Over-miR-155 (BMix)	1290
SU-DHL-10 Scramble (BMix)	1012
SU-DHL-10 (dCLIP)	4743

Table 5.3: Total number of identified clusters per condition: *Results show the amount of identified clusters with three different software of analysis (wavClusteR, BMix and dCLIP). dCLIP tool compares the difference between (i.e over) experimental and the control (i.e. scramble) cluster analysis.*
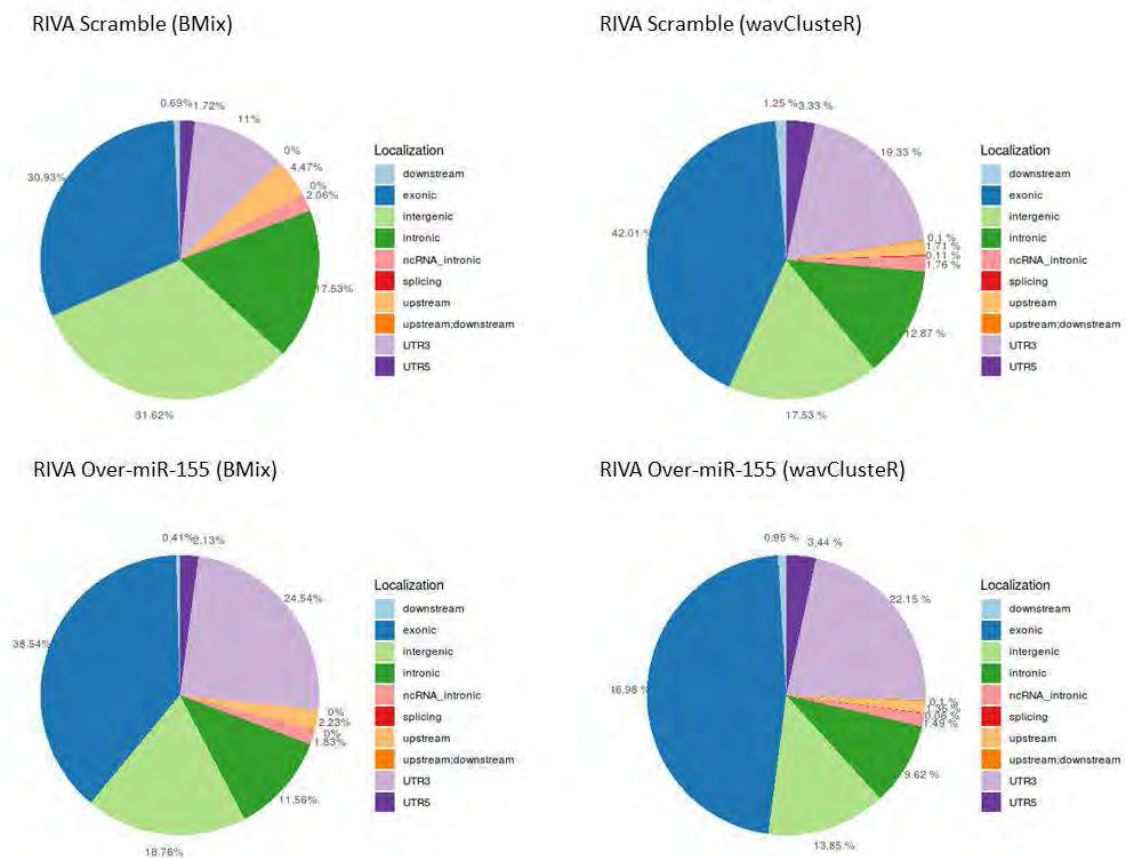
5. Elucidating the miRNA-targetome of DLBCL
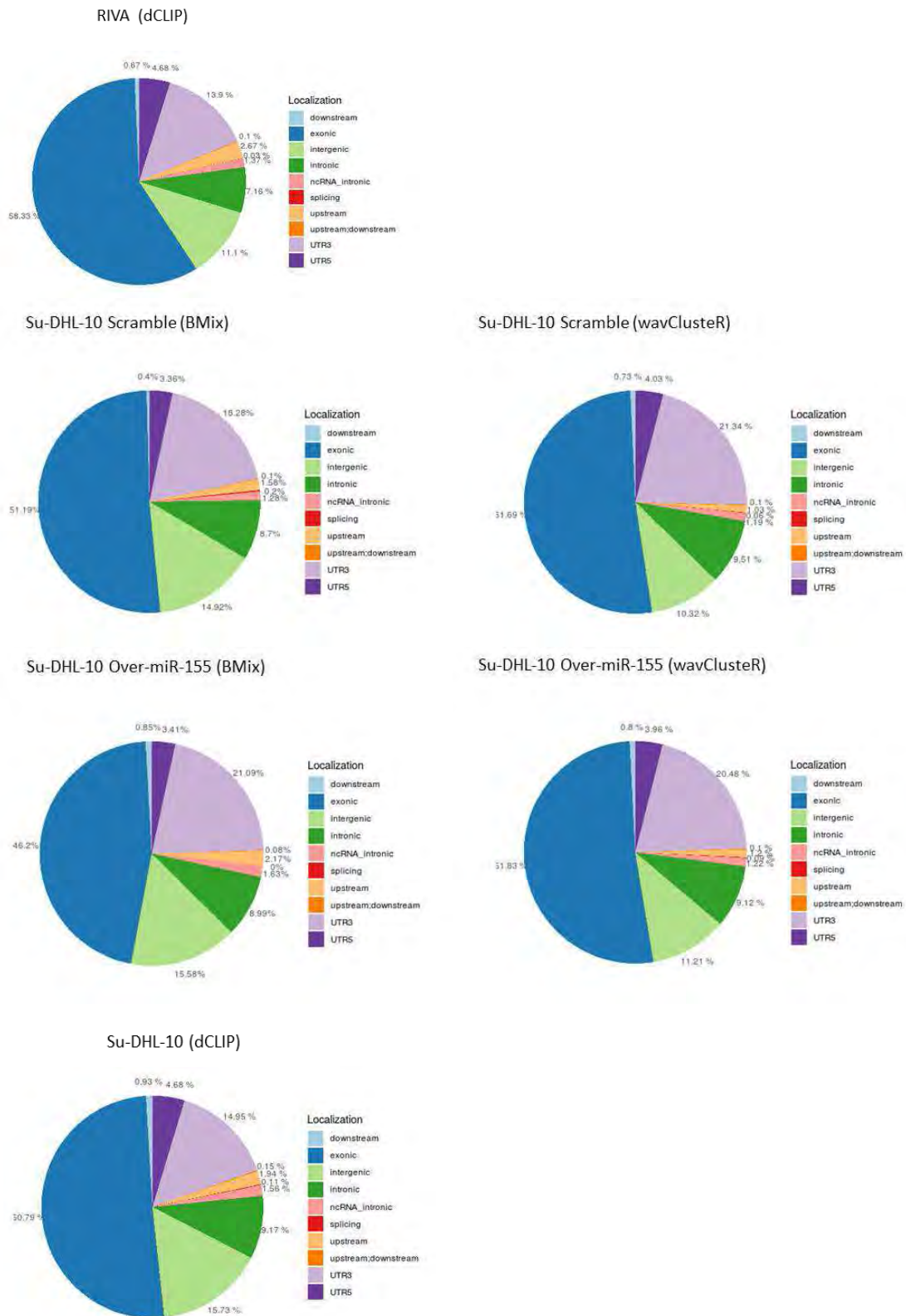
Figure 5.13: Genomic location of the identified clusters. *The figure shows the genomic location where the identified clusters align for each condition and each cell line using all the three softwares for the analysis. Only one graph appears for dCLIP analysis per cell line as it shows the differentially identified clusters between the over-miR-155 and the scramble condition.*

We found that wavClusteR identifies many more clusters than either BMix or dCLIP. By any of these algorithms, the SU-DHL10 cell line samples gave a higher number of clusters than the RIVA cell line samples.

Next, we analysed the genomic location of clusters and observed that the majority of them (average of 46.85%, range 30.93-58.33%) were located in exonic regions of protein coding genes, followed by 3'UTR regions (average of 18.7%, range 11-24.54%), intergenic regions (average of 16%, range 10.32-31.62%), intronic regions (average of 10.42%, range 7.16-17.53%), and 5'UTR regions (average of 3.47%, range 1.72-4.68%). A similar distribution (50% of the clusters identified in exonic regions) was observed in the original PAR-CLIP experiments [296, 326].

Then, we compared the identity of the identified genes using the three different pipelines. As can be seen in Figure 5.14 there was a high degree of overlap in the cluster calling. It can also be observed how wavClusteR is a more sensitive tool whereas BMix uses more stringent parameters for the cluster identification.

A) RIVA

- wavClusteR n = 3520
- BMix n = 537
- dCLIP n = 1415

A wavClusteR and dCLIP shared genes n = 1157
B wavClusteR and BMix shared genes n = 536
C dCLIP and BMix shared genes n = 308

B) SU-DHL-10

- wavClusteR n = 3752
- BMix n = 644
- dCLIP n = 1697

A wavClusteR and dCLIP shared genes n = 1502
B wavClusteR and BMix shared genes n = 642
C dCLIP and BMix shared genes n = 460
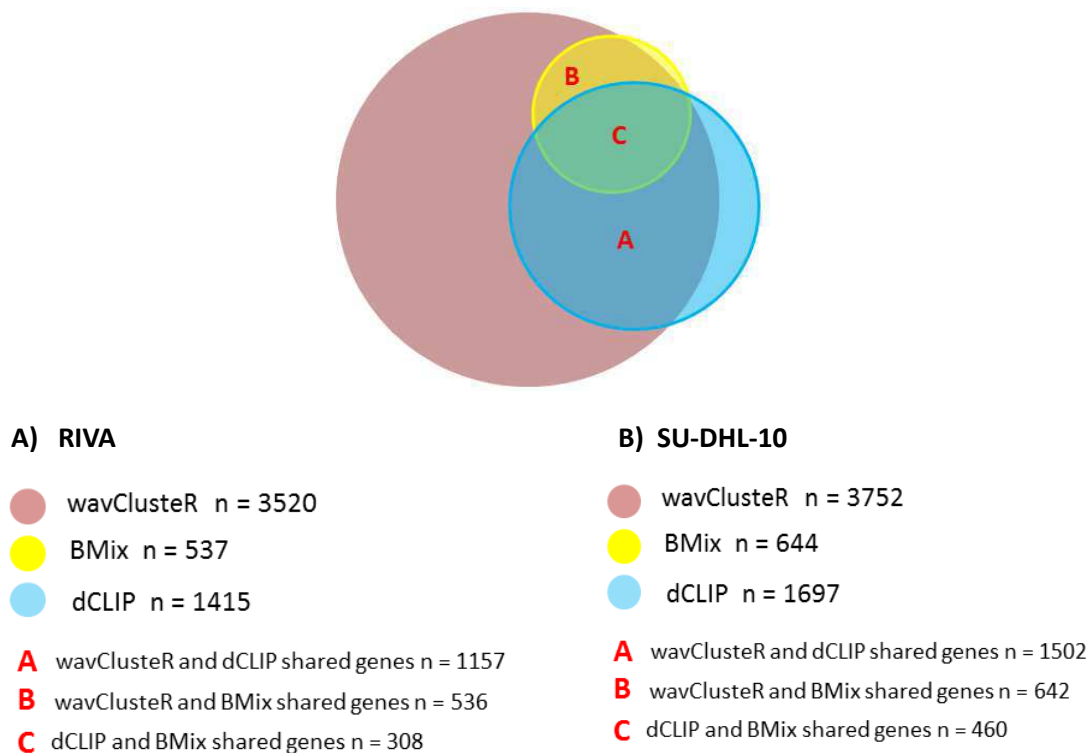
Figure 5.14: Number of identified genes with PAR-CLIP using three different softwares for cluster analysis. Three different softwares (waveClusteR, dCLIP and BMix) were employed in order to identify miRNA targeted genes in RIVA (A) and SU-DHL-10 (B) cell lines after PAR-CLIP

performance. The figure shows the total number of genes identified in over-miR-155 condition using wavClusteR and BMix and the differential cluster identification between over-miR-155 and scramble using dCLIP tool. Only clusters located in protein coding genes are considered. The figure also shows the overlapping between this gene identification.

The full data set for the PAR-CLIP experiments can be accessed from the following links:

A) Data for RIVA: https://maitenat.shinyapps.io/DLBCL_PARCLIP_RIVA_300M/
B) Data for SU-DHL-10: https://maitenat.shinyapps.io/DLBCL_PARCLIP_SU10_30050M/

5.3.8.3 Candidate gene selection

Considering the data of all the three tools combined (see links in previous section), we end up with a total of 4,376 genes identified in the RIVA cell line and 4,725 in the SU-DHL-10 cell line. We removed clusters that were located outside of protein-encoding genes (i.e. 5'UTR, exonic region and 3'UTR), resulting in 3,778 genes for RIVA cell line and 3,946 for SU-DHL-10 cell line.

We found that 82% of these genes were differentially identified between the experimental condition (over-miR-155) and the control condition (scramble) for RIVA cells, and 78% for SU-DHL-10 cells. Amongst these genes 63% in RIVA and 50% in SU-DHL-10 cells, were identified exclusively in cells that over-expressed *miR-155* but not the scramble-sequence controls. In case of discrepancy between BMIX and wavClusteR tools, data from BMix was prioritized as this tool is more specific than wavClusteR (Figure 5.14). The rest of the differentially identified genes were grouped as genes showing "minor evidence of differential identification" meaning that they were identified by dCLIP as differential even if they were also identified in the control dataset. 18% of the genes in RIVA and 21% in SU-DHL-10 were not differentially identified between the experimental (over-miR-155) and the control condition (scramble) (Figure 5.15) and were excluded from further analysis.

Ontology analysis of the differentially identified genes (n=3114 and n=3098 in RIVA and SU-DHL-10 respectively) were found to be significantly enriched for B-cell receptor signalling and cell cycle in both DLBCL subtypes. But, interestingly, only the genes identified in the ABC-DLBCL subtype (RIVA cell line) showed an enrichment for genes involved in FOXO (adjusted p value = 9.51 x 10^{-4}) and AMPK (adjusted p value = 3.32 x

10^{-2}) signalling pathways suggesting that genes involved in these pathways might be directly target by *miR-155* specifically in ABC-DLBCL subtype.

Even though only 20 (0.65%) and 14 (0.45%) of the differentially identified genes in RIVA (n=3085) and SU-DHL-10 (n=3098) respectively, had a complete seed match for *miR-155* binding, much higher numbers of these target genes had been experimentally validated as bona fide *miR-155* target genes (268 (8.7%) in RIVA, and 240 (7.75%) in SU-DHL-10) according to miRTarBase [379] (Figure 5.16). Five and two genes respectively presented a seed match and also are validated as *miR-155* targets. The differentially identified genes with a seed match for *miR-155* are listed in Table 5.4. SF1 and MICB genes were identified in both cell lines. Also, examples of cluster identification can be observed in Figure 5.18.
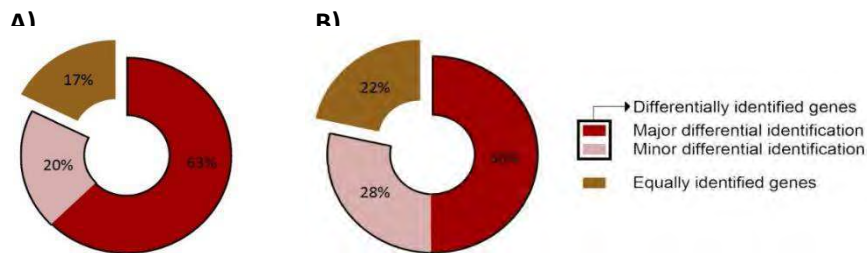
Figure 5.15: Differentially identified genes for RIVA and SU-DHL-10. The figure shows the differentially identified genes between the experimental (overexpression of miR-155) and the control condition (a scramble overexpression) after performing PAR-CLIP in RIVA (A) and SU-DHL-10 (B) cell lines.
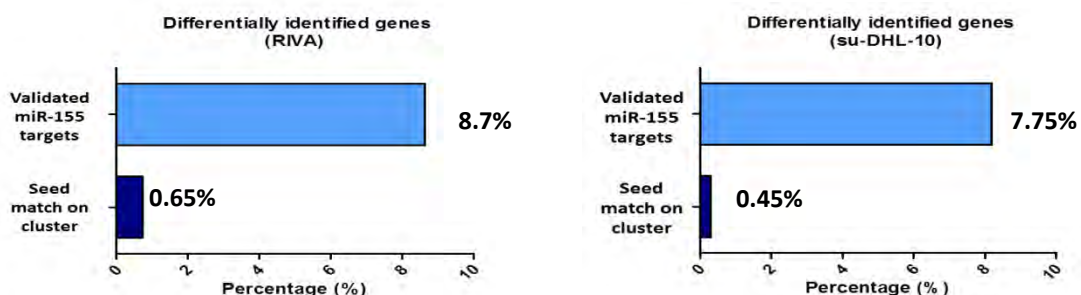
Figure 5.16: Differentially identified genes with clues for being a miR-155 target gene. The figure shows, among the differentially identified genes between the experimental and the control condition, the percentage of the identified genes that are predicted to be miR-155 targets, that have been experimentally validated as miR-155 target genes and the percentage of the genes that showed a seed sequence for the binding of miR-155 in the identified cluster.

5. Elucidating the miRNA-targetome of DLBCL

Gene	Cell line	Cluster calling tool	Location	Cluster lenght (bp)	Seed match	Evidence as *miR-155* target
CD22	SU-DHL-10	dCLIP	exonic, UTR3	282	yes	No evidence
EEF1G	SU-DHL-10	dCLIP	exonic	73	yes	No evidence
HIST1H2BK	SU-DHL-10	dCLIP	exonic	163	yes	No evidence
HMGXB3	SU-DHL-10	dCLIP	exonic	99	yes	No evidence
PIEZO1	SU-DHL-10	dCLIP	exonic, intronic	892	yes	No evidence
PXK	SU-DHL-10	dCLIP	UTR3	49	yes	No evidence
SF1	SU-DHL-10	dCLIP	UTR3, exonic	334	yes	No evidence
E2F2	SU-DHL-10	BMix	UTR3, exonic	130	yes	Validated
NFYC	SU-DHL-10	wavClusteR	intronic, UTR3	92	yes	Validated
MICB	SU-DHL-10	BMix	UTR3	36	yes	No evidence
DAP3	SU-DHL-10	wavClusteR	exonic	39	yes	No evidence
FAM32A	SU-DHL-10	BMix	exonic	53	yes	No evidence
STIP1	SU-DHL-10	BMix	exonic	168	yes	No evidence
UBB	SU-DHL-10	BMix	exonic, UTR3	120	yes	No evidence
CNDP2	RIVA	wavClusteR	exonic	67	yes	Validated
KANSL1	RIVA	wavClusteR	UTR3, exonic	55	yes	Validated
ZFP36	RIVA	wavClusteR	UTR3	37	yes	Validated
ZNF431	RIVA	wavClusteR	UTR3	37	yes	Validated
MEX3C	RIVA	dCLIP	UTR3, exonic	132	yes	Validated
BTAF1	RIVA	wavClusteR	exonic	26	yes	No evidence
CTNND1	RIVA	wavClusteR	UTR3	49	yes	No evidence
EIF2AK1	RIVA	wavClusteR	UTR3	76	yes	No evidence
KIAA0907	RIVA	wavClusteR	exonic	59	yes	No evidence
MCOLN2	RIVA	BMix	UTR3	30	yes	No evidence
MICB	RIVA	BMix	UTR3	91	yes	No evidence
PANK3	RIVA	wavClusteR	exonic	3424	yes	No evidence
RPL6	RIVA	BMix	exonic	35	yes	No evidence
TNRC6A	RIVA	BMix	exonic, UTR3	127	yes	No evidence
CXCR4	RIVA	dCLIP	exonic	74	yes	No evidence
HIST1H2BJ	RIVA	dCLIP	exonic	93	yes	No evidence
NREP	RIVA	dCLIP	UTR3	39	yes	No evidence
OGT	RIVA	dCLIP	exonic	119	yes	No evidence
RPL27A	RIVA	dCLIP	exonic	64	yes	No evidence
SF1	RIVA	dCLIP	exonic	280	yes	No evidence

Table 5.4: Candidate PAR-CLIP genes with a seed match for miR-155 in the identified cluster for RIVA and SU-DHL-10 cell lines. *The table shows the differentially identified genes with PAR-CLIP with a seed match for miR-155 in the identified cluster in the DLBCL cell lines. SF1 and MICB genes were identified in both cell lines.*

In addition, the list of the candidate PAR-CLIP genes (differentially identified genes) as *miR-155* targets was crossed with the differentially expressed genes in these cell lines analysed by microarrays upon *miR-155* modulation. Results led us to identify two candidate genes for RIVA and another two for SU-DHL-10 (Table 5.5).

Gene	Cell line	Cluster calling tool	Location	Cluster length (bp)	Seed match	Evidence as *miR-155* target
G2E3	SU-DHL-10	wavClusteR	3' UTR	21	no	Predicted
OSTN	SU-DHL-10	wavClusteR BMix	3' UTR	47	no	No evidence
SNORA51	RIVA	dCLIP	exonic	19	no	No evidence
HIST1H2BO	RIVA	dCLIP	exonic	78	no	No evidence

Table 5.5: Candidate genes as miR-155 target genes according to microarray and PAR-CLIP data. *Correlation of PAR-CLIP results and transcriptome analysis for RIVA and SU-DHL-10 cell lines with miR-155 modulation led us to prioritize two candidate genes as miR-155 targets in each cell line.*

A comparison of the degree of congruency between PAR-CLIP and bioinformatic prediction for miRNA target identification was carried out by comparing PAR-CLIP data, against predicted *miR-155* target genes using the TargetScan tool along with experimentally validated targets from miRTarBase. Results showed that PAR-CLIP was able to identify one third of the validated target genes (321/993) whereas TargetScan only one fifth (201/993). Also, the number of validated genes (n=993) is nearly twice as much (x 1.8) as the number of predicted genes with TargetScan (n=556) showing that a high number of validated *miR-155* target genes are not predicted with this tool (Figure 5.17).
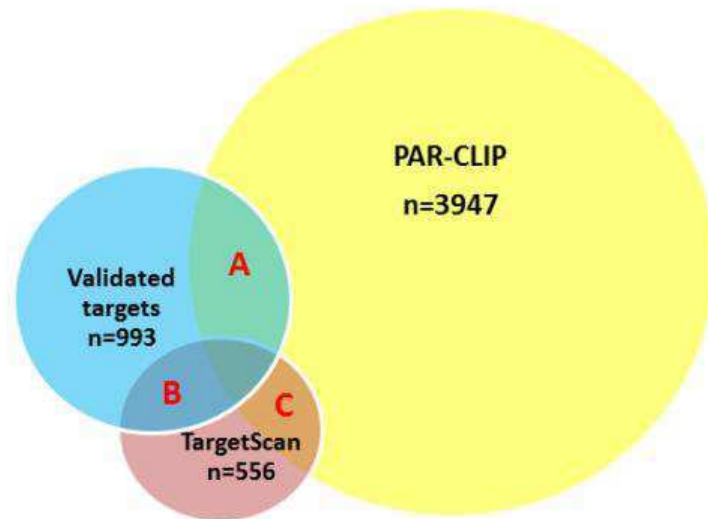
A   Validated genes as *miR-155* targets identified with PAR-CLIP n=321
B   Validated genes as *miR-155* targets predicted with TargetScan n = 201
C   Predicted *miR-155* target genes and identified with PAR-CLIP n = 124

Figure 5.17: Overlapping between the miR-155 predicted target genes and the identified genes with PAR-CLIP as well as the overlapping of both approaches with the experimentally validated miR-155 targets. *The PAR-CLIP data corresponds to the total amount of genes identified containing clusters exclusively in protein coding genes (5'UTR, exons or 3'UTR regions) from SU-DHL-10 over-miR-155 condition. Predicted miR-155 target genes were identified using the default parameters with TargetScan v.7.2. The validated miR-155 target genes were obtained from miRTarBase v.7.0.*

A) UBB: chr17:16285926-16285982, 3'UTR, + strand
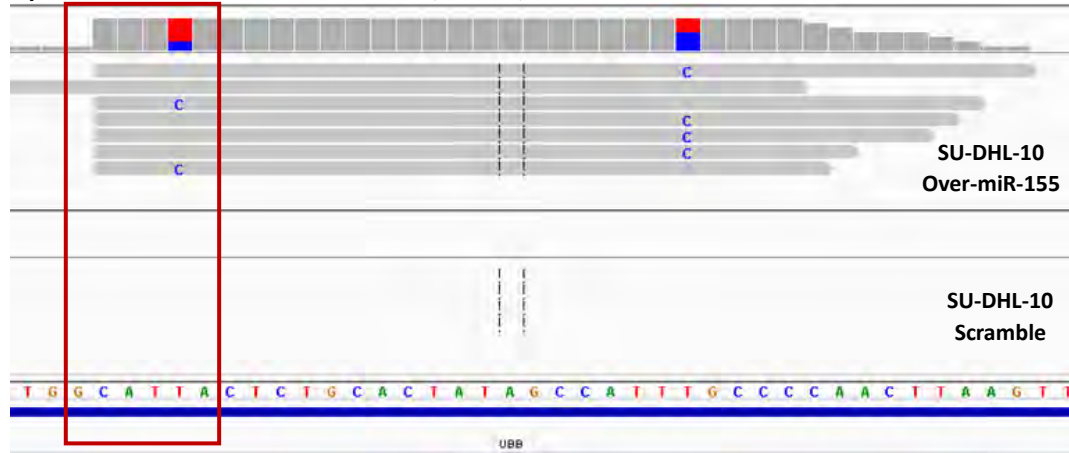
B) E2F2: chr1:23832932-23832955, 3'UTR, - strand
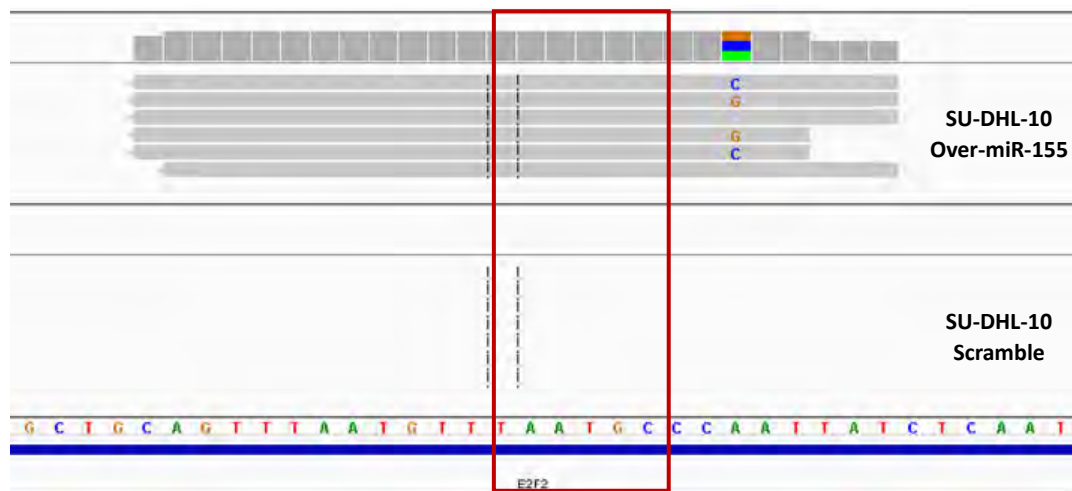
C) MICB: chr6:31478811-31478847, 3'UTR, + strand
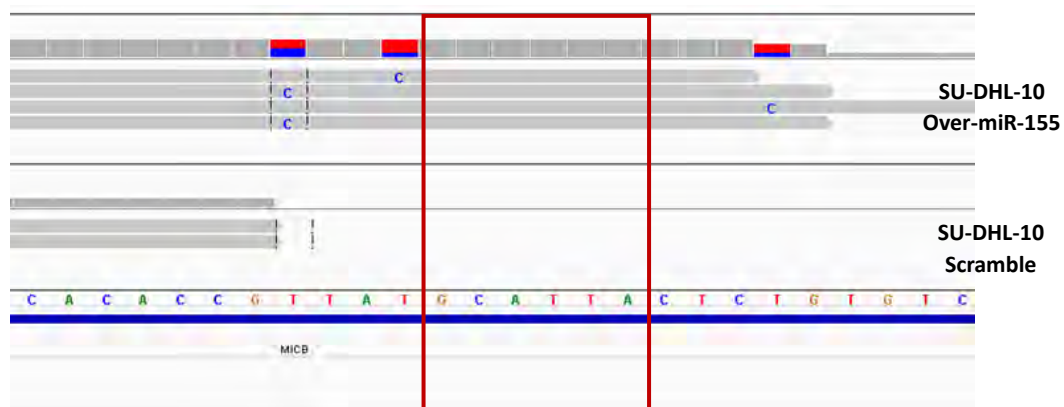
Figure 5.18: Differentially identified clusters in SU-DHL-10 cell line. *The figure shows some examples of the identified clusters for UBB (A), E2F2 (B) and MICB (C) genes exclusively in the over-miR-155 condition in SU-DHL-10 cell line. All the three clusters contain a seed match for miR-155 which is highlighted with a red square in the figure. The PAR-CLIP specific T-to-C mutation can be observed in all the three examples (A-to-G for E2F2 (B) as the gene is located in the negative strand).*

5.4 Discussion

Although previous work by ourselves, and others, have identified many microRNAs that are aberrantly expressed in lymphoma, the functional consequence of their dysfunctional expression to the pathogenesis of lymphoma, or indeed any cancer, is only very poorly understood. The primary reason for this is a paucity of knowledge about which genes are actually targeted by specific microRNAs and which of these genes are functionally important in specific cellular settings. With very few functionally annotated exceptions, current approaches to this problem rely upon the use of the many predictive computational algorithms available (see section 1.4.6). However, these algorithms typically predict hundreds or even thousands of target genes for each miRNA and in reality perform very poorly. Indeed when the most widely used algorithms were tested against experimentally validated miRNA-target gene interactions sensitivity ranged from just 1.3% to 48.8% and even when all five algorithms were used in union only 72% of experimentally validated miRNA-target gene interactions were predicted [380]. Additionally, it has been demonstrated that some experimentally proven targets of microRNAs (e.g. *KRAS* and *HRAS* targeting by let-7 [381], or *E2F2* and *MYC* targeting by *miR-24* [189]) are not predicted targets of these algorithms at all.

In light of these facts it is clear that to fully understand the role that aberrantly-expressed miRNAs play in lymphoma, we must firstly identify which genes are actually regulated by these microRNAs. One strategy to address this question is the use of gene expression arrays to elucidate which genes change in response to permutation of individual microRNAs [288]. However, this approach is limited, as miRNAs are well known to target protein levels without changing transcript levels. Another option is to analyze the proteome through mass spectrometry-based techniques (SILAC, super-SILAC) [382]. However, these techniques do not allow the observer to distinguish between direct and an indirect effect of aberrant miRNA expression. A more targeted approach would be to directly characterize the miRNA-target gene interface. We plan to do this by capturing the miRNA-mRNA population of cells using an Ago2 immunonoprecipitation (IP) technique followed by deep sequencing (see section 1.4.7). It has been reported that a high overlap of miRNAs loaded into Ago 1-4

isoforms suggesting that the use of Ago2 pull down is sufficient to have a representative sample over all Ago targets [383].

The goal of this project was to determine the target genes of *miR-155* in DLBCL cell models, a well-known miRNA implicated in lymphomagenesis, and study the biological effect that promotes a deregulation of this miRNA. We studied this effect in two cell models of DLBCL corresponding to two different subtypes of the disease that also differ in the expression levels of this miRNA aiming to find difference in the *miR-155* targetome between this subtypes that could somehow explain the existing appreciable phenotypic differences between them as well as exploring the possibility of finding biomarkers enabling the classification of these two molecular subtypes. To do this, we used a modified PAR-CLIP protocol. The original protocol we used as reference was optimized for performing this assay in HEK293 cell model genetically modified in order to overexpress an exogenous Ago2 protein [306].

Firstly, we had to adapt the protocol from the HEK293 to DLBCL cell lines. The major difference between them is that lymphoid cell lines are non-adherent and consequently more difficult to infect. Therefore, we did the infection starting with fairly high MOIs and we finally set a MOI of 15 as we demonstrated that it was already optimal and increasing the MOI would not be beneficial. Also, when we monitor *miR-155* expression in the infected cells, with the knockdown system used, we couldn't ensure a stable downregulation of the *miR-155* so this condition was discarded for the subsequent experiments. Consequently, we used the endogenous levels of *miR-155* expression of these cell lines as a control for the PAR-CLIP experiments (cells infected with a vector expressing a scramble miRNA sequence).

To adapt the protocol to DLBCL cells we first ensured that the endogenous expression levels of Ago2 protein were sufficient and also that the cells stay metabolically active during the 4-SU incorporation and that we can successfully incorporate this nucleoside into the RNA nascent molecules. We determined that the expression of endogenous Ago2 is high in DLBCL. Although, SU-DHL-10 expresses approximatley 30% less Ago2 than HEK293, RIVA twice as much Ago2 expression as SU-DHL-10, higher than even the FLAG-induced HEK293-Ago2 cells. This higher expression of Ago2 in the ABC-DLBCL

5. Elucidating the miRNA-targetome of DLBCL

(RIVA) cell line compared to the GCB-DLBCL (SU-DHL-10) cell line is consistent with previous observations [384].

We also looked at the effect of 4-SU on the DLBCL cell lines, as it has been reported that high levels of 4-SU can inhibit cell proliferation and production and processing of some RNAs [385]. We did indeed observe reduced cell viability at 200 µM, an effect that was even more pronounced at 400 µM. Using a concentration of 100 µM 4-SU did not appear to affect viability and we observed at that concentration that DLBCL cells successfully incorporated 4-SU in their nascent RNA molecules. Actually, the incorporation of 4-SU or RNA synthesis ratio was only 1.2 (RIVA) and 1.5 (SU-DHL-10) times less than that of HEK293. These data suggest that the quality of the recovered RNA might be better from the RIVA cells compared to SU-DHL-10 cells. We also checked if the IP conditions we were using were optimum as, in this modified PAR-CLIP assay, we used an antibody against Ago2 protein instead of an antibody for the tagged exogenous Ago2 (antibody against FLAG tag) used in the original protocol [306]. Results indicated that the employed IP conditions were appropriate for Ago2 capture.

A limitation of this technique might be the library preparation method used for NGS. We used *TruSeq Small RNA Sample Prep Kit* applying some modifications to the protocol to adapt it for low input RNA characteristic of CLIP protocols [386]. Specifically, we used a 1:10 dilution of the adapters to avoid self-ligation and we increased the number of PCR cycles from 11 to 17. Also, for the library purification the polyacrylamide percentage was increased from 6% to 8% in the gel to eliminate more efficiently the ligated empty adaptors. The sequencing results reflected the low amount of input RNA as more than 90% of the reads were clonal, probably due to PCR over- amplification. This might have promoted a biased in the sequencing results as it might be an over-representation of the small reads over the larger reads. In addition, the chemical reactions of the library preparation itself, such as the requirement of two independent intermolecular reactions for the ligation of the adaptors also affect the library preparation yield (Figure 5.19). The library should be a representative sample of the complexity of the input RNA but it might occur that not all of the RNA becomes part of the library. The high clonality of the libraries is a common fact in CLIP experiments. After removing the PCR duplicates the amount of data usually drops

162

dramatically and can be reduced up to <1/10 of its original size. Also, according to other PAR-CLIP data the percentage of unique mapped reads is similar to our results [386, 387]. Importantly, these reads cover the 0.22% of the genome indicating a specific RNA capture instead of widespread.
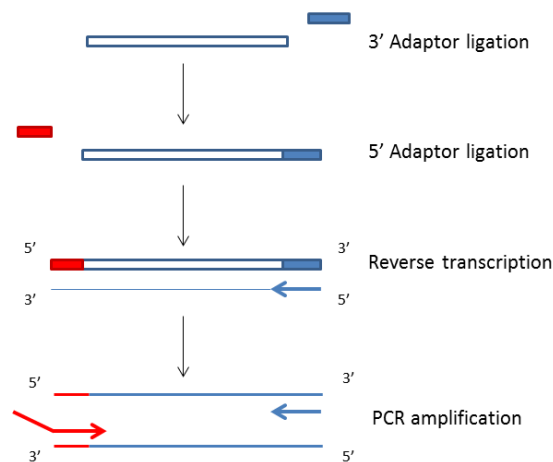
Figure 5.19: TruSeq Small RNA Library Preparation Kit chemistry. Figure shows a diagram of the steps followed for the library preparation in the PAR-CLIP samples.

PAR-CLIP technique improves signal-to-noise ratio over other CLIP techniques because allows, due to the characteristic T-to-C mutation, the specific identification of RNA cross-linked to Ago2 [292]. In addition, the RNA yield after crosslinking is increased compared to traditional UV-crosslinking [296]. However, this technique might be less sensitive than other CLIP technologies as RNA molecules without a 4-SU incorporation are not considered. A recent paper demonstrated that non-T-to-C clusters, frequently observed in PAR-CLIP experiments, exhibit functional miRNA-binding events and strong RNA accessibility [388].

The sequencing results indicated an enrichment in T-to-C changes over all the rest of possible nucleotide changes in the unique and correctly aligned reads suggesting that the recovered RNA was enriched for RNA crosslinked to Ago2 protein. This percentage of T-to-C changes was higher in the RIVA cell line than in the SU-DHL-10 cell line consistent with the higher expression of Ago2 and the higher amount of 4-SU incorporated in the RNA of these cells. This PAR-CLIP specific mutation occurs during

the cDNA synthesis due to the crosslink between the 4-SU residues incorporated in the RNA and the interacting protein aminoacids [296] (Figure 5.20). All the experimental conditions showed an enrichment in this T-to-C mutation. This is also a good indicator that the methodology was successful.
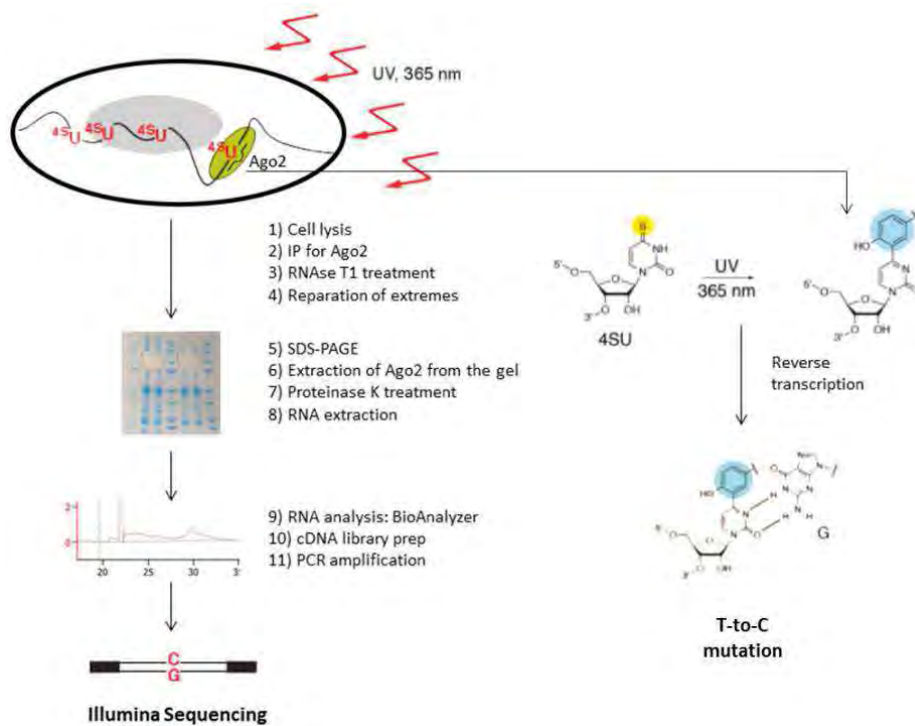
Figure 5.20: Modified PAR-CLIP protocol outline. *4SU-labelled transcripts were crosslinked to RBPs. Ago2-RNA complexes were immunopurified and RNA molecules were recovered and sequenced.*

For the cluster calling three different tools were used as there is not agreement about the most appropriate tool for the analysis of PAR-CLIP data. WavClusteR tool identified more clusters than any other tool. In contrast, much fewer clusters were identified by BMix, and nearly all the clusters identified were also present in the wavClusteR analysis, demonstrating a high degree of concordance between both tools. WavClusteR is a pipeline specific for PAR-CLIP data analysis written in R, that considers PAR-CLIP mutations only those that contain T-to-C changes, and normally excludes those with an extreme VAF percentage (too low or too high) [373]. In contrast, BMix, which is written in MATLAB and R considers multiple sources of background noise before the cluster calling resulting in a higher stringency or specificity [374]. DCLIP, the other algorithm

we used is written in Perl, is not specific for PAR-CLIP data and can be applied for the analysis of other CLIP techniques and so it does not dismiss clusters without T-to-C changes. dCLIP performs differential analysis between two conditions and finds differences in binding strengths rather than just the binary event of whether a binding site is shared or not [375]. Interestingly, when we compare the identity of the identified genes using these tools against predicted *miR-155* target genes along with experimentally validated genes, we determined that PAR-CLIP was able to identify more validated *miR-155* target genes than the in-silico prediction.

For the candidate selection, genes with clusters in 5'UTRs, exonic regions or 3'UTRs that were differentially identified between the experimental (over-miR-155) and the control condition (scramble) were considered. The genes identified in RIVA cell line, ABC-DLBCL subtype model, were enriched for genes involved in FOXO and AMPK signalling pathways suggesting that genes involved in these pathways might be directly regulated by *miR-155* in ABC-DLBCL subtype. AMPK signalling pathway coordinates cell growth and metabolism [389]. Activated AMPK inhibits mTOR pathway and consequently proliferation and survival of lymphoma cells [390, 391]. In addition, higher expression of AMPK is associated with better prognosis in NHL suggesting that downregulation of this pathway promotes the progression of NHL [392]. On the other hand, FOXO transcription factors are considered as tumor suppressors that limit cell proliferation and induce apoptosis [393] and FOXO signalling pathways are inactivated by oncogenic signals including miRNA regulation in multiple cancer types [394]. Our results suggest that *miR-155* might be targeting these pathways in ABC-DLBCL and therefore promoting lymphoma progression.

For the candidate selection as potential *miR-155* relevant target genes, those genes with a *miR-155* seed match in the identified cluster were prioritized. However, experimental evidence indicates that the majority (about 60 %) of the miRNA-binding activity is non-canonical [356, 357] and these has been also observed specifically for Ago-miR-155 binding sites [274]. Therefore, following these criteria we might be losing promising candidate genes following non-canonical binding rules. Similarly, the genes that were identified by PAR-CLIP which were downregulated comparing to the scramble control condition in a transcriptome analysis using microarrays were also

prioritized. However, we found only two genes in each cell line meeting these criteria. Different explanations for this are that miRNA regulation might just affect protein expression and not necessarily mRNA expression and also that the control cells express basal level of *miR-155*. In DLBCL cells, especially in RIVA (ABC subtype), these levels are already high. Therefore, it is understandable that we didn't find many differences in the microarray data.

Among the prioritized genes, some interesting candidates can be highlighted such as *E2F2, MICB* and *MCOLN2.* Even if some studies associate the upregulation of E2F2 transcription factor with cancer progression and poor prognosis [395, 396], in T cell lymphoma it has been described as a tumour suppressor. The inactivation of this transcription factor promotes lymphomagenesis and its reactivation resultes in increased apoptosis [397]. In addition, *E2F2* was reported as a direct target of *miR-155* in clear cell renal cell carcinoma and this miRNA was demonstrated to functions as a tumour promoting miRNA in this pathology [398]. Similarly, the upregulation of MICB, ligand of NKG2D receptor, enhances natural killer cell cytotoxicity in acute lymphoblastic leukemia (ALL) and NHL. Therefore, a targeted downregulation of this molecule might promote lymphoma progression [399]. In addition, *MCOLN2* regulatory region was found hypermethylated and downregulated in ALL indicating the potential tumor suppressor activity of this gene. Downregulation of this gene via *miR-155* targeting might be an alternative mechanism for controlling the expression of this gene in DLBCL [400].

In summary, in this chapter we successfully performed a PAR-CLIP experiment with a self-modified protocol avoiding the use of radioactivity and using the endogenously produced Ago2 protein for IP. We also used different experimental conditions, with overexpression and basal levels of *miR-155,* to facilitate the detection of noise. We identified a set of differentially expressed genes related with AMPK and FOXO signalling pathways exclusively in the ABC-DLBCL model suggesting that *miR-155* might be targeting these pathways in this DLBCL subtype and therefore promoting lymphoma progression. Also, among the genes presenting seed match sequence for *miR-155* some interesting candidates can be highlighted such as *E2F2*, *MICB* and *MCOLN2.* However,

these candidate genes should be validated with additional functional experiments. This might allow finding novel *miR-155* targets differentially expressed between the different DLBCL molecular subtypes and therefore identify potential biomarkers for their classification.

5. Elucidating the miRNA-targetome of DLBCL

6. DISCUSSION

Aberrant miRNA expression has been widely demonstrated in many cancer research studies [186, 203-205]. However, the significance of such deregulation remains poorly understood. The hypothesis of this thesis was that miRNA binding sites play important roles in the pathogenesis of non-Hodgkin B-cell lymphomas so, the study of the specific target genes of these miRNAs can led us to better understand the functional consequence of miRNA deregulation in these lymphomas. To do that, we employed two different strategies:

On one hand, we assessed whether mutations located in miRNA binding sites could function as pathological mechanisms by interfering the correct miRNA regulation in FL, and whether these mutations were associated with FL transformation and therefore serve as useful biomarkers for prognosis. For the identification of these mutations WGS data from FL patients was analyzed and a bespoke bioinformatics pipeline used for miRNA target site prediction. In contrast, we employed an experimental approach to analyse the targetome of *miR-155*, a well-known oncomiR in lymphomagenesis, in GCB and ABC DLBCL. This miRNA is differentially expressed between these two molecular subtypes [370, 371]. The aim of this study was to better understand the biological meaning of the *miR-155* overexpression in DLBCL and the differential expression between these subtypes by experimentally finding the specific target genes of these miRNA in situ.

The importance of miRNA function is clear as these regulatory molecules have been identified in multiple organisms: from unicellular and simple organisms to much more developed and complex systems [158, 401]. Reflecting the importance of these molecules to biology, miRNA sequences are highly conserved in evolution [262]. Mutations in miRNA coding sequences are rare [402], and the first evidence came from lymphoma where a region frequently deleted in CLL (13q14 locus) was shown to encode for the miRNAs *miR-15a* and *miR-16-1* [237]. Similarly, the first description of a miRNA SNP was in *pri-miR-16-1* in a patient with familial CLL [403]. In contrast, miRNA target sites do not show this high conservation so, the emergence of polymorphisms in these sites constitutes a source of variability promoting the formation of new target site for specific miRNAs as well as the disruption of exiting miRNA regulations. This mechanism might be acquired by cancer cells: through the insertion of alterations (i.e

point mutations) in specific miRNA binding sites cancer cells might be able to deregulate the expression of certain genes or signalling pathways in order to promote cell proliferation and survival.

Even if less conserved than miRNA encoding sequences, miRNA binding sequences are also more conserved than would be expected by chance. Indeed, many bioinformatics tools rely on the target site conservation for their miRNA target predictions and focus on those sites located in 3'UTRs and following canonical binding rules (see section 1.4.6). However, according to our experimental results in the *miR-155* targetome identification in DLBCL as well as in additional experimental studies [274, 356, 357], the majority of the identified miRNA binding follows non-canonical rules. This non-canonical or seedless binding has been demonstrated to be functional [189-191] even though some studies have reported that the majority of non-canonical binding mediates modest gene repression [191, 273-275] probably due to a lower binding affinity. Also, these sequences show modest sequence conservation across species [191, 273-275] suggesting that they may be evolutionary intermediates under selective pressure for a shift towards high affinity seed sites. Therefore, the biological meaning of non-canonical interactions could be interpreted as evolutional intermediates with slightly distinct functions and, the combination of canonical and non-canonical sites may provide a variety of spectra in the regulation of gene expression, enabling a fine-tuning of repression activity.

In the chapter four of this thesis we used bioinformatics prediction not to study the targetome itself, but to do a screening of possible candidate mutations in tumor samples. We showed that some of the predictions fail in the experimental validation and also that miRNA targeting is not always based on seed matching. We demonstrated that mutations in miRNA binding sites are not randomly distributed but located in genes associated with GC-like B-cell lymphomas. Importantly, we did demonstrate that mutations in miRNA binding sites in two key genes in FL such as *BCL2* and *EZH2* impair the miRNA binding promoting their own deregulation.

In contrast, in chapter five, we performed an experimental methodology (PAR-CLIP) to discover the targetome of *miR-155* in DLBCL models based on an Ago2 IP technique followed by deep sequencing. These methods are expensive and challenging and they

still have several limitations to overcome. For example, in a PAR-CLIP experiment, the amount of RNA cross-linked to Ago2 recovered for sequencing is usually very low [386]. Therefore, a library preparation methodology adapted to this input RNA is required. To address this need, library preparation kits specific for low input DNA/RNA are emerging [404] as it is highly important that the library becomes a representative sample of the complexity of the input RNA. A second limitation might be the lack of standardization of the methods for data analysis. Also, for an optimal PAR-CLIP assay is essential to have good cellular models and negative control samples to facilitate the identification of background noise in the data. A possible negative control is to do PAR-CLIP on a sample without adding 4-SU to the cells. This might help in the identification of true PAR-CLIP mutations. Another possible negative control is to do PAR-CLIP on a non-crosslinked sample as well as to use an isotype control antibody for the IP instead of an Ago2 antibody. This might help to distinguish the general background noise from the technique from your real data. Interestingly, a bioinformatics tool to identify common background presence in PAR-CLIP datasets called BackCLIP has been developed [405]. The improvement of CLIP technologies may allow an efficient analysis of the miRNA targetome in different cell models and probably using a reduced number of cells facilitating the transition to the performance of these methodologies in primary cell samples from patients. However, to reach this point is essential to overcome these technique limitations. This might allow an efficient identification of biomarkers for diagnosis and prognosis for B-cell lymphoma patients.

In summary, in this thesis we have used two complementary techniques to look at the role of miRNA-binding in B cell lymphomas. Both techniques suggest that not only are binding sites important and non-randomly distributed in genes associated with B cell lymphomagenesis but that mutations in these sites are also enriched in lymphoma progression, both facts pointing to the relevance of miRNA binding to lymphoma pathology and demonstrating that the work contained in this thesis has made a significant contribution to extending the field of miRNAs and showing the need to fully understand the relevance of dysfunctional miRNA expression in lymphomas. It is therefore hoped that this work, while clearing illustrating the amount of research still to be carried out and the technical difficulties involved in carrying out these studies,

6. Discussion

will serve as a first step in achieving this understanding and therefore help in the fight against these devastating diseases.

7. CONCLUSIONS

1. Mutations in putative miRNA-binding sites are frequent in FL and occur preferentially in 3'UTRs of genes associated with haematological malignancies.

2. Samples from FL patients at the transformed state have a higher number of mutations in putative miRNA-binding sites compared to the sample before transformation.

3. Targeted re-sequencing of the mutations in putative miRNA-binding sites using Ampliseq technology showed an 80% of overlapping with the mutations identified using WGS (Illumina technology) in FL patients.

4. Ampliseq sequencing detected a high amount of false positive recurrent mutations, approximately half of them, near repetitive and homopolymer regions.

5. The 27 recurrently mutated candidate genes in FL patients are highly enriched for GC-like B-cell lymphoma suggesting a biological meaning of these mutations in FL disease.

6. An association between the presence of these variants and the transformation event in FL couldn't be demonstrated due to the low amount of mutated cases for specific mutations.

7. Experimental validation of the predicted miRNA binding sites demonstrated that the majority (7/9) were functional as the luciferase activity was significantly reduced compared with the control condition when analysing the non-mutated or wild type target sites.

8. The position of the mutation or variant within the binding site is not associated with the effect of the variant on the miRNA-binding highlighting the importance of non-canonical miRNA binding.

9. The mutations in *EZH2* (Y641 codon) and *BCL2* (*hg19 chr18:g.[60793447G>A]*) identified in FL patients impair the *miR-144* and *miR-5008* binding respectively

demonstrating the functionality of this mechanism in two key genes in B cell lymphomagenesis.

10. *MiR-144* overexpression in DLBCL-derived cell lines induces the downregulation of *EZH2* at mRNA and protein levels in an EZH2 wild type model but not in a cell line carrying the Y641F mutation suggesting the relevance of this mechanism in lymphoid cell linage.

11. This potential novel mechanism in FL requires further investigation to assess its relevance to FL patients.

12. DLBCL cell lines express high levels of endogenous Ago2, especially RIVA cell line (ABC-DLBCL) that expresses higher levels than even the FLAG-induced HEK293-Ago2 cells.

13. 4-SU at a concentration of 100 µM in the cell culture media does not affect the viability of DLBCL cell lines. At this concentration cells remain metabolically active and successfully incorporated 4-SU in their nascent RNA molecules.

14. The RNA synthesis ratio or 4-SU incorporation ratio is similar in DLBCL cell lines and HEK293 cell line.

15. Sequencing results for the identification of the *miR-155* targetome indicated that the libraries from the PAR-CLIP samples were highly clonal (less than 10% of the reads were unique after removing PCR duplicates) consistent with other CLIP results.

16. The processed PAR-CLIP reads showed to cover 0.22% of the genome indicating that this RNA capture is specific instead of widespread.

17. The PAR-CLIP reads were enriched in T-to-C changes over all the rest of possible nucleotide changes in the unique and correctly aligned reads indicating that the recovered RNA was enriched for RNA crosslinked to Ago2 protein. This percentage of T-to-C changes was higher in the RIVA cell line than in the SU-

DHL-10 cell line consistent with the higher expression of Ago2 and the higher amount of 4-SU incorporated in the RNA of these cells.

18. The majority of the identified clusters were located in exonic regions followed by 3'UTR regions, intergenic regions, intronic regions and 5'UTR regions consistent with other PAR-CLIP data.

19. Ontology analysis of the differentially identified genes between the experimental condition (over-miR-155) and the control condition (scramble) were found to be significantly enriched for B-cell receptor signalling and cell cycle in both DLBCL subtypes. But, interestingly, only the genes identified in the aggressive ABC-DLBCL subtype (RIVA cell line) showed an enrichment for genes involved in FOXO (adjusted p value = 9.51×10^{-4}) and AMPK (adjusted p value = 3.32×10^{-2}) signalling pathways suggesting that genes involved in these pathways might be directly target by *miR-155* specifically in ABC-DLBCL subtype.

20. Experimental target identification (PAR-CLIP) identifies more effectively than computational target prediction (TargetScan algorithm) the already validated miR-155 target genes according to miRTarBase.

21. For the identification of potential biomarkers for DLBCL classification additional functional experiments over these candidate genes would be required.

8. SUMMARY IN SPANISH

8.1 Introducción

Los linfomas B no Hodgkin son un grupo heterogéneo de enfermedades causadas por una proliferación neoplásica de linfocitos B. Su clasificación viene determinada por características clínicas, morfológicas, inmunofenotípicas, genéticas, citogenéticas y también por su célula de origen u homólogo no maligno [4]. El linfoma folicular (FL) y el linfoma B difuso de célula grande (DLBCL) son los tipos más comunes [4]. En la patogénesis de estos linfomas, aparte de alteraciones en oncogenes y genes supresores tumorales también se han descrito alteraciones relacionadas con microRNAs.

Los microRNAs (miRNAs) son RNAs no codificantes de pequeño tamaño (19-24 nucleótidos) que regulan la expresión génica a nivel post-transcripcional [158]. La función de estas moléculas ha sido demostrada en diversos procesos fisiológicos, incluida la regulación de la hematopoyesis [222], pero también en diferentes tipos de cáncer, incluido los linfomas B [206-219]. La expresión aberrante de miRNAs en un fenómeno común en cáncer. Las causas para esta desregulación pueden ser diversas incluyendo alteraciones cromosómicas o desregulación epigenética. Sin embargo, otros factores también pueden tener efecto en la función de estas moléculas como la presencia de mutaciones en zonas que codifican miRNAs, en sitios de unión de miRNAs o en genes implicados en su biogénesis [253] [256].

De cualquier modo, para entender el significado de la expresión aberrante de miRNAs es necesario identificar los genes diana específicos en un contexto celular concreto. Las herramientas computacionales para la predicción de dianas de miRNAs son un recurso rápido y útil. Sin embargo, la validación experimental de estas dianas es necesaria ya que estas predicciones a menudo fallan. Por este motivo, en los últimos años se han desarrollado métodos experimentales para la identificación de dianas de miRNAs. Estos métodos, denominados CLIP (crosslinking and immunoprecipitation) identifican interacciones directas entre RNA y proteína. Consisten en la immunoprecipitación (IP) de la proteína Ago2 (proteína miembro del complejo RISC) y el aislamiento e identificación de los fragmentos de RNA covalentemente unidos a la proteína mediante irradiación con luz ultravioleta [294]. La variante PAR-CLIP (Photoactivatable ribonucleoside-enhanced CLIP) ha demostrado dos ventajas sustanciales respecto a

otras técnicas CLIP. En el protocolo PAR-CLIP se ha mejorado la eficiencia del entrecruzamiento RNA-proteína. Las células se incuban con un ribonucleósido fotoactivable que se incorpora en los ácidos nucleicos recién sintetizados y permite un entrecruzamiento específico de este residuo con las proteínas a una longitud de onda de menor penetrancia [296]. Además, la presencia de este análogo en el RNA capturado produce una mutación característica durante el proceso de síntesis de cDNA, lo que facilita su identificación [296, 306].

8.2 Hipótesis y objetivos

La hipótesis de esta tesis es que los sitios de unión de miRNAs tienen una función relevante en la fisiopatología de los linfomas B no Hodgkin. Para testar esta hipótesis utilizamos dos aproximaciones. En primer lugar identificamos mutaciones en sitios de unión de miRNAs en tumores de pacientes y el estudiamos su relevancia funcional. En segundo lugar utilizamos la técnica PAR-CLIP con el fin de dilucidar el targetoma de miRNAs comúnmente deregulados en estos linfomas como es el caso del *miR-155*.

Los objetivos que se planteraron para esta tesis fueron los siguientes:

• Identificar y analizar la funcionalidad de mutaciones en sitios de unión de miRNAs como posible mecanismo patogénico del linfoma folicular (FL)

• Determinar la relevancia biológica de mutaciones en sitios de unión de miRNAs y su asociación con la transformación a linfoma de alto grado.

• Evaluar si las mutaciones en sitios de unión de miRNAs pueden servir como biomarcadores pronóstico para la identificación de pacientes con FL con riesgo de transformación a linfoma de alto grado.

• Identificar el targetoma del *miR-155* en los subtipos de centro germinal (GCB) y de célula B activadas (ABC) del linfoma B difuso de célula grande (DLBCL) con el fin de determinar el significado de su expresión diferencial entre los dos subtipos.

• Identificar biomarcadores de clasificación para pacientes con DLBCL entre los dos subtipos moleculares GCB y ABC mediante el análisis del targetoma del *miR-155.*

8.3 Identificación y análisis de mutaciones en sitios de unión de miRNAs en linfoma folicular

El linfoma folicular (FL) es el linfoma B de bajo grado más común y representa el 20% de todos los linfomas diagnosticados [48]. A pesar de tratarse de un linfoma de progresión lenta, el 45% de los pacientes sufren una transformación a un linfoma mucho más agresivo y de peor pronóstico reduciendo drásticamente su esperanza de vida [49, 89]. A pesar de la gravedad de esta complicación, no existen a día de hoy biomarcadores para predecir el riesgo de sufrirla. Por tanto, el objetivo de este proyecto es identificar mutaciones en sitios de unión de microRNAs asociadas a la transformación que puedan resultar útiles como biomarcadores pronóstico al mismo tiempo que analizar el impacto biológico de esta pérdida de regulación en la fisiopatología es este linfoma. Para probar nuestra hipótesis, interrogamos datos de secuenciación del genoma completo (WGS) de muestras que provenían de 6 pacientes con linfoma folicular que habían sufrido un proceso de transformación y su material constitutivo correspondiente. Utilizando el algoritmo de predicción de TargetScan identificamos un total de 544 mutaciones presentes en sitios de unión de miRNAs. Los resultados mostraron un enriquecimiento de estas dianas en zonas 3'UTR y en genes relacionados con enfermedades hematológicas ($P = 2,18x10^{-4}$) sugiriendo que estas mutaciones no son al azar sino que es posible que tengan una función biológica.

La secuenciación Sanger de las mutaciones en 34 genes candidatos permitió validar un tercio de las mismas. Esto pudo deberse a la baja sensibilidad de esta técnica en comparación con las técnicas de secuenciación de próxima generación (NGS) ya que las frecuencias del alelo mutado menores al 15-20% no son detectables mediante esta técnica [351]. En cambio, empleando secuenciación dirigida se validaron el 80% de las mutaciones inicialmente identificadas. Al explorar su presencia en una cohorte de validación vimos que las variantes recurrentes (30 variantes en 27 genes) se encontraban en genes relacionados con linfomas B de centro germinal ($P = 4,39 \times 10^{-5}$) lo que nuevamente sugiere que estas variantes probablemente tengan un significado biológico relacionado con la patología. Sin embargo, ninguna de ellas se pudo asociar al proceso de transformación.

Además, pudimos validar este mecanismo patogénico mediante ensayos de luciferasa. Los resultados mostraron como las mutaciones en *BCL2* (*hg19 chr18:g.[60793447G>A])* y *EZH2* (*hg19 chr7:g.[148508727T>A/G, 148508728A>T])*), dos genes clave en linfomas B de centro germinal, interferían con la unión del miRNA impidiendo su correcta regulación. Seguidamente nos dispusimos a analizar *in-silico* la relevancia funcional de este mecanismo en pacientes con FL explorando datos de microarrays publicados en la base de datos GEO. Analizamos los niveles de expresión de *miR-144*, microRNA que demostramos que regulaba *EZH2*, y observamos que la expresión de ambos es variable en pacientes. Sin embargo, los datos no permitieron hacer una asociación entre los niveles de expresión del *miR-144, EZH2* y su genotipo *wild type* o mutado. El analisis *in-silico* de *BCL2* y *miR-5008* no se pudo realizar debido a la falta de datos disponibles.

EZH2 ya había sido descrito como diana de *miR-144* previamente en cáncer [359-361] y algunos estudios han demostrado una correlación inversa entre la expresión de *miR-144 y EZH2* [360, 361]. La expresión de *miR-144* se muestra variable en pacientes según el análisis *in-silico* realizado y además, su relevancia en pacientes con linfomas B ha sido demostrada [362]. Por otro lado, el mecanismo para *BCL2* sería particularmente interesante en pacientes con ausencia de la translocación t(14;18) ya que podría tratarse de un mecanismo de sobrexpresión de BCL2 alternativo [363] en aquellos pacientes con una alta expresión del *miR-5008*. Debido a la ausencia de datos no pudimos evaluar su expresión en pacientes pero ganancias cromosómicas del brazo 1q, donde se codifica el *miR-5008,* han sido descritas en un 20% de tumores de FL [95].

Por tanto, demostramos como estas mutaciones en EZH2 y BCL2 identificadas en pacientes con FL afectan negativamente a la unión del miRNA que los regula promoviendo la desregulación de los genes diana en los ensayos in-vitro realizados. Este potencial nuevo mecanismo en FL requiere investigación adicional para evaluar su relevancia en pacientes.

8.4. Estudio del targetoma de miRNAs en DLBCL

La expresión aberrante de microRNAs (miRNAs) es un fenómeno común en cáncer [186, 203-205]. Sin embargo, en la mayoría de los casos se desconoce el significado

funcional de dicha desregulación. Para comprender plenamente su función, es esencial identificar los genes diana de un miRNA concreto en un contexto celular relevante. Muchas de las metodologías actuales más utilizadas, basadas en algoritmos de predicción, no contemplan dicho contexto y en la práctica no funcionan del todo. Por ello, nos dispusimos a identificar el *targetoma* del *miR-155*, un regulador clave en la linfomagénesis y conocido oncomiR [243-246], en modelos celulares de linfoma B difuso de célula grande (DLBCL) mediante la técnica PAR-CLIP. Utilizamos para ello modelos celulares de los subtipos de GCB-DLBCL y ABC-DLBCL que difieren en los niveles de expresión de este miRNA [370, 371] y adaptamos el protocolo original [306] con el fin de evitar el uso de radioactividad y utilizar únicamente la proteína Ago2 endógena para la IP. Además, modulamos la expresión del *miR-155* en estas líneas celulares para mejorar la relación señal ruido al comparar el targetoma en células que sobrexpresar este miRNA con células *wild type* o células que lo inhiben. El objetivo de este proyecto es encontrar biomarcadores de clasificación para pacientes con DLBCL mediante la identificación del *targetoma* del *miR-155* en los dos subtipos a la par que determinar la relevancia biológica de esta expresión diferencial.

Inicialmente modulamos la expresión del *miR-155* en modelos celulares de GCB-DLBCL y ABC-DLBCL (líneas celulares SU-DHL-10 y RIVA respectivamente) utilizando infecciones lentivirales. Para ello, analizamos previamente la MOI óptima de infección y la resistencia intrínseca de estas células a la puromicina con el fin de maximizar la eficiencia de la infección y la selección. Las células fueron infectadas (MOI 15) para sobrexpresar la secuencia del *miR-155* (condición Over-miR-155), para sobrexpresar la secuencia complementaria al *miR-155* (condición anti-miR-155) y para sobrexpresar una secuencia control (condición scramble). Sin embargo, para los experimentos PAR-CLIP utilizamos las células con sobrexpresión del miRNA y, como control, su nivel de ezpresión basal (Scramble) ya que no conseguimos una modulación estable al infectar las células con el inhibidor del *miR-155*. Antes de realizar el ensayo PAR-CLIP tuvimos en cuenta factores como la expresión endógena de la proteína Ago2, la toxicidad del 4SU y su incorporación en el RNA de nueva síntesis en estas células. También evaluamos que, con el anticuerpo para reconocer Ago2 endógeno immunoprecipitábamos correctamente esta proteína con las condiciones que

habíamos establecido en la adaptación de nuestro protocolo. Tras optimizar todo el proceso realizamos los experimentos PAR-CLIP.

Los datos de secuenciación del RNA capturado mostraron que menos del 10% de las lecturas obtenidas eran únicas indicando una sobre amplificación del RNA aislado durante la fabricación de las librerías de secuenciación. De las lecturas únicas, el 21% alinearon correctamente contra el genoma de referencia (hg19/GRCh37). Estas lecturas cubren el 0.22% del genoma. Estos datos son comparables a los de otros ensayos CLIP donde es común obtener librerías de baja complejidad [386, 387]. Esto puede deberse a una limitación del sistema de fabricación de las librerías pero también podría ser un reflejo de la especificidad de la unión de Ago2 a motivos específicos.

El perfil mutacional de estos datos mostró un enriquecimiento para la mutación (T-to-C) característica de la técnica debida a la incorporación del análogo fotoactivable 4-SU lo que indica una buena captura de RNAs asociados a Ago2. Para la identificación de los clusters; es decir, lecturas no idénticas que se agrupan o solapan en una misma región, se utilizaron tres softwares especialidos en análisis de datos CLIP: wavClusteR [373], BMix [374] y dCLIP [375]. Los resultados mostraron una fuerte correlación entre los genes diana identificadas por los diferentes softwares aunque presentabas diferencias entre sí en cuanto a la sensibilidad y especificidad. En la distribución de los clusters identificados observamos que el 47% pertenecían a regiones codificantes mientras que el 19% a 3'UTRs y solo el 4% a 5'UTRs. El resto pertenecían a regiones intronicas e intergénicas. Este tipo de distribución, donde la mayoría de los clusters identificados alinean en zonas exónicas se ha observado también en otros ensayos CLIP [296, 326]. Se consideraron únicamente los clusters localizados en regiones codificantes de proteínas (5'UTR, exones y 3'UTRs) para la selección de genes candidatos.

Los resultados mostraron que el 82% de los genes se identificaron de forma diferencial entre la condición experimental (over-miR-155) y el control (scramble) en RIVA y el 78% en SU-DHL-10. Estos genes mostraron estar relacionados con la señalización del receptor B y con el ciclo celular en ambos subtipos. Sin embargo, solo los genes identificados en el subtipo ABC (línea celular RIVA) mostraron estar relacionados con

las vías de señalización FOXO (p valor ajustado = 9.51 x 10^{-4}) y AMPK (p valor ajustado = 3.32 x 10^{-2}) sugiriendo que los genes que forman parte de estas vías están directamente regulados por el *miR-155* exclusivamente en el subtipo ABC-DLBCL. La vía AMPK activada inhibe la vía mTOR y, por tanto, inhibe la proliferación y la viabilidad de las células de linfoma [390, 391] y se ha asociado a mejor pronóstico sugiriendo que, al regular negativamente esta vía se estaría favoreciendo la malignidad [392]. Por otro lado, los factores de transcripción FOXO se consideran supresores tumorales [393] y la vía FOXO de señalización es inactivada por señales oncogénicas incluyendo regulación por miRNAs en múltiples cánceres [394]. Por tanto, el miR-155 podría estar regulando negativamente estas vías y por consecuente promoviendo la progresión tumoral.

De cualquier modo, se requiere la validación de estos resultados y experimentos funcionales adicionales con el fin de identificar dianas noveles concretas del *miR-155* que se encuentren diferencialmente expresadas entre los dos subtipos moleculares de DLBCL. Esto podría llevar a la identificación de biomarcadores aplicables en la clínica para la correcta clasificación de estos linfomas.

8.5. Discusión

Para la realización de este trabajo empleamos dos aproximaciones complementarias con el fin de estudiar la funcion de los sitios de union de miRNAs en linfomas B. Los resultados de ambos estudios muestran que los sitios de unión de miRNAs no se encuentran distribuidos al azar sino en genes asociados con linfomagénesis y que la unión de miRNAs con frecuencia sigue reglas no canónicas. Al estudiar si la presencia de mutaciones en sitios de unión de miRNAs promueve una desregulación de los genes diana en un estudio no sesgado pudimos demostrar este mecanismo en dos genes clave en la biología del FL como son *BCL2* y *EZH2*. Estos hechos sugieren la relevancia de la función de los miRNAs en estas patologías y muestran la necesidad de entender la relevancia de la expresión aberrante de miRNAs en linfomas B.

8.5 Conclusiones

1. Las mutaciones en sitios de union de miRNAs son frecuentes en FL y ocurren preferencialmente en zonas 3'UTR de genes asociados con enfermedades hematológicas.

2. Los 27 genes seleccionados por presentar mutaciones recurrentes en sitios de union de miRNAs se encuentran altamente asociados con linfomas B de centro germinal.

3. No se pudo asociar ninguna de estas mutaciones con la transformación del FL debido al limitado número de pacientes con mutaciones concretas

4. Las mutaciones en *EZH2* (codon Y641) y la variante en *BCL2* (*hg19 chr18:g.[60793447G>A]*) impiden la correcta union del *miR-144* y *miR-5008* respectivamente demostrando la funcionalidad de este mecanismo en dos genes altamente relevantes en la linfomagenesis de célula B.

5. La sobrexpresión del *miR-144* en lineas celulares derivadas de DLBCL induce la regulación a la baja de *EZH2* a nivel de mRNA y proteína en células con *EZH2 wild type* pero no en células con la mutación Y641F. Esto sugiere la relevancia de este mecanismo en el linaje cellular linfoide.

6. En analisis del targetoma del *miR-155* mediante la técnica PAR-CLIP, las lecturas tras el procesamiento cubrían el 0.22% del genoma, indicando la especificidad en la captura del RNA.

7. La identificación experimental de dianas del miR-155 (PAR-CLIP) identifica más genes diana validados que los sistemas de predicción computacionales (algoritmo TargetScan).

8. Los genes diferencialmente identificados entre la condición experimental y control en el experiment PAR-CLIP demostraron estar significativamente enriquecidos para genes relacionados con la vía de señalización del receptor B y el ciclo cellular en ambos subtipos de DLBCL. Sin embargo, solo los genes

identificados en el subtipo agresivo ABC-DLBCL se encontraban significativamente involucrados en las vías de señalización AMPK y FOXO lo que sugiere que estas vías podrían estar directamente reguladas por el *miR-155* en este subtipo.

9. REFERENCES

1. Swerdlow, S.H., et al., *The 2016 revision of the World Health Organization classification of lymphoid neoplasms.* Blood, 2016. **127**(20): p. 2375-90.

2. *Global Cancer Observatory. International Agency for Research on Cancer*. 2018; Available from: gco.iarc.fr.

3. Hassan, M. and M. Abedi-Valugerdi, *Hematologic malignancies in elderly patients.* Haematologica, 2014. **99**(7): p. 1124-7.

4. Swerdlow, S., et al., *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. 2008, Lyon: International Agency for Research on Cancer (IARC).

5. Smith, A., et al., *Lymphoma incidence, survival and prevalence 2004-2014: sub-type analyses from the UK's Haematological Malignancy Research Network.* Br J Cancer, 2015. **112**(9): p. 1575-84.

6. Cartwright, R., et al., *The rise in incidence of lymphomas in Europe 1985-1992.* Eur J Cancer, 1999. **35**(4): p. 627-33.

7. Lunning, M.A. and J.M. Vose, *Management of indolent lymphoma: where are we now and where are we going.* Blood Rev, 2012. **26**(6): p. 279-88.

8. Jaffe, E.S., et al., *Classification of lymphoid neoplasms: the microscope as a tool for disease discovery.* Blood, 2008. **112**(12): p. 4384-99.

9. Busslinger, M., *Transcriptional control of early B cell development.* Annu Rev Immunol, 2004. **22**: p. 55-79.

10. Medina, K.L., et al., *Assembling a gene regulatory network for specification of the B cell fate.* Dev Cell, 2004. **7**(4): p. 607-17.

11. Inlay, M.A., et al., *Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development.* Genes Dev, 2009. **23**(20): p. 2376-81.

12. Li, Y.S., et al., *Identification of the earliest B lineage stage in mouse bone marrow.* Immunity, 1996. **5**(6): p. 527-35.

13. Milne, C.D. and C.J. Paige, *IL-7: a key regulator of B lymphopoiesis.* Semin Immunol, 2006. **18**(1): p. 20-30.

14. Lim, V.Y., et al., *A Chemoattractant-Guided Walk Through Lymphopoiesis: From Hematopoietic Stem Cells to Mature B Lymphocytes.* Adv Immunol, 2017. **134**: p. 47-88.

15. Ma, Y., et al., *Repair of double-strand DNA breaks by the human nonhomologous DNA end joining pathway: the iterative processing model.* Cell Cycle, 2005. **4**(9): p. 1193-200.

16. Schatz, D.G. and P.C. Swanson, *V(D)J recombination: mechanisms of initiation.* Annu Rev Genet, 2011. **45**: p. 167-202.

17. Schatz, D.G. and Y. Ji, *Recombination centres and the orchestration of V(D)J recombination.* Nat Rev Immunol, 2011. **11**(4): p. 251-63.

18. Agrawal, A. and D.G. Schatz, *RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination.* Cell, 1997. **89**(1): p. 43-53.

19. Lewis, S.M., *The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses.* Adv Immunol, 1994. **56**: p. 27-150.

20. Kuppers, R., et al., *Cellular origin of human B-cell lymphomas.* N Engl J Med, 1999. **341**(20): p. 1520-9.

21. Corcoran, L.M. and D.M. Tarlinton, *Regulation of germinal center responses, memory B cells and plasma cell formation-an update.* Curr Opin Immunol, 2016. **39**: p. 59-67.

22. Zhang, Y., L. Garcia-Ibanez, and K.M. Toellner, *Regulation of germinal center B-cell differentiation.* Immunol Rev, 2016. **270**(1): p. 8-19.

23. Janeway, C., et al., *Immunobiology: The Immune System in Health and Disease* 6th Edition ed. 2005: Garland Science.

9. References

24. Dent, A.L., et al., *Control of inflammation, cytokine expression, and germinal center formation by BCL-6.* Science, 1997. **276**(5312): p. 589-92.

25. Ye, B.H., et al., *The BCL-6 proto-oncogene controls germinal-centre formation and Th2-type inflammation.* Nat Genet, 1997. **16**(2): p. 161-70.

26. Phan, R.T. and R. Dalla-Favera, *The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells.* Nature, 2004. **432**(7017): p. 635-9.

27. Dickerson, S.K., et al., *AID mediates hypermutation by deaminating single stranded DNA.* J Exp Med, 2003. **197**(10): p. 1291-6.

28. Chaudhuri, J., et al., *Transcription-targeted DNA deamination by the AID antibody diversification enzyme.* Nature, 2003. **422**(6933): p. 726-30.

29. Allen, C.D., et al., *Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5.* Nat Immunol, 2004. **5**(9): p. 943-52.

30. Bannard, O., et al., *Germinal center centroblasts transition to a centrocyte phenotype according to a timed program and depend on the dark zone for effective selection.* Immunity, 2013. **39**(5): p. 912-24.

31. Klein, U. and R. Dalla-Favera, *Germinal centres: role in B-cell physiology and malignancy.* Nat Rev Immunol, 2008. **8**(1): p. 22-33.

32. Liu, Y.J., et al., *Germinal center cells express bcl-2 protein after activation by signals which prevent their entry into apoptosis.* Eur J Immunol, 1991. **21**(8): p. 1905-10.

33. Butch, A.W., et al., *Cytokine expression by germinal center cells.* J Immunol, 1993. **150**(1): p. 39-47.

34. Toellner, K.M., et al., *The human germinal centre cells, follicular dendritic cells and germinal centre T cells produce B cell-stimulating cytokines.* Cytokine, 1995. **7**(4): p. 344-54.

35. Linterman, M.A., et al., *Foxp3+ follicular regulatory T cells control the germinal center response.* Nat Med, 2011. **17**(8): p. 975-82.

36. Gordon, J., et al., *Central role of CD40 and its ligand in B lymphocyte responses to T-dependent antigens.* Cell Mol Biol (Noisy-le-grand), 1994. **40 Suppl 1**: p. 1-13.

37. Gulbranson-Judge, A., M. Casamayor-Palleja, and I.C. MacLennan, *Mutually dependent T and B cell responses in germinal centers.* Ann N Y Acad Sci, 1997. **815**: p. 199-210.

38. Tokoyoda, K., et al., *Organization of immunological memory by bone marrow stroma.* Nat Rev Immunol, 2010. **10**(3): p. 193-200.

39. Yoshida, T., et al., *Memory B and memory plasma cells.* Immunol Rev, 2010. **237**(1): p. 117-39.

40. Adachi, Y., et al., *Distinct germinal center selection at local sites shapes memory B cell response to viral escape.* J Exp Med, 2015. **212**(10): p. 1709-23.

41. Ise, W., et al., *Memory B cells contribute to rapid Bcl6 expression by memory follicular helper T cells.* Proc Natl Acad Sci U S A, 2014. **111**(32): p. 11792-7.

42. Fais, F., et al., *Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors.* J Clin Invest, 1998. **102**(8): p. 1515-25.

43. Hummel, M., et al., *Mantle cell (previously centrocytic) lymphomas express VH genes with no or very little somatic mutations like the physiologic cells of the follicle mantle.* Blood, 1994. **84**(2): p. 403-7.

44. Seifert, M., R. Scholtysik, and R. Kuppers, *Origin and pathogenesis of B cell lymphomas.* Methods Mol Biol, 2013. **971**: p. 1-25.

45. Rajewsky, K., *Clonal selection and learning in the antibody system.* Nature, 1996. **381**(6585): p. 751-8.

46. Lam, K.P., R. Kuhn, and K. Rajewsky, *In vivo ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death.* Cell, 1997. **90**(6): p. 1073-83.

47. Biagi, J.J. and J.F. Seymour, *Insights into the molecular pathogenesis of follicular lymphoma arising from analysis of geographic variation.* Blood, 2002. **99**(12): p. 4265-75.

48. Kahl, B.S. and D.T. Yang, *Follicular lymphoma: evolving therapeutic strategies.* Blood, 2016. **127**(17): p. 2055-63.

49. Wagner-Johnston, N.D., et al., *Outcomes of transformed follicular lymphoma in the modern era: a report from the National LymphoCare Study (NLCS).* Blood, 2015. **126**(7): p. 851-7.

50. Adam, P., et al., *Presence of preserved reactive germinal centers in follicular lymphoma is a strong histopathologic indicator of limited disease stage.* Am J Surg Pathol, 2005. **29**(12): p. 1661-4.

51. Horn, H., et al., *Follicular lymphoma grade 3B is a distinct neoplasm according to cytogenetic and immunohistochemical profiles.* Haematologica, 2011. **96**(9): p. 1327-34.

52. Salaverria, I. and R. Siebert, *Follicular lymphoma grade 3B.* Best Pract Res Clin Haematol, 2011. **24**(2): p. 111-9.

53. Boyd, S.D., et al., *Selective immunophenotyping for diagnosis of B-cell neoplasms: immunohistochemistry and flow cytometry strategies and results.* Appl Immunohistochem Mol Morphol, 2013. **21**(2): p. 116-31.

54. Olweny, C.L., *Cotswolds modification of the Ann Arbor staging system for Hodgkin's disease.* J Clin Oncol, 1990. **8**(9): p. 1598.

55. Michallet, A.S., et al., *Early stage follicular lymphoma: what is the clinical impact of the first-line treatment strategy?* J Hematol Oncol, 2013. **6**: p. 45.

56. Marcus, R., et al., *Phase III study of R-CVP compared with cyclophosphamide, vincristine, and prednisone alone in patients with previously untreated advanced follicular lymphoma.* J Clin Oncol, 2008. **26**(28): p. 4579-86.

57. Herold, M., et al., *Rituximab added to first-line mitoxantrone, chlorambucil, and prednisolone chemotherapy followed by interferon maintenance prolongs survival in patients with advanced follicular lymphoma: an East German Study Group Hematology and Oncology Study.* J Clin Oncol, 2007. **25**(15): p. 1986-92.

58. Solal-Celigny, P., et al., *Follicular lymphoma international prognostic index.* Blood, 2004. **104**(5): p. 1258-65.

59. Numata, A., et al., *Retrospective study of the utility of FLIPI/FLIPI-2 for follicular lymphoma patients treated with R-CHOP.* J Clin Exp Hematop, 2012. **52**(1): p. 77-9.

60. Federico, M., et al., *Follicular lymphoma international prognostic index 2: a new prognostic index for follicular lymphoma developed by the international follicular lymphoma prognostic factor project.* J Clin Oncol, 2009. **27**(27): p. 4555-62.

61. Pastore, A., et al., *Integration of gene mutations in risk prognostication for patients receiving first-line immunochemotherapy for follicular lymphoma: a retrospective analysis of a prospective clinical trial and validation in a population-based registry.* Lancet Oncol, 2015. **16**(9): p. 1111-1122.

62. Casulo, C., et al., *Early Relapse of Follicular Lymphoma After Rituximab Plus Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone Defines Patients at High Risk for Death: An Analysis From the National LymphoCare Study.* J Clin Oncol, 2015. **33**(23): p. 2516-22.

63. Castellino, A., et al., *Follicular Lymphoma: The Management of Elderly Patient.* Mediterr J Hematol Infect Dis, 2017. **9**(1): p. e2017009.

64. Roulland, S., et al., *Early steps of follicular lymphoma pathogenesis.* Adv Immunol, 2011. **111**: p. 1-46.

65. Tsujimoto, Y., et al., *The t(14;18) chromosome translocations involved in B-cell neoplasms result from mistakes in VDJ joining.* Science, 1985. **229**(4720): p. 1390-3.

9. References

66. Sungalee, S., et al., *Germinal center reentries of BCL2-overexpressing B cells drive follicular lymphoma progression.* J Clin Invest, 2014. **124**(12): p. 5337-51.

67. Roulland, S., et al., *Follicular lymphoma-like B cells in healthy individuals: a novel intermediate step in early lymphomagenesis.* J Exp Med, 2006. **203**(11): p. 2425-31.

68. Schuler, F., et al., *Prevalence and frequency of circulating t(14;18)-MBR translocation carrying cells in healthy individuals.* Int J Cancer, 2009. **124**(4): p. 958-63.

69. Kridel, R., L.H. Sehn, and R.D. Gascoyne, *Pathogenesis of follicular lymphoma.* J Clin Invest, 2012. **122**(10): p. 3424-31.

70. Schwaenen, C., et al., *Microarray-based genomic profiling reveals novel genomic aberrations in follicular lymphoma which associate with patient survival and gene expression status.* Genes Chromosomes Cancer, 2009. **48**(1): p. 39-54.

71. Cheung, K.J., et al., *Genome-wide profiling of follicular lymphoma by array comparative genomic hybridization reveals prognostically significant DNA copy number imbalances.* Blood, 2009. **113**(1): p. 137-48.

72. Cheung, K.J., et al., *Acquired TNFRSF14 mutations in follicular lymphoma are associated with worse prognosis.* Cancer Res, 2010. **70**(22): p. 9166-74.

73. Oricchio, E., et al., *The Eph-receptor A7 is a soluble tumor suppressor for follicular lymphoma.* Cell, 2011. **147**(3): p. 554-64.

74. Zhang, J., et al., *Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis.* Nat Med, 2015. **21**(10): p. 1190-8.

75. Ortega-Molina, A., et al., *The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development.* Nat Med, 2015. **21**(10): p. 1199-208.

76. Pasqualucci, L., et al., *Inactivating mutations of acetyltransferase genes in B-cell lymphoma.* Nature, 2011. **471**(7337): p. 189-95.

77. Beguelin, W., et al., *EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation.* Cancer Cell, 2013. **23**(5): p. 677-92.

78. Margueron, R. and D. Reinberg, *The Polycomb complex PRC2 and its mark in life.* Nature, 2011. **469**(7330): p. 343-9.

79. van Galen, J.C., et al., *Distinct expression patterns of polycomb oncoproteins and their binding partners during the germinal center reaction.* Eur J Immunol, 2004. **34**(7): p. 1870-81.

80. Bodor, C., et al., *EZH2 mutations are frequent and represent an early event in follicular lymphoma.* Blood, 2013. **122**(18): p. 3165-8.

81. Basso, K. and R. Dalla-Favera, *Roles of BCL6 in normal and transformed germinal center B cells.* Immunol Rev, 2012. **247**(1): p. 172-83.

82. Okosun, J., et al., *Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma.* Nat Genet, 2014. **46**(2): p. 176-181.

83. Okosun, J., et al., *Corrigendum: Recurrent mTORC1-activating RRAGC mutations in follicular lymphoma.* Nat Genet, 2016. **48**(6): p. 700.

84. Tweeddale, M.E., et al., *The presence of clonogenic cells in high-grade malignant lymphoma: a prognostic factor.* Blood, 1987. **69**(5): p. 1307-14.

85. Scott, D.W. and R.D. Gascoyne, *The tumour microenvironment in B cell lymphomas.* Nat Rev Cancer, 2014. **14**(8): p. 517-34.

86. Kiaii, S., et al., *Follicular lymphoma cells induce changes in T-cell gene expression and function: potential impact on survival and risk of transformation.* J Clin Oncol, 2013. **31**(21): p. 2654-61.

87. Pangault, C., et al., *Follicular lymphoma cell niche: identification of a preeminent IL-4-dependent T(FH)-B cell axis.* Leukemia, 2010. **24**(12): p. 2080-9.

88.   Ame-Thomas, P., et al., *Characterization of intratumoral follicular helper T cells in follicular lymphoma: role in the survival of malignant B cells.* Leukemia, 2012. **26**(5): p. 1053-63.

89.   Lossos, I.S. and R.D. Gascoyne, *Transformation of follicular lymphoma.* Best Pract Res Clin Haematol, 2011. **24**(2): p. 147-63.

90.   Okosun, J., S. Montoto, and J. Fitzgibbon, *The routes for transformation of follicular lymphoma.* Curr Opin Hematol, 2016. **23**(4): p. 385-91.

91.   Pasqualucci, L., et al., *Genetics of follicular lymphoma transformation.* Cell Rep, 2014. **6**(1): p. 130-40.

92.   Ngo, V.N., et al., *Oncogenically active MYD88 mutations in human lymphoma.* Nature, 2011. **470**(7332): p. 115-9.

93.   Compagno, M., et al., *Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma.* Nature, 2009. **459**(7247): p. 717-21.

94.   Kridel, R., et al., *Cell of origin of transformed follicular lymphoma.* Blood, 2015. **126**(18): p. 2118-27.

95.   Bouska, A., et al., *Genome-wide copy-number analyses reveal genomic abnormalities involved in transformation of follicular lymphoma.* Blood, 2014. **123**(11): p. 1681-90.

96.   Correia, C., et al., *BCL2 mutations are associated with increased risk of transformation and shortened survival in follicular lymphoma.* Blood, 2015. **125**(4): p. 658-67.

97.   Kridel, R., et al., *Histological Transformation and Progression in Follicular Lymphoma: A Clonal Evolution Study.* PLoS Med, 2016. **13**(12): p. e1002197.

98.   Shim, H., et al., *Prognostic impact of concordant and discordant cytomorphology of bone marrow involvement in patients with diffuse, large, B-cell lymphoma treated with R-CHOP.* J Clin Pathol, 2013. **66**(5): p. 420-5.

99.   Li, S., K.H. Young, and L.J. Medeiros, *Diffuse large B-cell lymphoma.* Pathology, 2018. **50**(1): p. 74-87.

100.   Wight, J.C., et al., *Prognostication of diffuse large B-cell lymphoma in the molecular era: moving beyond the IPI.* Blood Rev, 2018.

101.   Klapper, W., et al., *Patient age at diagnosis is associated with the molecular characteristics of diffuse large B-cell lymphoma.* Blood, 2012. **119**(8): p. 1882-7.

102.   Alizadeh, A.A., et al., *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.* Nature, 2000. **403**(6769): p. 503-11.

103.   Hans, C.P., et al., *Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray.* Blood, 2004. **103**(1): p. 275-82.

104.   Choi, W.W., et al., *A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy.* Clin Cancer Res, 2009. **15**(17): p. 5494-502.

105.   Visco, C., et al., *Comprehensive gene expression profiling and immunohistochemical studies support application of immunophenotypic algorithm for molecular subtype classification in diffuse large B-cell lymphoma: a report from the International DLBCL Rituximab-CHOP Consortium Program Study.* Leukemia, 2012. **26**(9): p. 2103-13.

106.   Barrans, S., et al., *Rearrangement of MYC is associated with poor prognosis in patients with diffuse large B-cell lymphoma treated in the era of rituximab.* J Clin Oncol, 2010. **28**(20): p. 3360-5.

107.   Quesada, A.E., et al., *Increased MYC copy number is an independent prognostic factor in patients with diffuse large B-cell lymphoma.* Mod Pathol, 2017. **30**(12): p. 1688-1697.

108.   Lu, T.X., et al., *MYC or BCL2 copy number aberration is a strong predictor of outcome in patients with diffuse large B-cell lymphoma.* Oncotarget, 2015. **6**(21): p. 18374-88.

109.   Visco, C., et al., *Patients with diffuse large B-cell lymphoma of germinal center origin with BCL2 translocations have poor outcome, irrespective of MYC status: a report from*

9. References

an International DLBCL rituximab-CHOP Consortium Program Study. Haematologica, 2013. **98**(2): p. 255-63.

110. Rosenthal, A. and A. Younes, *High grade B-cell lymphoma with rearrangements of MYC and BCL2 and/or BCL6: Double hit and triple hit lymphomas and double expressing lymphoma.* Blood Rev, 2017. **31**(2): p. 37-42.

111. Lin, P. and L.J. Medeiros, *High-grade B-cell lymphoma/leukemia associated with t(14;18) and 8q24/MYC rearrangement: a neoplasm of germinal center immunophenotype with poor prognosis.* Haematologica, 2007. **92**(10): p. 1297-301.

112. Johnson, N.A., et al., *Concurrent expression of MYC and BCL2 in diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone.* J Clin Oncol, 2012. **30**(28): p. 3452-9.

113. Hu, S., et al., *MYC/BCL2 protein coexpression contributes to the inferior survival of activated B-cell subtype of diffuse large B-cell lymphoma and demonstrates high-risk gene expression signatures: a report from The International DLBCL Rituximab-CHOP Consortium Program.* Blood, 2013. **121**(20): p. 4021-31; quiz 4250.

114. Xu-Monette, Z.Y., et al., *Clinical and biological significance of de novo CD5+ diffuse large B-cell lymphoma in Western countries.* Oncotarget, 2015. **6**(8): p. 5615-33.

115. Xu-Monette, Z.Y., et al., *Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP: report from an International DLBCL Rituximab-CHOP Consortium Program Study.* Blood, 2012. **120**(19): p. 3986-96.

116. Hu, S., et al., *CD30 expression defines a novel subgroup of diffuse large B-cell lymphoma with favorable prognosis and distinct gene expression signature: a report from the International DLBCL Rituximab-CHOP Consortium Program Study.* Blood, 2013. **121**(14): p. 2715-24.

117. *A predictive model for aggressive non-Hodgkin's lymphoma.* N Engl J Med, 1993. **329**(14): p. 987-94.

118. Zhou, Z., et al., *An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era.* Blood, 2014. **123**(6): p. 837-42.

119. Gine, E. and L.H. Sehn, *Diffuse Large B-Cell Lymphoma: Should Limited-Stage Patients Be Treated Differently?* Hematol Oncol Clin North Am, 2016. **30**(6): p. 1179-1194.

120. Kumar, A., et al., *Limited Stage Aggressive Non-Hodgkin Lymphoma: What Is Optimal Therapy?* Curr Treat Options Oncol, 2016. **17**(9): p. 45.

121. Chiappella, A., A. Castellino, and U. Vitolo, *State-of-the-art Therapy for Advanced-stage Diffuse Large B-cell Lymphoma.* Hematol Oncol Clin North Am, 2016. **30**(6): p. 1147-1162.

122. Feugier, P., et al., *Long-term results of the R-CHOP study in the treatment of elderly patients with diffuse large B-cell lymphoma: a study by the Groupe d'Etude des Lymphomes de l'Adulte.* J Clin Oncol, 2005. **23**(18): p. 4117-26.

123. Crump, M., *Management of Relapsed Diffuse Large B-cell Lymphoma.* Hematol Oncol Clin North Am, 2016. **30**(6): p. 1195-1213.

124. Hancock, C., et al., *A Retrospective Study of Double-hit Lymphomas in Elderly Patients (Aged > 70 Years): Overall Outcomes.* Clin Lymphoma Myeloma Leuk, 2018. **18**(4): p. 280-285.

125. Morin, R.D., et al., *Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin.* Nat Genet, 2010. **42**(2): p. 181-5.

126. Sneeringer, C.J., et al., *Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas.* Proc Natl Acad Sci U S A, 2010. **107**(49): p. 20980-5.

127. Yap, D.B., et al., *Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation.* Blood, 2011. **117**(8): p. 2451-9.

128. Morin, R.D., et al., *Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma.* Nature, 2011. **476**(7360): p. 298-303.

129. Pasqualucci, L., et al., *Analysis of the coding genome of diffuse large B-cell lymphoma.* Nat Genet, 2011. **43**(9): p. 830-7.

130. Cerchietti, L.C., et al., *BCL6 repression of EP300 in human diffuse large B cell lymphoma cells provides a basis for rational combinatorial therapy.* J Clin Invest, 2010. **120**(12): p. 4569-82.

131. Nogai, H., et al., *IkappaB-zeta controls the constitutive NF-kappaB target gene network and survival of ABC DLBCL.* Blood, 2013. **122**(13): p. 2242-50.

132. Young, R.M., et al., *B-cell receptor signaling in diffuse large B-cell lymphoma.* Semin Hematol, 2015. **52**(2): p. 77-85.

133. Phelan, J.D., et al., *A multiprotein supercomplex controlling oncogenic signalling in lymphoma.* Nature, 2018.

134. Knittel, G., et al., *Rewired NFkappaB signaling as a potentially actionable feature of activated B-cell-like diffuse large B-cell lymphoma.* Eur J Haematol, 2016. **97**(6): p. 499-510.

135. Davis, R.E., et al., *Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma.* Nature, 2010. **463**(7277): p. 88-92.

136. Lenz, G., et al., *Oncogenic CARD11 mutations in human diffuse large B cell lymphoma.* Science, 2008. **319**(5870): p. 1676-9.

137. Martins, G.A., et al., *Transcriptional repressor Blimp-1 regulates T cell homeostasis and function.* Nat Immunol, 2006. **7**(5): p. 457-65.

138. Nguyen, L., P. Papenhausen, and H. Shao, *The Role of c-MYC in B-Cell Lymphomas: Diagnostic and Molecular Aspects.* Genes (Basel), 2017. **8**(4).

139. Schmidt-Hansen, M., et al., *Does cell-of-origin or MYC, BCL2 or BCL6 translocation status provide prognostic information beyond the International Prognostic Index score in patients with diffuse large B-cell lymphoma treated with rituximab and chemotherapy? A systematic review.* Leuk Lymphoma, 2017. **58**(10): p. 2403-2418.

140. Bu, R., et al., *Bortezomib inhibits proteasomal degradation of IkappaBalpha and induces mitochondrial dependent apoptosis in activated B-cell diffuse large B-cell lymphoma.* Leuk Lymphoma, 2014. **55**(2): p. 415-24.

141. Wilson, W.H., et al., *Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma.* Nat Med, 2015. **21**(8): p. 922-6.

142. Kotla, V., et al., *Mechanism of action of lenalidomide in hematological malignancies.* J Hematol Oncol, 2009. **2**: p. 36.

143. Witzig, T.E., et al., *An international phase II trial of single-agent lenalidomide for relapsed or refractory aggressive B-cell non-Hodgkin's lymphoma.* Ann Oncol, 2011. **22**(7): p. 1622-7.

144. Chiappella, A., et al., *Integrating novel drugs to chemoimmunotherapy in diffuse large B-cell lymphoma.* Expert Rev Hematol, 2017. **10**(8): p. 697-705.

145. Zhang, J., L.J. Medeiros, and K.H. Young, *Cancer Immunotherapy in Diffuse Large B-Cell Lymphoma.* Front Oncol, 2018. **8**: p. 351.

146. Su, Y., et al., *Therapeutic strategy with artificially-designed i-lncRNA targeting multiple oncogenic microRNAs exhibits effective antitumor activity in diffuse large B-cell lymphoma.* Oncotarget, 2016. **7**(31): p. 49143-49155.

147. Cheng, C.J., et al., *MicroRNA silencing for cancer therapy targeted to the tumour microenvironment.* Nature, 2015. **518**(7537): p. 107-10.

9. References

148. Babar, I.A., et al., *Nanoparticle-based therapy in an in vivo microRNA-155 (miR-155)-dependent mouse model of lymphoma.* Proc Natl Acad Sci U S A, 2012. **109**(26): p. E1695-704.

149. Craig, V.J., et al., *Systemic microRNA-34a delivery induces apoptosis and abrogates growth of diffuse large B-cell lymphoma in vivo.* Leukemia, 2012. **26**(11): p. 2421-4.

150. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. **75**(5): p. 843-54.

151. Wightman, B., I. Ha, and G. Ruvkun, *Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans.* Cell, 1993. **75**(5): p. 855-62.

152. Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.* Nature, 2000. **403**(6772): p. 901-6.

153. Slack, F.J., et al., *The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor.* Mol Cell, 2000. **5**(4): p. 659-69.

154. Pasquinelli, A.E., et al., *Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA.* Nature, 2000. **408**(6808): p. 86-9.

155. Lee, R.C. and V. Ambros, *An extensive class of small RNAs in Caenorhabditis elegans.* Science, 2001. **294**(5543): p. 862-4.

156. Lau, N.C., et al., *An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans.* Science, 2001. **294**(5543): p. 858-62.

157. Lagos-Quintana, M., et al., *Identification of novel genes coding for small expressed RNAs.* Science, 2001. **294**(5543): p. 853-8.

158. Bhaskaran, M. and M. Mohan, *MicroRNAs: history, biogenesis, and their evolving role in animal development and disease.* Vet Pathol, 2014. **51**(4): p. 759-74.

159. Saini, H.K., S. Griffiths-Jones, and A.J. Enright, *Genomic analysis of human microRNA transcripts.* Proc Natl Acad Sci U S A, 2007. **104**(45): p. 17719-24.

160. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.

161. Yeom, K.H., et al., *Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing.* Nucleic Acids Res, 2006. **34**(16): p. 4622-9.

162. Lund, E., et al., *Nuclear export of microRNA precursors.* Science, 2004. **303**(5654): p. 95-8.

163. Bohnsack, M.T., K. Czaplinski, and D. Gorlich, *Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs.* RNA, 2004. **10**(2): p. 185-91.

164. Grishok, A., et al., *Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing.* Cell, 2001. **106**(1): p. 23-34.

165. Knight, S.W. and B.L. Bass, *A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans.* Science, 2001. **293**(5538): p. 2269-71.

166. Lee, Y., et al., *The role of PACT in the RNA silencing pathway.* EMBO J, 2006. **25**(3): p. 522-32.

167. Okamura, K., N. Liu, and E.C. Lai, *Distinct mechanisms for microRNA strand selection by Drosophila Argonautes.* Mol Cell, 2009. **36**(3): p. 431-44.

168. Meister, G., *Argonaute proteins: functional insights and emerging roles.* Nat Rev Genet, 2013. **14**(7): p. 447-59.

169. Liu, J., et al., *Argonaute2 is the catalytic engine of mammalian RNAi.* Science, 2004. **305**(5689): p. 1437-41.

170. Pratt, A.J. and I.J. MacRae, *The RNA-induced silencing complex: a versatile gene-silencing machine.* J Biol Chem, 2009. **284**(27): p. 17897-901.

171. Zhang, J., et al., *Oncogenic role of microRNA-532-5p in human colorectal cancer via targeting of the 5'UTR of RUNX3.* Oncol Lett, 2018. **15**(5): p. 7215-7220.

172. Akhtar, N., M.S. Makki, and T.M. Haqqi, *MicroRNA-602 and microRNA-608 regulate sonic hedgehog expression via target sites in the coding region in human chondrocytes.* Arthritis Rheumatol, 2015. **67**(2): p. 423-34.

173. Lawrie, C.H., *MicroRNAs and haematology: small molecules, big function.* Br J Haematol, 2007. **137**(6): p. 503-12.

174. Ruby, J.G., C.H. Jan, and D.P. Bartel, *Intronic microRNA precursors that bypass Drosha processing.* Nature, 2007. **448**(7149): p. 83-6.

175. Westholm, J.O. and E.C. Lai, *Mirtrons: microRNA biogenesis via splicing.* Biochimie, 2011. **93**(11): p. 1897-904.

176. Daugaard, I. and T.B. Hansen, *Biogenesis and Function of Ago-Associated RNAs.* Trends Genet, 2017. **33**(3): p. 208-219.

177. Yang, J.S., et al., *Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis.* Proc Natl Acad Sci U S A, 2010. **107**(34): p. 15163-8.

178. Hansen, T.B., et al., *Argonaute-associated short introns are a novel class of gene regulators.* Nat Commun, 2016. **7**: p. 11538.

179. Bartel, D.P., *MicroRNAs: target recognition and regulatory functions.* Cell, 2009. **136**(2): p. 215-33.

180. Chekulaeva, M. and W. Filipowicz, *Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells.* Curr Opin Cell Biol, 2009. **21**(3): p. 452-60.

181. Guo, H., et al., *Mammalian microRNAs predominantly act to decrease target mRNA levels.* Nature, 2010. **466**(7308): p. 835-40.

182. Eichhorn, S.W., et al., *mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues.* Mol Cell, 2014. **56**(1): p. 104-15.

183. Djuranovic, S., A. Nahvi, and R. Green, *miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay.* Science, 2012. **336**(6078): p. 237-40.

184. Yang, J.S. and E.C. Lai, *Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants.* Mol Cell, 2011. **43**(6): p. 892-903.

185. Baek, D., et al., *The impact of microRNAs on protein output.* Nature, 2008. **455**(7209): p. 64-71.

186. Dragomir, M., et al., *Using microRNA Networks to Understand Cancer.* Int J Mol Sci, 2018. **19**(7).

187. Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.* Cell, 2005. **120**(1): p. 15-20.

188. Lewis, B.P., et al., *Prediction of mammalian microRNA targets.* Cell, 2003. **115**(7): p. 787-98.

189. Lal, A., et al., *miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements.* Mol Cell, 2009. **35**(5): p. 610-25.

190. Flamand, M.N., et al., *A non-canonical site reveals the cooperative mechanisms of microRNA-mediated silencing.* Nucleic Acids Res, 2017. **45**(12): p. 7212-7225.

191. Helwak, A., et al., *Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding.* Cell, 2013. **153**(3): p. 654-65.

192. Ni, W.J. and X.M. Leng, *miRNA-Dependent Activation of mRNA Translation.* Microrna, 2016. **5**(2): p. 83-86.

9. References

193. Yoon, J.H., K. Abdelmohsen, and M. Gorospe, *Functional interactions among microRNAs and long noncoding RNAs.* Semin Cell Dev Biol, 2014. **34**: p. 9-14.

194. Majid, S., et al., *MicroRNA-205-directed transcriptional activation of tumor suppressor genes in prostate cancer.* Cancer, 2010. **116**(24): p. 5637-49.

195. Seviour, E.G., et al., *Functional proteomics identifies miRNAs to target a p27/Myc/phospho-Rb signature in breast and ovarian cancer.* Oncogene, 2016. **35**(6): p. 691-701.

196. Place, R.F., et al., *MicroRNA-373 induces expression of genes with complementary promoter sequences.* Proc Natl Acad Sci U S A, 2008. **105**(5): p. 1608-13.

197. Korla, K., P. Arrigo, and C.K. Mitra, *Promoters, toll like receptors and microRNAs: a strange association.* Indian J Biochem Biophys, 2013. **50**(3): p. 169-76.

198. Eiring, A.M., et al., *miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts.* Cell, 2010. **140**(5): p. 652-65.

199. Balkhi, M.Y., et al., *miR-29 acts as a decoy in sarcomas to protect the tumor suppressor A20 mRNA from degradation by HuR.* Sci Signal, 2013. **6**(286): p. ra63.

200. Wang, Y., et al., *Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation.* Nat Genet, 2008. **40**(12): p. 1478-83.

201. Melton, C. and R. Blelloch, *MicroRNA Regulation of Embryonic Stem Cell Self-Renewal and Differentiation.* Adv Exp Med Biol, 2010. **695**: p. 105-17.

202. Zhao, Y., E. Samal, and D. Srivastava, *Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis.* Nature, 2005. **436**(7048): p. 214-20.

203. Robertus, J.L., et al., *MiRNA profiling in B non-Hodgkin lymphoma: a MYC-related miRNA profile characterizes Burkitt lymphoma.* Br J Haematol, 2010. **149**(6): p. 896-9.

204. Lawrie, C.H., et al., *Noncoding RNA Expression and Targeted Next-Generation Sequencing Distinguish Tubulocystic Renal Cell Carcinoma (TC-RCC) from Other Renal Neoplasms.* J Mol Diagn, 2018. **20**(1): p. 34-45.

205. Bockmeyer, C.L., et al., *MicroRNA profiles of healthy basal and luminal mammary epithelial cells are distinct and reflected in different breast cancer subtypes.* Breast Cancer Res Treat, 2011. **130**(3): p. 735-45.

206. Chen, E., et al., *Small but Heavy Role: MicroRNAs in Hepatocellular Carcinoma Progression.* Biomed Res Int, 2018. **2018**: p. 6784607.

207. Hu, W., et al., *Functional miRNAs in breast cancer drug resistance.* Onco Targets Ther, 2018. **11**: p. 1529-1541.

208. Fattore, L., et al., *MicroRNAs in melanoma development and resistance to target therapy.* Oncotarget, 2017. **8**(13): p. 22262-22278.

209. Xiao, C., et al., *Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes.* Nat Immunol, 2008. **9**(4): p. 405-14.

210. le Sage, C., et al., *Regulation of the p27(Kip1) tumor suppressor by miR-221 and miR-222 promotes cancer cell proliferation.* Embo j, 2007. **26**(15): p. 3699-708.

211. Sarkar, S., et al., *Down-regulation of miR-221 inhibits proliferation of pancreatic cancer cells through up-regulation of PTEN, p27(kip1), p57(kip2), and PUMA.* Am J Cancer Res, 2013. **3**(5): p. 465-77.

212. Tsukerman, P., J. Enk, and O. Mandelboim, *Metastamir-mediated immune evasion: miR-10b downregulates the stress-induced molecule MICB, hence avoid recognition by NKG2D receptor.* Oncoimmunology, 2013. **2**(1): p. e22245.

213. Ma, L., J. Teruya-Feldstein, and R.A. Weinberg, *Tumour invasion and metastasis initiated by microRNA-10b in breast cancer.* Nature, 2007. **449**(7163): p. 682-8.

214. Dang, K. and K.A. Myers, *The role of hypoxia-induced miR-210 in cancer progression.* Int J Mol Sci, 2015. **16**(3): p. 6353-72.

215. Sasahira, T., et al., *Downregulation of miR-126 induces angiogenesis and lymphangiogenesis by activation of VEGF-A in oral cancer.* Br J Cancer, 2012. **107**(4): p. 700-6.

216. Wurdinger, T., et al., *miR-296 regulates growth factor receptor overexpression in angiogenic endothelial cells.* Cancer Cell, 2008. **14**(5): p. 382-93.

217. Bommer, G.T., et al., *p53-mediated activation of miRNA34 candidate tumor-suppressor genes.* Curr Biol, 2007. **17**(15): p. 1298-307.

218. Chang, T.C., et al., *Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis.* Mol Cell, 2007. **26**(5): p. 745-52.

219. Hashemi, Z.S., et al., *Inhibition of breast cancer metastasis by co-transfection of miR-31/193b-mimics.* Iran J Basic Med Sci, 2018. **21**(4): p. 427-433.

220. Ventura, A., et al., *Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters.* Cell, 2008. **132**(5): p. 875-86.

221. Chen, C.Z., et al., *MicroRNAs modulate hematopoietic lineage differentiation.* Science, 2004. **303**(5654): p. 83-6.

222. Lwin, T., et al., *Follicular dendritic cell-dependent drug resistance of non-Hodgkin lymphoma involves cell adhesion-mediated Bim down-regulation through induction of microRNA-181a.* Blood, 2010. **116**(24): p. 5228-36.

223. Mehta, A., et al., *The microRNA-212/132 cluster regulates B cell development by targeting Sox4.* J Exp Med, 2015. **212**(10): p. 1679-92.

224. Hu, H., et al., *Foxp1 is an essential transcriptional regulator of B cell development.* Nat Immunol, 2006. **7**(8): p. 819-26.

225. Rao, D.S., et al., *MicroRNA-34a perturbs B lymphocyte development by repressing the forkhead box transcription factor Foxp1.* Immunity, 2010. **33**(1): p. 48-59.

226. Xiao, C., et al., *MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb.* Cell, 2007. **131**(1): p. 146-59.

227. Chaudhuri, A.A., et al., *Oncomir miR-125b regulates hematopoiesis by targeting the gene Lin28A.* Proc Natl Acad Sci U S A, 2012. **109**(11): p. 4233-8.

228. Vigorito, E., et al., *microRNA-155 regulates the generation of immunoglobulin class-switched plasma cells.* Immunity, 2007. **27**(6): p. 847-59.

229. Frasca, D., et al., *MicroRNAs miR-155 and miR-16 Decrease AID and E47 in B Cells from Elderly Individuals.* J Immunol, 2015. **195**(5): p. 2134-40.

230. de Yebenes, V.G., et al., *miR-181b negatively regulates activation-induced cytidine deaminase in B cells.* J Exp Med, 2008. **205**(10): p. 2199-206.

231. Zhang, J., et al., *Patterns of microRNA expression characterize stages of human B-cell differentiation.* Blood, 2009. **113**(19): p. 4586-94.

232. Gururajan, M., et al., *MicroRNA 125b inhibition of B cell differentiation in germinal centers.* Int Immunol, 2010. **22**(7): p. 583-92.

233. Xu, S., et al., *Mir-17-92 regulates bone marrow homing of plasma cells and production of immunoglobulin G2c.* Nat Commun, 2015. **6**: p. 6764.

234. Minges Wols, H.A., et al., *The role of bone marrow-derived stromal cells in the maintenance of plasma cell longevity.* J Immunol, 2002. **169**(8): p. 4213-21.

235. Coffre, M. and S.B. Koralov, *miRNAs in B Cell Development and Lymphomagenesis.* Trends Mol Med, 2017. **23**(8): p. 721-736.

236. Lawrie, C.H., et al., *Expression of microRNAs in diffuse large B cell lymphoma is associated with immunophenotype, survival and transformation from follicular lymphoma.* J Cell Mol Med, 2009. **13**(7): p. 1248-60.

237. Calin, G.A., et al., *Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia.* Proc Natl Acad Sci U S A, 2002. **99**(24): p. 15524-9.

238. He, L., et al., *A microRNA polycistron as a potential human oncogene.* Nature, 2005. **435**(7043): p. 828-33.

9. References

239. Craig, V.J., et al., *Myc-mediated repression of microRNA-34a promotes high-grade transformation of B-cell lymphoma by dysregulation of FoxP1.* Blood, 2011. **117**(23): p. 6227-36.

240. Boysen, J., et al., *The tumor suppressor axis p53/miR-34a regulates Axl expression in B-cell chronic lymphocytic leukemia: implications for therapy in p53-defective CLL patients.* Leukemia, 2014. **28**(2): p. 451-5.

241. Marques, S.C., et al., *High miR-34a expression improves response to doxorubicin in diffuse large B-cell lymphoma.* Exp Hematol, 2016. **44**(4): p. 238-46 e2.

242. Mraz, M., et al., *miR-150 influences B-cell receptor signaling in chronic lymphocytic leukemia by regulating expression of GAB1 and FOXP1.* Blood, 2014. **124**(1): p. 84-95.

243. Costinean, S., et al., *Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice.* Proc Natl Acad Sci U S A, 2006. **103**(18): p. 7024-9.

244. Jiang, D. and R.C. Aguiar, *MicroRNA-155 controls RB phosphorylation in normal and malignant B lymphocytes via the noncanonical TGF-beta1/SMAD5 signaling module.* Blood, 2014. **123**(1): p. 86-93.

245. Liu, Q., et al., *The inositol polyphosphate 5-phosphatase ship is a crucial negative regulator of B cell antigen receptor signaling.* J Exp Med, 1998. **188**(7): p. 1333-42.

246. Li, X.D., et al., *MiR-155 regulates lymphoma cell proliferation and apoptosis through targeting SOCS3/JAK-STAT3 signaling pathway.* Eur Rev Med Pharmacol Sci, 2017. **21**(22): p. 5153-5159.

247. Rossi, S., et al., *microRNA fingerprinting of CLL patients with chromosome 17p deletion identify a miR-21 score that stratifies early survival.* Blood, 2010. **116**(6): p. 945-52.

248. Li, J., et al., *miR-21 expression predicts prognosis in diffuse large B-cell lymphoma.* Int J Clin Exp Pathol, 2015. **8**(11): p. 15019-24.

249. Bai, H., et al., *MicroRNA-21 regulates the sensitivity of diffuse large B-cell lymphoma cells to the CHOP chemotherapy regimen.* Int J Hematol, 2013. **97**(2): p. 223-31.

250. Li, G., et al., *Epigenetic silencing of miR-125b is required for normal B-cell development.* Blood, 2018. **131**(17): p. 1920-1930.

251. So, A.Y., et al., *Dual mechanisms by which miR-125b represses IRF4 to induce myeloid and B-cell leukemias.* Blood, 2014. **124**(9): p. 1502-12.

252. Sole, C., et al., *miRNAs in B-cell lymphoma: Molecular mechanisms and biomarker potential.* Cancer Lett, 2017. **405**: p. 79-89.

253. Croce, C.M., *Causes and consequences of microRNA dysregulation in cancer.* Nat Rev Genet, 2009. **10**(10): p. 704-14.

254. O'Donnell, K.A., et al., *c-Myc-regulated microRNAs modulate E2F1 expression.* Nature, 2005. **435**(7043): p. 839-43.

255. Boerma, E.G., et al., *Translocations involving 8q24 in Burkitt lymphoma and other malignant lymphomas: a historical review of cytogenetics in the light of todays knowledge.* Leukemia, 2009. **23**(2): p. 225-34.

256. Guo, Z., et al., *Identification of diagnostic and prognostic biomarkers for cancer: Focusing on genetic variations in microRNA regulatory pathways (Review).* Mol Med Rep, 2016. **13**(3): p. 1943-52.

257. Sung, H., et al., *Common genetic polymorphisms of microRNA biogenesis pathway genes and risk of breast cancer: a case-control study in Korea.* Breast Cancer Res Treat, 2011. **130**(3): p. 939-51.

258. Sung, H., et al., *Common genetic polymorphisms of microRNA biogenesis pathway genes and breast cancer survival.* BMC Cancer, 2012. **12**: p. 195.

259. Guo, Z., et al., *A microRNA-related single nucleotide polymorphism of the XPO5 gene is associated with survival of small cell lung cancer patients.* Biomed Rep, 2013. **1**(4): p. 545-548.

260. Mishra, P.J. and J.R. Bertino, *MicroRNA polymorphisms: the future of pharmacogenomics, molecular epidemiology and individualized medicine.* Pharmacogenomics, 2009. **10**(3): p. 399-416.

261. Cipollini, M., S. Landi, and F. Gemignani, *MicroRNA binding site polymorphisms as biomarkers in cancer management and research.* Pharmgenomics Pers Med, 2014. **7**: p. 173-91.

262. Friedman, R.C., et al., *Most mammalian mRNAs are conserved targets of microRNAs.* Genome Res, 2009. **19**(1): p. 92-105.

263. Agarwal, V., et al., *Predicting effective microRNA target sites in mammalian mRNAs.* Elife, 2015. **4**.

264. Maragkakis, M., et al., *DIANA-microT web server: elucidating microRNA functions through target prediction.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W273-6.

265. Paraskevopoulou, M.D., et al., *DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W169-73.

266. Enright, A.J., et al., *MicroRNA targets in Drosophila.* Genome Biol, 2003. **5**(1): p. R1.

267. Kertesz, M., et al., *The role of site accessibility in microRNA target recognition.* Nat Genet, 2007. **39**(10): p. 1278-84.

268. Grun, D., et al., *microRNA target predictions across seven Drosophila species and comparison to mammalian targets.* PLoS Comput Biol, 2005. **1**(1): p. e13.

269. Miranda, K.C., et al., *A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes.* Cell, 2006. **126**(6): p. 1203-17.

270. Rehmsmeier, M., et al., *Fast and effective prediction of microRNA/target duplexes.* Rna, 2004. **10**(10): p. 1507-17.

271. Riffo-Campos, A.L., I. Riquelme, and P. Brebi-Mieville, *Tools for Sequence-Based miRNA Target Prediction: What to Choose?* Int J Mol Sci, 2016. **17**(12).

272. Mourelatos, Z., *Small RNAs: The seeds of silence.* Nature, 2008. **455**(7209): p. 44-5.

273. Chi, S.W., G.J. Hannon, and R.B. Darnell, *An alternative mode of microRNA target recognition.* Nat Struct Mol Biol, 2012. **19**(3): p. 321-7.

274. Loeb, G.B., et al., *Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting.* Mol Cell, 2012. **48**(5): p. 760-70.

275. Moore, M.J., et al., *miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity.* Nat Commun, 2015. **6**: p. 8864.

276. Xia, Z., et al., *Molecular dynamics simulations of Ago silencing complexes reveal a large repertoire of admissible 'seed-less' targets.* Sci Rep, 2012. **2**: p. 569.

277. Vlachos, I.S., et al., *DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions.* Nucleic Acids Res, 2015. **43**(Database issue): p. D153-9.

278. Seok, H., et al., *MicroRNA Target Recognition: Insights from Transcriptome-Wide Non-Canonical Interactions.* Mol Cells, 2016. **39**(5): p. 375-81.

279. Mathews, D.H., et al., *Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.* J Mol Biol, 1999. **288**(5): p. 911-40.

280. Lorenz, R., et al., *ViennaRNA Package 2.0.* Algorithms Mol Biol, 2011. **6**: p. 26.

281. Schnall-Levin, M., et al., *Unusually effective microRNA targeting within repeat-rich coding regions of mammalian mRNAs.* Genome Res, 2011. **21**(9): p. 1395-403.

282. Grimson, A., et al., *MicroRNA targeting specificity in mammals: determinants beyond seed pairing.* Mol Cell, 2007. **27**(1): p. 91-105.

283. Nam, J.W., et al., *Global analyses of the effect of different cellular contexts on microRNA targeting.* Mol Cell, 2014. **53**(6): p. 1031-1043.

284. Hsu, J.B., et al., *miRTar: an integrated system for identifying miRNA-target interactions in human.* BMC Bioinformatics, 2011. **12**: p. 300.

9. References

285. Xu, W., et al., *Identifying microRNA targets in different gene regions.* BMC Bioinformatics, 2014. **15 Suppl 7**: p. S4.

286. Clark, P.M., et al., *Argonaute CLIP-Seq reveals miRNA targetome diversity across tissue types.* Sci Rep, 2014. **4**: p. 5947.

287. Lossner, C., et al., *Quantitative proteomics identify novel miR-155 target proteins.* PLoS One, 2011. **6**(7): p. e22146.

288. Xu, G., et al., *Transcriptome and targetome analysis in MIR155 expressing cells using RNA-seq.* Rna, 2010. **16**(8): p. 1610-22.

289. Tenenbaum, S.A., et al., *Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays.* Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14085-90.

290. Zhao, J., et al., *Genome-wide identification of polycomb-associated RNAs by RIP-seq.* Mol Cell, 2010. **40**(6): p. 939-53.

291. McHugh, C.A., P. Russell, and M. Guttman, *Methods for comprehensive experimental identification of RNA-protein interactions.* Genome Biol, 2014. **15**(1): p. 203.

292. Ascano, M., et al., *Identification of RNA-protein interaction networks using PAR-CLIP.* Wiley Interdiscip Rev RNA, 2012. **3**(2): p. 159-77.

293. Milek, M., E. Wyler, and M. Landthaler, *Transcriptome-wide analysis of protein-RNA interactions using high-throughput sequencing.* Semin Cell Dev Biol, 2012. **23**(2): p. 206-12.

294. Lee, F.C.Y. and J. Ule, *Advances in CLIP Technologies for Studies of Protein-RNA Interactions.* Mol Cell, 2018. **69**(3): p. 354-369.

295. Ule, J., et al., *CLIP identifies Nova-regulated RNA networks in the brain.* Science, 2003. **302**(5648): p. 1212-5.

296. Hafner, M., et al., *Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP.* Cell, 2010. **141**(1): p. 129-41.

297. Mittal, N. and M. Zavolan, *Seq and CLIP through the miRNA world.* Genome Biol, 2014. **15**(1): p. 202.

298. Chi, S.W., et al., *Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps.* Nature, 2009. **460**(7254): p. 479-86.

299. Farazi, T.A., et al., *Identification of distinct miRNA target regulation between breast cancer molecular subtypes using AGO2-PAR-CLIP and patient datasets.* Genome Biol, 2014. **15**(1): p. R9.

300. Wang, T., et al., *A model-based approach to identify binding sites in CLIP-Seq data.* PLoS One, 2014. **9**(4): p. e93248.

301. Ule, J., et al., *CLIP: a method for identifying protein-RNA interaction sites in living cells.* Methods, 2005. **37**(4): p. 376-86.

302. Sanford, J.R., et al., *Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF.* PLoS One, 2008. **3**(10): p. e3369.

303. Licatalosi, D.D., et al., *HITS-CLIP yields genome-wide insights into brain alternative RNA processing.* Nature, 2008. **456**(7221): p. 464-9.

304. Yeo, G.W., et al., *An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells.* Nat Struct Mol Biol, 2009. **16**(2): p. 130-7.

305. Granneman, S., et al., *Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs.* Proc Natl Acad Sci U S A, 2009. **106**(24): p. 9613-8.

306. Hafner, M., et al., *PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins.* J Vis Exp, 2010(41).

307. Konig, J., et al., *iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution.* J Vis Exp, 2011(50).

308. Wang, Z., et al., *iCLIP predicts the dual splicing effects of TIA-RNA interactions.* PLoS Biol, 2010. **8**(10): p. e1000530.

309. Huppertz, I., et al., *iCLIP: protein-RNA interactions at nucleotide resolution.* Methods, 2014. **65**(3): p. 274-87.

310. Weyn-Vanhentenryck, S.M., et al., *HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism.* Cell Rep, 2014. **6**(6): p. 1139-1152.

311. Flynn, R.A., et al., *Dissecting noncoding and pathogen RNA-protein interactomes.* RNA, 2015. **21**(1): p. 135-43.

312. Zarnegar, B.J., et al., *irCLIP platform for efficient characterization of protein-RNA interactions.* Nat Methods, 2016. **13**(6): p. 489-92.

313. Van Nostrand, E.L., et al., *Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP).* Nat Methods, 2016. **13**(6): p. 508-14.

314. Van Nostrand, E.L., et al., *Variation in single-nucleotide sensitivity of eCLIP derived from reverse transcription conditions.* Methods, 2017. **126**: p. 29-37.

315. Aktas, T., et al., *DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome.* Nature, 2017. **544**(7648): p. 115-119.

316. Brugiolo, M., et al., *Fractionation iCLIP detects persistent SR protein binding to conserved, retained introns in chromatin, nucleoplasm and cytoplasm.* Nucleic Acids Res, 2017. **45**(18): p. 10452-10465.

317. Kargapolova, Y., et al., *sCLIP-an integrated platform to study RNA-protein interactomes in biomedical research: identification of CSTF2tau in alternative processing of small nuclear RNAs.* Nucleic Acids Res, 2017. **45**(10): p. 6074-6086.

318. Rosenberg, M., et al., *Denaturing CLIP, dCLIP, Pipeline Identifies Discrete RNA Footprints on Chromatin-Associated Proteins and Reveals that CBX7 Targets 3' UTRs to Regulate mRNA Expression.* Cell Syst, 2017. **5**(4): p. 368-385 e15.

319. Kudla, G., et al., *Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast.* Proc Natl Acad Sci U S A, 2011. **108**(24): p. 10010-5.

320. Sugimoto, Y., et al., *hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1.* Nature, 2015. **519**(7544): p. 491-4.

321. Hwang, H.W., et al., *PAPERCLIP Identifies MicroRNA Targets and a Role of CstF64/64tau in Promoting Non-canonical poly(A) Site Usage.* Cell Rep, 2016. **15**(2): p. 423-35.

322. Hwang, H.W., et al., *cTag-PAPERCLIP Reveals Alternative Polyadenylation Promotes Cell-Type Specific Protein Diversity and Shifts Araf Isoforms with Microglia Activation.* Neuron, 2017. **95**(6): p. 1334-1349 e5.

323. Hussain, S., et al., *NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs.* Cell Rep, 2013. **4**(2): p. 255-61.

324. Linder, B., et al., *Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome.* Nat Methods, 2015. **12**(8): p. 767-72.

325. Stefani, G., et al., *A novel mechanism of LIN-28 regulation of let-7 microRNA expression revealed by in vivo HITS-CLIP in C. elegans.* RNA, 2015. **21**(5): p. 985-96.

326. Spengler, R.M., et al., *Elucidation of transcriptome-wide microRNA binding sites in human cardiac tissues by Ago2 HITS-CLIP.* Nucleic Acids Res, 2016. **44**(15): p. 7120-31.

327. Wang, W., et al., *MicroRNA profiling of follicular lymphoma identifies microRNAs related to cell proliferation and tumor response.* Haematologica, 2012. **97**(4): p. 586-94.

328. Gebauer, N., et al., *MicroRNA signatures in subtypes of follicular lymphoma.* Anticancer Res, 2014. **34**(5): p. 2105-11.

329. Malpeli, G., et al., *MicroRNA signatures and Foxp3(+) cell count correlate with relapse occurrence in follicular lymphoma.* Oncotarget, 2018. **9**(28): p. 19961-19979.

330. Thompson, M.A., et al., *miR-31 and miR-17-5p levels change during transformation of follicular lymphoma.* Hum Pathol, 2016. **50**: p. 118-26.

331. Hezaveh, K., et al., *Alterations of microRNA and microRNA-regulated messenger RNA expression in germinal center B-cell lymphomas determined by integrative sequencing analysis.* Haematologica, 2016. **101**(11): p. 1380-1389.

9. References

332. Lopes-Ramos, C.M., et al., *E2F1 somatic mutation within miRNA target site impairs gene regulation in colorectal cancer.* PLoS One, 2017. **12**(7): p. e0181153.

333. Amato, F., et al., *Gene mutation in microRNA target sites of CFTR gene: a novel pathogenetic mechanism in cystic fibrosis?* PLoS One, 2013. **8**(3): p. e60448.

334. Morales, S., et al., *Genetic Variants in pre-miR-146a, pre-miR-499, pre-miR-125a, pre-miR-605, and pri-miR-182 Are Associated with Breast Cancer Susceptibility in a South American Population.* Genes (Basel), 2018. **9**(9).

335. Dai, Z.J., et al., *Five common functional polymorphisms in microRNAs (rs2910164, rs2292832, rs11614913, rs3746444, rs895819) and the susceptibility to breast cancer: evidence from 8361 cancer cases and 8504 controls.* Curr Pharm Des, 2015. **21**(11): p. 1455-63.

336. Brewster, B.L., et al., *Identification of fifteen novel germline variants in the BRCA1 3'UTR reveals a variant in a breast cancer case that introduces a functional miR-103 target site.* Hum Mutat, 2012. **33**(12): p. 1665-75.

337. Li, Y., et al., *G-A variant in miR-200c binding site of EFNA1 alters susceptibility to gastric cancer.* Mol Carcinog, 2014. **53**(3): p. 219-29.

338. Du, Y., et al., *Association of miRNA-122-binding site polymorphism at the interleukin-1 alpha gene and its interaction with hepatitis B virus mutations with hepatocellular carcinoma risk.* Front Med, 2014. **8**(2): p. 217-26.

339. Peckham-Gregory, E.C., et al., *MicroRNA-related polymorphisms and non-Hodgkin lymphoma susceptibility in the Multicenter AIDS Cohort Study.* Cancer Epidemiol, 2016. **45**: p. 47-57.

340. Yang, B., et al., *A polymorphism at the microRNA binding site in the 3' untranslated region of C14orf101 is associated with non-Hodgkin lymphoma overall survival.* Cancer Genet, 2014. **207**(4): p. 141-6.

341. Xie, Y., et al., *A miR-SNP of the KRT81 gene is associated with the prognosis of non-Hodgkin's lymphoma.* Gene, 2014. **539**(2): p. 198-202.

342. Ziebarth, J.D., A. Bhattacharya, and Y. Cui, *Integrative analysis of somatic mutations altering microRNA targeting in cancer genomes.* PLoS One, 2012. **7**(10): p. e47137.

343. Bhattacharya, A., J.D. Ziebarth, and Y. Cui, *SomamiR: a database for somatic mutations impacting microRNA function in cancer.* Nucleic Acids Res, 2013. **41**(Database issue): p. D977-82.

344. Kozomara, A. and S. Griffiths-Jones, *miRBase: integrating microRNA annotation and deep-sequencing data.* Nucleic Acids Res, 2011. **39**(Database issue): p. D152-7.

345. Kinsella, R.J., et al., *Ensembl BioMarts: a hub for data retrieval across taxonomic space.* Database (Oxford), 2011. **2011**: p. bar030.

346. Kramer, A., et al., *Causal analysis approaches in Ingenuity Pathway Analysis.* Bioinformatics, 2014. **30**(4): p. 523-30.

347. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res, 2001. **29**(1): p. 308-11.

348. Kamioka, M., et al., *Testican 3 expression in adult T-cell leukemia.* Leuk Res, 2009. **33**(7): p. 913-8.

349. Huang, B.B., et al., *Down-regulation of SENP1 expression increases apoptosis of Burkitt lymphoma cells.* Asian Pac J Cancer Prev, 2012. **13**(5): p. 2045-9.

350. Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets--update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D991-5.

351. Tsiatis, A.C., et al., *Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications.* J Mol Diagn, 2010. **12**(4): p. 425-32.

352. Ying, C.Y., et al., *MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma.* Nat Immunol, 2013. **14**(10): p. 1084-92.

353. Fujita, S., et al., *Single nucleotide variant sequencing errors in whole exome sequencing using the Ion Proton System.* Biomed Rep, 2017. **7**(1): p. 17-20.

354. Laehnemann, D., A. Borkhardt, and A.C. McHardy, *Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction.* Brief Bioinform, 2016. **17**(1): p. 154-79.

355. Rothberg, J.M., et al., *An integrated semiconductor device enabling non-optical genome sequencing.* Nature, 2011. **475**(7356): p. 348-52.

356. Bottini, S., et al., *From benchmarking HITS-CLIP peak detection programs to a new method for identification of miRNA-binding sites from Ago2-CLIP data.* Nucleic Acids Res, 2017. **45**(9): p. e71.

357. Ghosal, S., et al., *miRepress: modelling gene expression regulation by microRNA with non-conventional binding sites.* Sci Rep, 2016. **6**: p. 22334.

358. Sahasrabuddhe, A.A., et al., *Oncogenic Y641 mutations in EZH2 prevent Jak2/beta-TrCP-mediated degradation.* Oncogene, 2015. **34**(4): p. 445-54.

359. Guo, Y., et al., *miR-144 downregulation increases bladder cancer cell proliferation by targeting EZH2 and regulating Wnt signaling.* FEBS J, 2013. **280**(18): p. 4531-8.

360. Cao, J., et al., *TUG1 promotes osteosarcoma tumorigenesis by upregulating EZH2 expression via miR-144-3p.* Int J Oncol, 2017. **51**(4): p. 1115-1123.

361. Lin, L., et al., *MicroRNA-144 suppresses tumorigenesis and tumor progression of astrocytoma by targeting EZH2.* Hum Pathol, 2015. **46**(7): p. 971-80.

362. Wang, H., et al., *A Critical Role of miR-144 in Diffuse Large B-cell Lymphoma Proliferation and Invasion.* Cancer Immunol Res, 2016. **4**(4): p. 337-44.

363. Leich, E., et al., *Similar clinical features in follicular lymphomas with and without breaks in the BCL2 locus.* Leukemia, 2016. **30**(4): p. 854-60.

364. Albinger-Hegyi, A., et al., *High frequency of t(14;18)-translocation breakpoints outside of major breakpoint and minor cluster regions in follicular lymphomas: improved polymerase chain reaction protocols for their detection.* Am J Pathol, 2002. **160**(3): p. 823-32.

365. Weinberg, O.K., et al., *''Minor'' BCL2 breakpoints in follicular lymphoma: frequency and correlation with grade and disease presentation in 236 cases.* J Mol Diagn, 2007. **9**(4): p. 530-7.

366. Spitzer, J., et al., *PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins.* Methods Enzymol, 2014. **539**: p. 113-61.

367. Maatz, H., et al., *Transcriptome-wide Identification of RNA-binding Protein Binding Sites Using Photoactivatable-Ribonucleoside-Enhanced Crosslinking Immunoprecipitation (PAR-CLIP).* Curr Protoc Mol Biol, 2017. **118**: p. 27 6 1-27 6 19.

368. Jin, H.Y., et al., *MicroRNA-17~92 plays a causative role in lymphomagenesis by coordinating multiple oncogenic pathways.* EMBO J, 2013. **32**(17): p. 2377-91.

369. Gottwein, E., et al., *Viral microRNA targetome of KSHV-infected primary effusion lymphoma cell lines.* Cell Host Microbe, 2011. **10**(5): p. 515-26.

370. Lawrie, C.H., et al., *MicroRNA expression distinguishes between germinal center B cell-like and activated B cell-like subtypes of diffuse large B cell lymphoma.* Int J Cancer, 2007. **121**(5): p. 1156-61.

371. Iqbal, J., et al., *Global microRNA expression profiling uncovers molecular markers for classification and prognosis in aggressive B-cell lymphoma.* Blood, 2015. **125**(7): p. 1137-45.

372. Carvalho, B., et al., *Exploration, normalization, and genotype calls of high-density oligonucleotide SNP array data.* Biostatistics, 2007. **8**(2): p. 485-99.

373. Comoglio, F., C. Sievers, and R. Paro, *Sensitive and highly resolved identification of RNA-protein interaction sites in PAR-CLIP data.* BMC Bioinformatics, 2015. **16**: p. 32.

9. References

374. Golumbeanu, M., P. Mohammadi, and N. Beerenwinkel, *BMix: probabilistic modeling of occurring substitutions in PAR-CLIP data.* Bioinformatics, 2016. **32**(7): p. 976-83.

375. Wang, T., Y. Xie, and G. Xiao, *dCLIP: a computational approach for comparative CLIP-seq analyses.* Genome Biol, 2014. **15**(1): p. R11.

376. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.

377. Robinson, J.T., et al., *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.

378. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* OMICS, 2012. **16**(5): p. 284-7.

379. Hsu, S.D., et al., *miRTarBase: a database curates experimentally validated microRNA-target interactions.* Nucleic Acids Res, 2011. **39**(Database issue): p. D163-9.

380. Sethupathy, P., M. Megraw, and A.G. Hatzigeorgiou, *A guide through present computational approaches for the identification of mammalian microRNA targets.* Nat Methods, 2006. **3**(11): p. 881-6.

381. Johnson, S.M., et al., *RAS is regulated by the let-7 microRNA family.* Cell, 2005. **120**(5): p. 635-47.

382. Huang, X., et al., *Quantitative proteomics reveals that miR-155 regulates the PI3K-AKT pathway in diffuse large B-cell lymphoma.* Am J Pathol, 2012. **181**(1): p. 26-33.

383. Hafner, M., et al., *Genome-wide identification of miRNA targets by PAR-CLIP.* Methods, 2012. **58**(2): p. 94-105.

384. Ye, Z., H. Jin, and Q. Qian, *Argonaute 2: A Novel Rising Star in Cancer Research.* J Cancer, 2015. **6**(9): p. 877-82.

385. Burger, K., et al., *4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response.* RNA Biol, 2013. **10**(10): p. 1623-30.

386. Wheeler, E.C., E.L. Van Nostrand, and G.W. Yeo, *Advances and challenges in the detection of transcriptome-wide protein-RNA interactions.* Wiley Interdiscip Rev RNA, 2018. **9**(1).

387. Wang, T., et al., *Design and bioinformatics analysis of genome-wide CLIP experiments.* Nucleic Acids Res, 2015. **43**(11): p. 5263-74.

388. Paraskevopoulou, M.D., et al., *microCLIP super learning framework uncovers functional transcriptome-wide miRNA interactions.* Nat Commun, 2018. **9**(1): p. 3601.

389. Mihaylova, M.M. and R.J. Shaw, *The AMPK signalling pathway coordinates cell growth, autophagy and metabolism.* Nat Cell Biol, 2011. **13**(9): p. 1016-23.

390. Shi, W.Y., et al., *Therapeutic metformin/AMPK activation blocked lymphoma cell growth via inhibition of mTOR pathway and induction of autophagy.* Cell Death Dis, 2012. **3**: p. e275.

391. Riva, B., et al., *Celecoxib inhibits proliferation and survival of chronic myelogeous leukemia (CML) cells via AMPK-dependent regulation of beta-catenin and mTORC1/2.* Oncotarget, 2016. **7**(49): p. 81555-81570.

392. Hoffman, A.E., et al., *Association of AMP-activated protein kinase with risk and progression of non-Hodgkin lymphoma.* Cancer Epidemiol Biomarkers Prev, 2013. **22**(4): p. 736-44.

393. Coomans de Brachene, A. and J.B. Demoulin, *FOXO transcription factors in cancer development and therapy.* Cell Mol Life Sci, 2016. **73**(6): p. 1159-72.

394. Liu, Z., et al., *Identification of Hub Genes and Key Pathways Associated with Two Subtypes of Diffuse Large B-Cell Lymphoma Based on Gene Expression Profiling via Integrated Bioinformatics.* Biomed Res Int, 2018. **2018**: p. 3574534.

395. Xie, L., T. Li, and L.H. Yang, *E2F2 induces MCM4, CCNE2 and WHSC1 upregulation in ovarian cancer and predicts poor overall survival.* Eur Rev Med Pharmacol Sci, 2017. **21**(9): p. 2150-2156.

396. Feliciano, A., et al., *miR-99a reveals two novel oncogenic proteins E2F2 and EMR2 and represses stemness in lung cancer.* Cell Death Dis, 2017. **8**(10): p. e3141.

397. Opavsky, R., et al., *Specific tumor suppressor function for E2F2 in Myc-induced T cell lymphomagenesis.* Proc Natl Acad Sci U S A, 2007. **104**(39): p. 15400-5.

398. Gao, Y., et al., *miR-155 regulates the proliferation and invasion of clear cell renal cell carcinoma cells by targeting E2F2.* Oncotarget, 2016. **7**(15): p. 20324-37.

399. Satwani, P., et al., *Upregulation of NKG2D ligands in acute lymphoblastic leukemia and non-Hodgkin lymphoma cells by romidepsin and enhanced in vitro and in vivo natural killer cell cytotoxicity.* Cytotherapy, 2014. **16**(10): p. 1431-40.

400. Almamun, M., et al., *Integrated methylome and transcriptome analysis reveals novel regulatory elements in pediatric acute lymphoblastic leukemia.* Epigenetics, 2015. **10**(9): p. 882-90.

401. Voshall, A., et al., *Identification of AGO3-associated miRNAs and computational prediction of their targets in the green alga Chlamydomonas reinhardtii.* Genetics, 2015. **200**(1): p. 105-21.

402. Chen, K. and N. Rajewsky, *Natural selection on human microRNA binding sites inferred from SNP data.* Nat Genet, 2006. **38**(12): p. 1452-6.

403. Calin, G.A., et al., *A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia.* N Engl J Med, 2005. **353**(17): p. 1793-801.

404. Wu, W.W., et al., *Robust Sub-nanomolar Library Preparation for High Throughput Next Generation Sequencing.* BMC Genomics, 2018. **19**(1): p. 326.

405. Reyes-Herrera, P.H., et al., *BackCLIP: a tool to identify common background presence in PAR-CLIP datasets.* Bioinformatics, 2015. **31**(22): p. 3703-5.

10. ANNEX

	Sample ID	Diagnosis	Biopsy	Preservation	Source	WGS (Discovery cohort)	Sanger Sequencing (Preliminary extended cohort)	Deep Targeted Resequencing (extended cohort)	Taqman Genotyping cohort
	S1FL	FL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S2FL1	FL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S2FL2	FL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S2FL3	FL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S3FL	FL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S4FL	FL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S5FL	FL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S6FL	FL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S1tFL	tFL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S2tFL	tFL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S3tFL	tFL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S4tFL	tFL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S5tFL	tFL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
	S6tFL	tFL	lymph node	Frozen	Barts Cancer Institute	Y	Y	Y	N
paired FL-tFL	09-5129	FL	lymph node	FFPE	Cruces University Hospital	N	Y	Y	N
	09-6449-1X	tFL	lymph node	FFPE	Cruces University Hospital	N	Y	Y	N
	09-9938A	FL	lymph node	FFPE	Cruces University Hospital	N	Y	Y	N
	09-23799	tFL	lymph node	FFPE	Cruces University Hospital	N	Y	Y	N
	05-.2226	FL	lymph node	FFPE	Cruces University Hospital	N	Y	Y	N
	11-15173	tFL	lymph node	FFPE	Cruces University Hospital	N	Y	N	N
	10-.2347	FL	lymph node	FFPE	Cruces University Hospital	N	Y	N	N
	10-17189-1	tFL	lymph node	FFPE	Cruces University Hospital	N	Y	N	N

04-15557	FL	lymph node	FFPE	Cruces University Hospital	N	Y	Y	N
13-4811	tFL	lymph node	FFPE	Cruces University Hospital	N	Y	N	N
FL1	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL2	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL3	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL4	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL5	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL7	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL9	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL10	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL11	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL12	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL13	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL14	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL15	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL16	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL23	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL25	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL26	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL34	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
FL35	FL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB1	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB2	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB3	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB4	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB5	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB7	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N

DB9	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB10	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB11	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB12	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB13	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB14	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB15	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB16	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB23	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB25	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB26	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB34	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
DB35	tFL	lymph node	Frozen	Barts Cancer Institute	N	N	Y	N
THD15T0024PF1	FL	Ovary	FFPE	Donostia University Hospital	N	N	Y	N
THD15T0025CM1	tFL	Submaxillary gland	Frozen	Donostia University Hospital	N	N	Y	N
1357	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
2613	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
1289	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
1287	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
2105	FL ?	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
1286	FL ?	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
586	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
994	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
2655	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
990	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
1754	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y

817	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
1710	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	N
623	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
1655	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
525	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
1640	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
175	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
1637	FL	lymph node	Frozen	Barts Cancer Institute	N	Y	N	Y
626	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
685	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
804	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1152	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1153	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1165	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
FL18	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1171	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1180	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1246	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1558	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
FL19	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
FL20	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1641	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1645	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1646	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1647	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
7260	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1666	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y

1668	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1669	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2663	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
6640	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1699	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1701	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1702	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1792	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1801	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1987	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2373	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2542	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
D01	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
7823	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
464	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
465	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
519	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
585	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
665	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1169	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1182	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1183	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1202	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1204	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1231	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1281	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1282	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y

1283	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1644	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1649	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1652	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1671	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1642	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1674	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1700	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1803	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1643	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2115	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2119	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2120	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2121	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2229	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2296	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2299	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1508	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2594	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2601	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2618	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
717	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
805	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
850	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1012	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1124	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1156	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y

1162	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1164	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1627	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1638	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2399	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2651	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2653	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1154	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1166	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1663	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1181	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1151	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1639	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
666	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1648	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1670	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1675	FL	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2094	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1122	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1177	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1636	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1665	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1667	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
2200	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
1800	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
849	FL Relapse	lymph node	Frozen	Barts Cancer Institute	N	N	N	Y
T120012	FL	lymph node	Frozen	Donostia University Hospital	N	Y	Y	N

	T120014	FL	lymph node	Frozen	Donostia University Hospital	N	Y	Y	N
	T130021	FL	lymph node	Frozen	Donostia University Hospital	N	Y	Y	N
	T130023	FL	lymph node	Frozen	Donostia University Hospital	N	Y	Y	N
	T130024	FL	lymph node	Frozen	Donostia University Hospital	N	Y	Y	N
	T130027	FL	Ileum	Frozen	Donostia University Hospital	N	Y	Y	N
	T130032	FL	lymph node	Frozen	Donostia University Hospital	N	Y	Y	N
	THD13T0034	FL	lymph node	Frozen	Donostia University Hospital	N	Y	Y	N
	THD13T0038	FL	lymph node	Frozen	Donostia University Hospital	N	Y	Y	N
	THD15T0010	FL	lymph node	Frozen	Donostia University Hospital	N	N	Y	N
	THD15T0022CM1	FL	lymph node	Frozen	Donostia University Hospital	N	N	Y	N
	THD15T0023CM1	FL	lymph node	Frozen	Donostia University Hospital	N	N	Y	N
	T130070.1	FL	lymph node	Frozen	Donostia University Hospital	N	N	Y	N
	THD15T0004	FL	lymph node	Frozen	Donostia University Hospital	N	N	Y	N
	THD15T0009	FL	lymph node	Frozen	Donostia University Hospital	N	N	Y	N
	THD15T0011	FL	lymph node	Frozen	Donostia University Hospital	N	N	Y	N
	THD15T0012	FL	lymph node	Frozen	Donostia University Hospital	N	N	Y	N
	THD15T0014	FL	lymph node	Frozen	Donostia University Hospital	N	N	Y	N
	THD15T0016	FL	lymph node	Frozen	Donostia University Hospital	N	N	Y	N
	THD15T0017	FL	lymph node	Frozen	Donostia University Hospital	N	N	Y	N
	THD15T0018	FL	lymph node	Frozen	Donostia University Hospital	N	N	Y	N
not-paired tFL	T130019	FL (pre-transf)	lymph node	Frozen	Donostia University Hospital	N	Y	Y	N
	THD13T0033	FL (pre-transf)	lymph node	Frozen	Donostia University Hospital	N	Y	Y	N
	THD13T0035	tFL (DLBCL)	lymph node	Frozen	Donostia University Hospital	N	Y	Y	N
	THD15T0026PF1	tFL (DLBCL + Burkitt)	Skin	FFPE	Donostia University Hospital	N	N	Y	N

DLBCL *de novo*	T120019	DLBCL	lymph node	Frozen	Donostia University Hospital	N	Y	N	N
	T130025	DLBCL	lymph node	Frozen	Donostia University Hospital	N	Y	N	N
	T130026	DLBCL	Skin	Frozen	Donostia University Hospital	N	Y	N	N
	T130028	DLBCL	lymph node	Frozen	Donostia University Hospital	N	Y	N	N
	T130030	DLBCL	Spleen	Frozen	Donostia University Hospital	N	Y	N	N
	THD13T0036	DLBCL	lymph node	Frozen	Donostia University Hospital	N	Y	N	N
	THD13T0037	DLBCL	lymph node	Frozen	Donostia University Hospital	N	Y	N	N
Lymphoma cell lines	FL-18	FL cell line	-	-	-	-	-	Y	N
	RIVA	DLBCL cell line	-	-	-	-	-	Y	N
	OCI-Ly10	DLBCL cell line	-	-	-	-	-	Y	N
	suDHL4	DLBCL cell line	-	-	-	-	-	Y	N
	suDHL10	DLBCL cell line	-	-	-	-	-	Y	N
	OCI-Ly1	DLBCL cell line	-	-	-	-	-	Y	N
	OCI-Ly8	DLBCL cell line	-	-	-	-	-	Y	N
	suDHL1	DLBCL cell line	-	-	-	-	-	Y	N

Table 1-Annex: Detailed information about each sample. This information includes the diagnosis, biopsied tissue, preservation of the sample, hospital and in what cohort the sample was included.

Chr	Position	Strand	miRNA	WT	Mutated	Gene	Location	Patient
18	60793447	-	hsa-mir-5008	G	GTCCAAA…	BCL2	3utr	S3_FL
18	60793448	-	hsa-mir-5008	G	AGGAGAG…	BCL2	3utr	S3_FL
18	60793641	-	hsa-mir-6071	A	C	BCL2	3utr	S3_tFL
18	60793653	-	hsa-mir-138	T	C	BCL2	3utr	S3_FL,S3_tFL
18	60793906	-	hsa-mir-27a	G	A	BCL2	3utr	S3_FL,S3_tFL
20	55745511	-	hsa-mir-1537	A	T	BMP7	3utr	S1_FL,S1_tFL
20	55745512	-	hsa-mir-1537	T	A	BMP7	3utr	S1_FL,S1_tFL
3	58703311	-	hsa-mir-1303	G	A	C3orf67	3utr	S3_FL,S3_tFL
3	58727863	-	hsa-mir-141	TTGAA	T	C3orf67	3utr	S1_FL,S1_tFL
3	58847570	-	hsa-mir-129	C	G	C3orf67	3utr	S1_tFL
7	106299866	-	hsa-mir-3685	A	G	RP5-892G19.1	3utr	S5_tFL
7	106299908	-	hsa-mir-28	T	A	RP5-892G19.1	3utr	S5_tFL
X	100518659	+	hsa-mir-3614	A	G	DRP2	3utr	S3_tFL
X	100518668	+	hsa-mir-3662	A	G	DRP2	3utr	S3_tFL
20	26064296	+	hsa-mir-17	T	G	FAM182A	3utr	S1_FL
20	26067008	+	hsa-mir-10a	C	A	FAM182A	3utr	S5_tFL
8	87480573	-	hsa-mir-376a	G	A	FAM82B	3utr	S5_tFL
8	87486431	-	hsa-mir-224	C	T	FAM82B	3utr	S2_FL2,S2_FL3
4	46923421	-	hsa-mir-5681a	A	C	GABRA4	3utr	S2_FL2,S2_FL3
4	46927520	-	hsa-mir-3157	G	A	GABRA4	3utr	S1_tFL
17	9818018	-	hsa-mir-3133	T	A	GAS7	3utr	S3_tFL
17	9820289	-	hsa-mir-3942	C	A	GAS7	3utr	S3_FL,S3_tFL
10	5807657	-	hsa-mir-15a	A	C	GDI2	3utr	S5_tFL
10	5807705	-	hsa-mir-1231	A	G	GDI2	3utr	S5_tFL
10	5807722	-	hsa-mir-33a	A	C	GDI2	3utr	S5_tFL
12	42480760	-	hsa-mir-1229	A	C	GXYLT1	3utr	S2_tFL

12	42480767	-	hsa-mir-181c	A	G	GXYLT1	3utr	S2_tFL
12	42480798	-	hsa-mir-320a	C	A	GXYLT1	3utr	S2_tFL
12	42480799	-	hsa-mir-320a	G	A	GXYLT1	3utr	S2_tFL
12	110606884	+	hsa-mir-1252	T	C	IFT81	3utr	S6_FL,S6_tFL
12	110606885	+	hsa-mir-1252	C	T	IFT81	3utr	S6_FL,S6_tFL
12	110606888	+	hsa-mir-3125	C	T	IFT81	3utr	S2_FL3
12	110606889	+	hsa-mir-3125	T	C	IFT81	3utr	S2_FL3
2	9546301	-	hsa-mir-146a	A	C	ITGB1BP1	3utr	S5_tFL
2	9546324	-	hsa-mir-1245b	GTGTT	G	ITGB1BP1	3utr	S5_tFL
18	47349754	-	hsa-mir-101	C	T	MYO5B	3utr	S4_tFL
18	47350291	-	hsa-mir-1323	C	A	MYO5B	3utr	S2_tFL
18	47350304	-	hsa-mir-144	C	T	MYO5B	3utr	S4_tFL
18	47350307	-	hsa-mir-144	G	A	MYO5B	3utr	S4_tFL
18	47350444	-	hsa-mir-4311	A	G	MYO5B	3utr	S4_tFL
18	47350494	-	hsa-mir-433	A	G	MYO5B	3utr	S4_tFL
4	30732884	+	hsa-mir-495	A	G	PCDH7	3utr	S1_FL,S1_tFL
4	30732983	+	hsa-mir-329	GTA	G	PCDH7	3utr	S3_tFL
9	78808412	+	hsa-mir-4287	A	C	PCSK5	3utr	S2_tFL
9	78975278	+	hsa-mir-5009	A	T	PCSK5	3utr	S4_tFL
1	160247610	-	hsa-mir-3189	T	C	PEX19	3utr	S1_FL
1	160247627	-	hsa-mir-125b	T	A	PEX19	3utr	S1_FL
2	197705581	-	hsa-mir-758	T	C	PGAP1	3utr	S4_FL,S4_tFL
2	197712238	-	hsa-mir-3134	T	A	PGAP1	3utr	S5_tFL
2	86365110	+	hsa-mir-3152	C	T	PTCD3	3utr	S2_FL3
2	86365781	+	hsa-mir-2355	AG	A	PTCD3	3utr	S1_tFL
4	57801657	+	hsa-mir-3129	A	T	REST	3utr	S5_FL
4	57801952	+	hsa-mir-1279	A	G	REST	3utr	S2_FL2

2	85569666	-	hsa-mir-146a	C	T	RETSAT	3utr	S5_FL
2	85570186	-	hsa-mir-1207	G	A	RETSAT	3utr	S3_FL
18	47350459	-	hsa-mir-3140	C	CTTTATGT	MYO5B	3utr	S4_tFL
18	47352742	-	hsa-mir-216b	T	G	MYO5B	3utr	S1_tFL
18	47352754	-	hsa-mir-2681	A	G	MYO5B	3utr	S1_tFL
19	55899833	+	hsa-mir-4761	G	A	RPL28	3utr	S3_tFL
19	55903097	+	hsa-mir-4265	A	G	RPL28	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
X	142711405	-	hsa-mir-3163	A	C	SLITRK4	3utr	S4_FL
X	142712252	-	hsa-mir-31	A	C	SLITRK4	3utr	S4_FL,S4_tFL
18	40848310	-	hsa-mir-3606	A	T	SYT4	3utr	S1_FL,S1_tFL
18	40848626	-	hsa-mir-1294	G	T	SYT4	3utr	S1_FL,S1_tFL
18	40848861	-	hsa-mir-371a	A	C	SYT4	3utr	S1_FL,S1_tFL
17	76849749	-	hsa-mir-105	A	C	TIMP2	3utr	S3_FL
3	171574589	+	hsa-mir-142	C	CTT	TMEM212	3utr	S6_FL
3	171574683	+	hsa-mir-4666a	G	A	TMEM212	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
7	19735117	-	hsa-mir-3140	T	C	TWISTNB	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
7	19735141	-	hsa-mir-1267	A	G	TWISTNB	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
19	44886986	-	hsa-mir-149	A	ATCAAATGAC...	ZNF285	3utr	S3_FL
19	44890193	-	hsa-mir-1225	T	C	ZNF285	3utr	S6_FL
17	66864089	-	hsa-mir-1295b	G	A	ABCA8	3utr	S3_tFL
10	101611732	+	hsa-mir-1910	A	G	ABCC2	3utr	S6_FL
X	153009482	+	hsa-mir-147a	C	G	ABCD1	3utr	S4_FL
2	204259279	+	hsa-mir-205	A	T	RAPH1	3utr	S2_FL2,S2_FL3
1	179072649	-	hsa-mir-4759	T	G	ABL2	3utr	S3_tFL
7	150558458	+	hsa-mir-654	G	A	ABP1	3utr	S4_FL,S4_tFL
2	37430689	+	hsa-mir-135a	C	A	AC007390.5	3utr	S2_FL2,S2_FL3
15	100255955	-	hsa-mir-3613	A	C	AC022692.1	3utr	S2_FL2,S2_FL3

2	132021476	+	hsa-mir-942	C	T	AC131180.2	3utr	S4_tFL
15	78463428	-	hsa-mir-155	G	A	ACSBG1	3utr	S1_FL,S1_tFL
14	105213399	+	hsa-mir-134	G	C	ADSSL1	3utr	S1_tFL
5	132215395	-	hsa-mir-6505	T	C	AFF4	3utr	S1_FL,S1_tFL
1	230838449	-	hsa-mir-181a	C	T	AGT	3utr	S4_FL,S4_tFL
1	243665741	-	hsa-mir-1238	C	G	AKT3	3utr	S2_tFL
1	16863809	+	hsa-mir-125a	C	T	AL355149.1	3utr	S4_tFL
6	35056661	+	hsa-mir-3163	A	G	ANKS1A	3utr	S2_FL3
4	80823431	-	hsa-mir-208a	A	C	ANTXR2	3utr	S1_FL,S1_tFL
10	75881949	-	hsa-mir-424	T	C	AP3M1	3utr	S5_FL,S5_tFL
22	36623188	-	hsa-mir-128	C	T	APOL2	3utr	S5_tFL
9	33441702	-	hsa-mir-146b	A	AC	AQP3	3utr	S6_FL
5	142604597	+	hsa-mir-4287	G	A	ARHGAP26	3utr	S6_FL
3	119139312	+	hsa-mir-3146	C	T	ARHGAP31	3utr	S5_FL
13	50205385	+	hsa-mir-126	C	A	ARL11	3utr	S5_FL
7	102739179	+	hsa-mir-222	A	G	ARMC10	3utr	S5_FL
12	27575606	+	hsa-mir-141	GT	G	ARNTL2	3utr	S3_FL,S3_tFL
X	69501023	+	hsa-mir-130a	C	T	ARR3	3utr	S6_tFL
5	151125317	-	hsa-mir-223	G	C	ATOX1	3utr	S2_FL3
1	203709122	+	hsa-mir-2682	A	C	ATP2B4	3utr	S1_tFL
20	50217352	-	hsa-mir-302b	C	A	ATP9A	3utr	S1_tFL
1	85732702	-	hsa-mir-495	A	C	BCL10	3utr	S4_FL,S4_tFL
14	99639429	-	hsa-mir-299	A	T	BCL11B	3utr	S2_FL2,S2_FL3
15	80253404	-	hsa-mir-122	T	C	BCL2A1	3utr	S2_tFL
6	136582252	-	hsa-mir-126	TCCTAA	T	BCLAF1	3utr	S4_FL
12	70048260	-	hsa-mir-135a	A	G	BEST3	3utr	S1_FL,S1_tFL
2	127805769	-	hsa-mir-1915	G	A	BIN1	3utr	S5_FL

14	54416456	-	hsa-mir-4724	T	G	BMP4	3utr	S1_FL,S1_tFL
2	203429766	+	hsa-mir-1273f	C	T	BMPR2	3utr	S1_tFL
19	15360063	-	hsa-mir-1299	C	T	BRD4	3utr	S3_tFL
4	75670630	-	hsa-mir-125b	A	C	BTC	3utr	S2_tFL
1	52553074	+	hsa-mir-148a	C	T	BTF3L4	3utr	S6_FL
10	104623993	+	hsa-mir-3682	C	T	C10orf32	3utr	S2_FL2
11	2317556	-	hsa-mir-3064	A	T	C11orf21	3utr	S3_FL
11	131529982	-	hsa-mir-4457	C	T	C11orf39	3utr	S1_FL,S1_tFL
11	109296866	+	hsa-mir-217	T	C	C11orf87	3utr	S1_tFL
11	111405233	+	hsa-mir-2115	A	G	C11orf88	3utr	S1_tFL
12	25146497	-	hsa-mir-1261	A	G	C12orf77	3utr	S1_FL,S1_tFL
1	226790202	+	hsa-mir-30a	T	C	C1orf95	3utr	S3_tFL
20	55093928	+	hsa-mir-1253	T	G	C20orf43	3utr	S2_FL2,S2_FL3,S2_tFL
21	34169475	+	hsa-mir-3189	G	C	C21orf49	3utr	S4_tFL
2	24253440	-	hsa-mir-302a	T	G	C2orf44	3utr	S3_tFL
4	128959426	+	hsa-mir-411	A	G	C4orf29	3utr	S4_tFL
6	170103788	+	hsa-mir-1180	G	A	C6orf120	3utr	S1_FL,S1_tFL
2	188210007	-	hsa-mir-2681	T	C	CALCRL	3utr	S1_FL,S1_tFL
11	34122120	+	hsa-mir-758	T	G	CAPRIN1	3utr	S6_FL,S6_tFL
1	19665861	-	hsa-mir-4779	A	G	CAPZB	3utr	S6_FL,S6_tFL
12	54626806	-	hsa-mir-29a	T	A	CBX5	3utr	S3_tFL
2	219883372	-	hsa-mir-639	T	A	CCDC108	3utr	S4_tFL
2	159027687	-	hsa-mir-1587	T	A	CCDC148	3utr	S6_FL,S6_tFL
9	15972581	+	hsa-mir-1292	C	A	C9orf93	3utr	S1_FL,S1_tFL
3	128758938	+	hsa-mir-1205	T	G	CCDC48	3utr	S2_FL2,S2_FL3
17	34198769	-	hsa-mir-758	G	A	CCL5	3utr	S1_tFL
17	72608386	-	hsa-mir-3120	G	C	CD300E	3utr	S2_FL1

11	60894894	+	hsa-mir-1914	C	A	CD5	3utr	S6_FL
17	7485372	+	hsa-mir-153	A	T	CD68	3utr	S6_FL
19	541881	+	hsa-mir-132	G	A	CDC34	3utr	S5_FL
18	63531068	+	hsa-mir-150	T	A	CDH7	3utr	S3_FL,S3_tFL
16	29870063	-	hsa-mir-1286	C	T	CDIPT	3utr	S6_tFL
19	42311427	+	hsa-mir-4797	T	C	CEACAM3	3utr	S5_FL
10	11376834	+	hsa-mir-4254	G	A	CELF2	3utr	S1_FL,S1_tFL
9	80887336	+	hsa-mir-4299	A	G	CEP78	3utr	S2_tFL
15	78916746	-	hsa-mir-146b	G	A	CHRNB4	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
10	73768236	+	hsa-mir-3184	C	T	CHST3	3utr	S4_FL,S4_tFL
4	103810130	+	hsa-mir-2682	GCAGA	G	CISD2	3utr	S4_tFL
8	121244027	+	hsa-mir-1290	C	A	COL14A1	3utr	S3_tFL
4	73923659	-	hsa-mir-106a	G	A	COX18	3utr	S1_FL,S1_tFL
20	6018638	+	hsa-mir-5007	G	C	CRLS1	3utr	S4_tFL
5	148875666	-	hsa-mir-4729	T	C	CSNK1A1	3utr	S6_FL
X	70836029	-	hsa-mir-185	A	C	CXCR3	3utr	S3_tFL
22	42523897	-	hsa-mir-1972	G	A	CYP2D6	3utr	S4_tFL
10	70744144	+	hsa-mir-4670	C	T	DDX21	3utr	S3_tFL
1	115126456	-	hsa-mir-548at	AAC	A	DENND2C	3utr	S3_tFL
1	244869281	+	hsa-mir-103a	T	A	PPPDE1	3utr	S3_FL
1	3801141	+	hsa-mir-10a	T	G	DFFB	3utr	S6_FL,S6_tFL
14	80665370	-	hsa-mir-5196	A	C	DIO2	3utr	S1_tFL
9	1056564	+	hsa-mir-382	T	G	DMRT2	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
18	67509881	+	hsa-mir-1272	A	T	DOK6	3utr	S1_FL
12	63953768	-	hsa-mir-1303	T	C	DPY19L2	3utr	S3_tFL
6	7586855	+	hsa-mir-3146	A	G	DSP	3utr	S2_tFL
1	1271039	-	hsa-mir-185	C	A	DVL1	3utr	S3_tFL

3	183889010	+	hsa-mir-222	C	G	DVL3	3utr	S6_FL,S6_tFL
1	184660899	-	hsa-mir-4279	T	G	EDEM3	3utr	S6_FL
5	106716920	-	hsa-mir-3667	G	T	EFNA5	3utr	S2_tFL
X	68061256	+	hsa-mir-4708	T	G	EFNB1	3utr	S6_FL
7	55278583	+	hsa-mir-1295b	C	T	EGFR	3utr	S5_tFL
8	117657097	-	hsa-mir-135a	G	A	EIF3H	3utr	S1_FL,S1_tFL
1	201985852	+	hsa-mir-1262	C	T	ELF3	3utr	S2_tFL
1	19545113	-	hsa-mir-495	A	G	KIAA0090	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
16	10623868	-	hsa-mir-4471	G	T	EMP2	3utr	S3_FL,S3_tFL
19	14845672	-	hsa-mir-4438	C	T	EMR2	3utr	S1_FL,S1_tFL
11	94863498	+	hsa-mir-3136	TG	T	ENDOD1	3utr	S2_tFL
17	77078803	+	hsa-mir-1228	A	C	ENGASE	3utr	S2_tFL
10	25273615	-	hsa-mir-4307	GT	G	ENKUR	3utr	S1_tFL
6	132215786	+	hsa-mir-1202	C	A	ENPP1	3utr	S1_FL,S1_tFL
9	111936168	-	hsa-mir-1290	A	G	EPB41L4B	3utr	S6_FL
X	71424805	-	hsa-mir-1292	G	A	ERCC6L	3utr	S6_FL
9	98730787	+	hsa-mir-185	A	C	C9orf102	3utr	S1_FL,S1_tFL
5	137843353	-	hsa-mir-106a	T	C	ETF1	3utr	S1_FL,S1_tFL
14	100604645	+	hsa-mir-2467	G	A	EVL	3utr	S5_FL
12	50261026	-	hsa-mir-296	G	A	FAIM2	3utr	S3_FL,S3_tFL
7	22982689	-	hsa-mir-593	C	G	FAM126A	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
8	139142643	-	hsa-mir-3163	T	TA	FAM135B	3utr	S2_FL2,S2_FL3,S2_tFL
2	70523109	-	hsa-mir-3150a	G	TATCAATTTC…	FAM136A	3utr	S6_FL
8	11280997	-	hsa-mir-4436b	C	G	FAM167A	3utr	S1_tFL
14	45543318	+	hsa-mir-3668	A	G	FAM179B	3utr	S5_FL
7	102449275	+	hsa-mir-1293	A	C	FAM185A	3utr	S4_tFL
6	24806326	-	hsa-mir-552	T	C	FAM65B	3utr	S1_FL,S1_tFL

10	90773889	+	hsa-mir-140	T	TA	FAS	3utr	S2_FL3
1	224346009	+	hsa-mir-1277	A	C	FBXO28	3utr	S4_tFL
11	33762796	-	hsa-mir-4255	T	C	FBXO3	3utr	S3_tFL
11	72548986	-	hsa-mir-4263	A	C	FCHSD2	3utr	S3_tFL
5	108532414	+	hsa-mir-1272	TTCA	T	FER	3utr	S5_tFL
7	76824492	-	hsa-mir-3162	A	C	FGL2	3utr	S6_FL
11	27018026	+	hsa-mir-3661	T	A	FIBIN	3utr	S3_FL
2	24286356	+	hsa-mir-124	A	G	FKBP1B	3utr	S3_tFL
14	86094878	+	hsa-mir-1238	G	A	FLRT2	3utr	S3_FL,S3_tFL
X	147030640	+	hsa-mir-495	G	C	FMR1	3utr	S4_FL,S4_tFL
2	114258286	+	hsa-mir-4263	TAGAA	T	FOXD4L1	3utr	S4_FL
17	80553855	+	hsa-mir-345	G	A	FOXK2	3utr	S1_FL,S1_tFL
13	41132162	-	hsa-mir-2861	C	T	FOXO1	3utr	S1_tFL
4	48500769	-	hsa-mir-4423	C	G	FRYL	3utr	S6_FL
5	160716287	-	hsa-mir-1179	T	C	GABRB2	3utr	S1_tFL
12	51747877	-	hsa-mir-146b	G	T	GALNT6	3utr	S1_FL,S1_tFL
3	128199008	-	hsa-mir-586	T	A	GATA2	3utr	S6_FL,S6_tFL
7	92086889	+	hsa-mir-3191	G	C	GATAD1	3utr	S1_tFL
1	89473256	-	hsa-mir-3148	A	C	GBP3	3utr	S6_FL
20	34022953	+	hsa-mir-155	A	T	GDF5OS	3utr	S3_tFL
5	37813104	-	hsa-mir-1234	T	TG	GDNF	3utr	S1_FL,S1_tFL
7	935183	+	hsa-mir-3195	G	A	GET4	3utr	S5_FL
5	179727693	-	hsa-mir-3613	A	C	GFPT2	3utr	S3_tFL
16	23476271	-	hsa-mir-1269a	CACACTTA...	C	GGA2	3utr	S5_tFL
6	121769806	+	hsa-mir-3145	ATATCAT	A	GJA1	3utr	S1_tFL
15	35044128	-	hsa-mir-518a	A	T	GJD2	3utr	S1_tFL
9	3827358	-	hsa-mir-3612	A	T	GLIS3	3utr	S4_FL,S4_tFL

17	9793627	+	hsa-mir-2278	G	A	GLP2R	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
4	175560784	-	hsa-mir-7	G	C	GLRA3	3utr	S3_tFL
2	70107498	+	hsa-mir-126	T	C	GMCL1	3utr	S1_tFL
18	11881569	+	hsa-mir-3202	C	G	GNAL	3utr	S1_tFL
7	93539407	+	hsa-mir-135a	G	C	GNGT1	3utr	S6_FL
4	44709057	-	hsa-mir-2054	G	T	GNPDA2	3utr	S1_FL,S1_tFL
10	113916242	-	hsa-mir-103a	T	A	GPAM	3utr	S4_FL,S4_tFL
1	202098358	+	hsa-mir-3614	A	G	GPR37L1	3utr	S2_tFL
12	13068992	+	hsa-mir-1263	C	T	GPRC5A	3utr	S6_FL
7	50659032	-	hsa-mir-127	CTAGGCTACG	C	GRB10	3utr	S1_FL
9	104331953	-	hsa-mir-186	A	G	GRIN3A	3utr	S2_FL3
4	7061934	-	hsa-mir-4517	A	G	GRPEL1	3utr	S5_tFL
10	1063493	+	hsa-mir-3676	G	C	GTPBP4	3utr	S5_FL,S5_tFL
4	156651457	+	hsa-mir-106a	A	C	GUCY1A3	3utr	S3_tFL
12	14930580	+	hsa-mir-3129	T	C	H2AFJ	3utr	S2_tFL
4	2231239	-	hsa-mir-1200	G	T	HAUS3	3utr	S4_FL,S4_tFL
1	236713460	-	hsa-mir-1244	T	C	HEATR1	3utr	S4_tFL
10	69682490	-	hsa-mir-218	C	A	HERC4	3utr	S6_FL,S6_tFL
8	80676728	-	hsa-mir-3129	T	A	HEY1	3utr	S3_tFL
6	26157317	+	hsa-mir-1276	C	G	HIST1H1E	3utr	S1_FL,S1_tFL
6	26123680	-	hsa-mir-3176	C	G	HIST1H2BC	3utr	S5_tFL
1	149859044	-	hsa-mir-1257	G	GGTGA	HIST2H2AB	3utr	S6_tFL
12	14923694	-	hsa-mir-1289	C	G	HIST4H4	3utr	S5_tFL
7	27202235	-	hsa-mir-376a	T	A	HOXA9	3utr	S2_tFL
12	54405190	+	hsa-mir-1234	C	T	HOXC8	3utr	S3_tFL
4	11398139	-	hsa-mir-6515	C	T	HS3ST1	3utr	S4_tFL
9	115232919	+	hsa-mir-1277	A	G	HSDL2	3utr	S5_FL

5	132441053	+	hsa-mir-141	G	T	HSPA4	3utr	S3_tFL
1	228369247	+	hsa-mir-3140	C	A	IBA57	3utr	S6_FL,S6_tFL
11	1756496	-	hsa-mir-1271	G	A	IFITM10	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
19	46623584	-	hsa-mir-346	T	G	IGFL3	3utr	S6_tFL
1	206941015	-	hsa-mir-194	C	T	IL10	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
7	1510196	-	hsa-mir-4443	G	A	INTS1	3utr	S4_FL
11	77635688	-	hsa-mir-4484	A	G	INTS4	3utr	S1_tFL
3	49763336	-	hsa-mir-3675	C	G	IP6K1	3utr	S3_tFL
5	131818706	-	hsa-mir-192	T	C	IRF1	3utr	S5_tFL
1	209959244	-	hsa-mir-1292	G	A	IRF6	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
5	50690217	+	hsa-mir-1275	C	A	ISL1	3utr	S5_FL
14	93404640	-	hsa-mir-1915	C	A	ITPK1	3utr	S3_tFL
1	65299513	-	hsa-mir-148b	A	C	JAK1	3utr	S6_FL
5	146970782	-	hsa-mir-4286	G	GT	JAKMIP2	3utr	S1_FL,S1_tFL
17	74710001	-	hsa-mir-1291	G	A	JMJD6	3utr	S6_tFL
8	75147151	-	hsa-mir-101	A	G	JPH1	3utr	S6_FL,S6_tFL
13	30777622	-	hsa-mir-106a	A	G	KATNAL1	3utr	S4_tFL
1	160009034	-	hsa-mir-1304	C	A	KCNJ10	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
1	233807805	+	hsa-mir-376c	T	C	KCNK1	3utr	S1_FL,S1_tFL
8	99443066	+	hsa-mir-4635	C	G	KCNS2	3utr	S3_tFL
16	29917422	-	hsa-mir-1207	A	C	ASPHD1	3utr	S2_FL2
7	66106891	+	hsa-mir-3614	G	C	KCTD7	3utr	S1_tFL
12	393639	-	hsa-mir-3166	T	C	KDM5A	3utr	S3_FL,S3_tFL
6	138663138	+	hsa-mir-1226	A	G	KIAA1244	3utr	S5_tFL
3	128689434	-	hsa-mir-1277	C	A	KIAA1257	3utr	S2_tFL
6	39302545	-	hsa-mir-3127	AGT	A	KIF6	3utr	S3_FL
4	55604964	+	hsa-mir-2115	C	A	KIT	3utr	S1_FL,S1_tFL

12	88889607	-	hsa-mir-3133	A	G	KITLG	3utr	S1_FL,S1_tFL
4	88083167	-	hsa-mir-3119	T	A	KLHL8	3utr	S4_tFL
3	122145587	-	hsa-mir-1290	G	T	KPNA1	3utr	S5_tFL
12	52862416	-	hsa-mir-2115	C	T	KRT6C	3utr	S1_FL,S1_tFL
21	46021936	+	hsa-mir-1263	A	G	KRTAP10-7	3utr	S3_FL
5	154197002	+	hsa-mir-300	G	A	LARP1	3utr	S3_tFL
15	59500519	+	hsa-mir-410	TTA	T	LDHAL6B	3utr	S4_tFL
7	127897316	+	hsa-mir-520d	T	C	LEP	3utr	S6_FL
X	111873998	-	hsa-mir-5011	A	C	LHFPL1	3utr	S6_FL
19	54744436	-	hsa-mir-139	G	A	LILRA6	3utr	S5_FL
7	156473795	-	hsa-mir-548ae	T	A	LMBR1	3utr	S6_FL
3	8609402	+	hsa-mir-1229	G	A	LMCD1	3utr	S5_FL
4	54326315	-	hsa-mir-146a	C	A	FIP1L1	3utr	S3_FL,S3_tFL
16	55617432	+	hsa-mir-10b	T	C	LPCAT2	3utr	S1_FL,S1_tFL
3	197612864	+	hsa-mir-2276	G	T	LRCH3	3utr	S1_FL,S1_tFL
3	120049715	-	hsa-mir-136	C	G	LRRC58	3utr	S3_tFL
2	80525445	-	hsa-mir-3125	T	G	LRRTM1	3utr	S4_FL,S4_tFL
7	1580362	+	hsa-mir-1227	G	T	MAFK	3utr	S3_FL
17	61771305	+	hsa-mir-1184	G	T	MAP3K3	3utr	S3_FL,S3_tFL
2	20192847	-	hsa-mir-30a	T	A	MATN3	3utr	S4_FL,S4_tFL
13	113751454	+	hsa-mir-1827	G	A	MCF2L	3utr	S5_tFL
11	28353434	+	hsa-mir-4313	G	A	METTL15	3utr	S4_tFL
12	31821887	+	hsa-mir-183	G	A	METTL20	3utr	S1_tFL
3	127409018	-	hsa-mir-1285	T	G	MGLL	3utr	S3_FL
1	67413810	+	hsa-mir-132	TG	T	MIER1	3utr	S1_tFL
10	27963059	-	hsa-mir-5700	C	T	MKX	3utr	S1_tFL
1	2522886	-	hsa-mir-1266	T	G	C1orf93	3utr	S4_FL

10	88696329	-	hsa-mir-1248	G	C	MMRN2	3utr	S1_tFL
1	158819031	+	hsa-mir-4684	A	G	MNDA	3utr	S5_FL,S5_tFL
22	50529515	+	hsa-mir-103a	A	G	MOV10L1	3utr	S2_FL1
1	11846026	-	hsa-mir-1245a	G	T	MTHFR	3utr	S1_tFL
6	74216870	+	hsa-mir-103a	G	C	MTO1	3utr	S1_FL
12	81102767	+	hsa-mir-141	C	T	MYF6	3utr	S4_FL
7	45002610	-	hsa-mir-1226	G	A	MYO1G	3utr	S5_tFL
19	17323594	+	hsa-mir-140	T	C	MYO9B	3utr	S5_FL
21	30244911	-	hsa-mir-1277	C	T	N6AMT1	3utr	S6_FL,S6_tFL
14	57877459	+	hsa-mir-3658	A	T	NAA30	3utr	S5_FL
11	2966359	-	hsa-mir-1273f	A	T	NAP1L4	3utr	S5_tFL
1	24933789	+	hsa-mir-4781	G	A	C1orf130	3utr	S2_FL2
6	41318457	+	hsa-mir-1225	T	G	NCR2	3utr	S6_FL
9	132996079	+	hsa-mir-3200	T	G	NCS1	3utr	S1_tFL
5	141533407	+	hsa-let-7a	G	A	NDFIP1	3utr	S2_tFL
8	125580414	+	hsa-mir-1255b	C	T	NDUFB9	3utr	S4_FL,S4_tFL
18	70411236	-	hsa-mir-154	A	G	NETO1	3utr	S2_FL1,S2_FL3,S2_tFL
16	69736839	+	hsa-mir-30a	C	G	NFAT5	3utr	S3_tFL
14	77732256	-	hsa-mir-4632	C	T	NGB	3utr	S2_tFL
4	164245890	-	hsa-mir-18a	GT	G	NPY1R	3utr	S6_FL
6	24146956	+	hsa-mir-182	T	A	NRSN1	3utr	S6_FL,S6_tFL
15	88403334	-	hsa-mir-380	G	T	NTRK3	3utr	S3_FL,S3_tFL
12	86276428	+	hsa-mir-943	T	G	NTS	3utr	S6_FL,S6_tFL
7	44425397	-	hsa-mir-4691	C	T	NUDCD3	3utr	S1_tFL
6	118031596	+	hsa-mir-129	A	G	NUS1	3utr	S2_tFL
X	123511018	-	hsa-mir-1	T	C	ODZ1	3utr	S2_FL3
X	70795324	+	hsa-mir-4705	A	G	OGT	3utr	S6_FL

11	132288788	-	hsa-mir-128	C	G	OPCML	3utr	S5_tFL
11	4702538	-	hsa-mir-1292	C	T	OR51E2	3utr	S5_FL,S5_tFL
12	50517503	+	hsa-mir-1273g	A	T	AC074032.1	3utr	S1_FL
18	21743050	-	hsa-mir-335	T	A	OSBPL1A	3utr	S1_FL,S1_tFL
8	90938856	+	hsa-mir-410	T	G	OSGIN2	3utr	S5_tFL
13	97637989	-	hsa-mir-340	A	G	OXGR1	3utr	S6_FL
16	50263455	+	hsa-mir-4446	A	G	PAPD5	3utr	S6_tFL
16	3022129	+	hsa-mir-1207	C	T	PAQR4	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
1	226548564	-	hsa-mir-1252	A	C	PARP1	3utr	S1_FL,S1_tFL
3	122356648	+	hsa-mir-374a	A	C	PARP15	3utr	S4_tFL
11	12550853	+	hsa-mir-4700	C	T	PARVA	3utr	S2_tFL
9	36840138	-	hsa-mir-4264	C	G	PAX5	3utr	S3_FL,S3_tFL
5	141323278	-	hsa-mir-1273e	G	C	PCDH12	3utr	S5_tFL
13	58302443	+	hsa-mir-130a	T	C	PCDH17	3utr	S1_FL,S1_tFL
5	140232167	+	hsa-mir-3613	A	T	PCDHA9	3utr	S6_FL
8	52731707	-	hsa-mir-3689a	A	G	PCMTD1	3utr	S3_FL
8	66634574	-	hsa-mir-4762	C	T	PDE7A	3utr	S1_FL
22	39620429	-	hsa-mir-3173	T	A	PDGFB	3utr	S3_tFL
X	153068647	-	hsa-mir-1266	G	A	PDZD4	3utr	S6_FL
16	21995228	-	hsa-mir-1278	C	CA	PDZD9	3utr	S5_tFL
15	77401915	-	hsa-mir-133a	G	C	AC087465.1	3utr	S3_tFL
7	94296494	+	hsa-mir-561	C	T	PEG10	3utr	S5_FL
X	77224046	-	hsa-mir-1204	T	TG	PGAM4	3utr	S2_tFL
19	3637102	-	hsa-mir-1184	C	G	PIP5K1C	3utr	S6_tFL
10	115541346	+	hsa-mir-3122	A	C	C10orf81	3utr	S6_FL
7	131812650	-	hsa-mir-802	C	T	PLXNA4	3utr	S3_tFL
10	118236708	+	hsa-mir-3190	A	G	PNLIPRP3	3utr	S6_FL

20	30805120	+	hsa-mir-3920	A	G	POFUT1	3utr	S5_FL,S5_tFL
3	121150668	-	hsa-mir-138	C	G	POLQ	3utr	S6_FL,S6_tFL
2	128605190	-	hsa-mir-1254	A	T	POLR2D	3utr	S1_FL,S1_tFL
7	72419066	+	hsa-mir-4311	C	CTG	POM121	3utr	S3_FL,S4_tFL
3	87308576	-	hsa-mir-1306	A	C	POU1F1	3utr	S3_FL,S3_tFL
5	149226197	+	hsa-mir-3124	A	G	PPARGC1B	3utr	S6_FL
10	133772780	+	hsa-mir-3191	G	C	PPP2R2D	3utr	S2_tFL
1	3351616	+	hsa-mir-1249	G	A	PRDM16	3utr	S3_FL,S3_tFL
10	54057140	+	hsa-mir-633	T	TA	PRKG1	3utr	S2_tFL
20	33764826	+	hsa-mir-2116	G	T	PROCR	3utr	S1_FL,S1_tFL
X	37315619	+	hsa-mir-103a	T	G	PRRG1	3utr	S4_FL
7	141537926	-	hsa-mir-1299	G	A	PRSS37	3utr	S1_tFL
16	31142966	-	hsa-mir-4436b	T	C	PRSS8	3utr	S5_FL,S5_tFL
19	43413365	-	hsa-mir-3157	T	C	PSG6	3utr	S1_tFL
1	79004088	+	hsa-mir-483	G	A	PTGFR	3utr	S2_tFL
9	112139908	-	hsa-mir-199a	G	C	PTPN3	3utr	S3_tFL
11	18749965	-	hsa-mir-1237	A	G	PTPN5	3utr	S4_FL,S4_tFL
6	128290472	-	hsa-mir-1284	A	T	PTPRK	3utr	S3_tFL
7	44920885	-	hsa-mir-323b	T	C	PURB	3utr	S4_tFL
19	45165618	+	hsa-mir-1233	C	T	PVR	3utr	S4_FL,S4_tFL
12	21622061	+	hsa-mir-103b	A	G	PYROXD1	3utr	S4_tFL
6	163989284	+	hsa-mir-1275	T	C	QKI	3utr	S1_FL,S1_tFL
19	11433019	-	hsa-mir-1289	A	C	RAB3D	3utr	S2_tFL
3	133544679	-	hsa-mir-3126	A	G	RAB6B	3utr	S2_FL2,S2_FL3,S2_tFL
15	63555876	+	hsa-mir-1262	A	C	RAB8B	3utr	S6_FL,S6_tFL
12	65089357	+	hsa-mir-3140	G	A	RASSF3	3utr	S6_tFL
1	89446130	-	hsa-mir-129	A	T	RBMXL1	3utr	S6_FL

1	173901940	-	hsa-mir-548an	AAAT	A	RC3H1	3utr	S4_FL
8	86573416	-	hsa-mir-105	A	G	REXO1L1	3utr	S4_tFL
6	153332308	-	hsa-mir-1271	G	C	RGS17	3utr	S2_FL2,S2_FL3
5	96496744	-	hsa-mir-3653	A	T	RIOK2	3utr	S3_tFL
13	51530854	+	hsa-mir-2053	A	T	RNASEH2B	3utr	S6_FL
11	65485939	-	hsa-mir-3616	A	C	RNASEH2C	3utr	S1_FL,S1_tFL
3	149679188	+	hsa-mir-200b	C	T	RNF13	3utr	S2_FL2
13	79189635	-	hsa-mir-3149	T	G	RNF219	3utr	S3_FL,S3_tFL
15	60787016	-	hsa-mir-4318	C	T	RORA	3utr	S1_FL,S1_tFL
8	105907686	+	hsa-mir-3692	G	T	RP11-127H5.1	3utr	S6_FL,S6_tFL
17	16679597	+	hsa-mir-3686	A	T	CCDC144A	3utr	S5_tFL
2	210885346	+	hsa-mir-3152	T	G	RPE	3utr	S4_tFL
17	62316	-	hsa-mir-4289	C	T	RPH3AL	3utr	S6_FL
19	49995219	+	hsa-mir-34a	G	T	RPL13A	3utr	S1_tFL
6	166825162	-	hsa-mir-106a	C	T	RPS6KA2	3utr	S6_FL
2	10270375	+	hsa-mir-188	A	G	RRM2	3utr	S1_FL,S1_tFL
6	43639843	+	hsa-mir-3190	C	T	RSPH9	3utr	S1_FL,S1_tFL
1	38078213	-	hsa-mir-147a	G	A	RSPO1	3utr	S3_FL,S3_tFL
3	127800007	-	hsa-mir-1238	C	A	RUVBL1	3utr	S1_FL,S1_tFL
6	147889639	+	hsa-mir-106a	G	T	SAMD5	3utr	S3_FL,S3_tFL
7	80373350	-	hsa-mir-200b	T	A	SEMA3C	3utr	S6_tFL
7	84625565	-	hsa-mir-5701	T	G	SEMA3D	3utr	S5_FL
12	48438318	-	hsa-mir-196a	A	G	SENP1	3utr	S5_tFL
15	72433236	+	hsa-mir-1179	T	C	SENP8	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
1	67874055	-	hsa-mir-150	C	G	SERBP1	3utr	S1_tFL
20	43127775	-	hsa-mir-6507	G	A	SERINC3	3utr	S5_FL
9	135138179	-	hsa-mir-1245b	A	C	SETX	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL

17	27282195	-	hsa-mir-3150b	A	C	SEZ6	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
2	223423656	+	hsa-mir-202	G	T	SGPP2	3utr	S1_tFL
X	80553667	+	hsa-mir-4662a	T	A	SH3BGRL	3utr	S3_tFL
18	43422407	+	hsa-mir-342	T	A	SIGLEC15	3utr	S3_FL
20	45188535	-	hsa-mir-1238	C	T	SLC13A3	3utr	S1_tFL
6	25783234	-	hsa-mir-197	A	C	SLC17A1	3utr	S6_FL,S6_tFL
14	37149553	-	hsa-mir-4709	A	G	SLC25A21	3utr	S2_FL2,S2_FL3,S2_tFL
12	105198038	-	hsa-mir-3653	A	G	SLC41A2	3utr	S6_FL
12	299853	-	hsa-mir-34a	A	G	SLC6A12	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
3	170184758	-	hsa-mir-1287	A	G	SLC7A14	3utr	S2_FL2,S2_FL3,S2_tFL
12	21421596	-	hsa-mir-3153	G	A	SLCO1A2	3utr	S1_FL
17	33594728	+	hsa-mir-139	G	T	SLFN5	3utr	S1_tFL
10	98760081	-	hsa-mir-1245b	T	C	ARHGAP19	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
17	38783119	-	hsa-mir-4649	G	T	SMARCE1	3utr	S5_FL
17	18167760	+	hsa-mir-4285	C	T	SMCR7	3utr	S6_FL,S6_tFL
16	18819330	-	hsa-mir-3150a	GC	G	SMG1	3utr	S2_FL1,S2_FL2,S2_tFL
9	33043722	-	hsa-mir-4316	G	GA	SMU1	3utr	S1_FL,S1_tFL
7	17922385	-	hsa-mir-200a	G	A	SNX13	3utr	S3_tFL
15	75949584	+	hsa-mir-17	C	T	SNX33	3utr	S1_tFL
10	107023790	+	hsa-mir-33a	C	T	SORCS3	3utr	S4_FL,S4_tFL
7	16501292	-	hsa-mir-330	AT	A	SOSTDC1	3utr	S1_FL,S1_tFL
6	21598724	+	hsa-mir-361	GA	G	SOX4	3utr	S3_FL,S3_tFL
12	23682820	-	hsa-mir-138	C	T	SOX5	3utr	S1_FL,S1_tFL
11	15993470	-	hsa-mir-4484	T	C	SOX6	3utr	S5_FL
11	124567404	+	hsa-mir-203a	A	C	SPA17	3utr	S6_FL,S6_tFL
X	57161584	-	hsa-mir-1291	A	C	SPIN2A	3utr	S2_FL1
4	167655217	-	hsa-mir-192	A	T	SPOCK3	3utr	S2_FL1,S2_tFL

2	139328316	+	hsa-mir-3974	A	G	SPOPL	3utr	S2_FL3
17	36687237	-	hsa-mir-1286	A	G	SRCIN1	3utr	S1_FL
3	156258306	-	hsa-mir-1827	A	G	SSR3	3utr	S5_FL
20	43708134	+	hsa-mir-1270	A	C	STK4	3utr	S3_tFL
9	93659684	+	hsa-mir-3064	A	G	SYK	3utr	S2_FL2,S2_FL3
6	86267774	-	hsa-mir-369	T	C	SNX14	3utr	S6_FL
6	158519568	+	hsa-mir-2355	T	G	SYNJ2	3utr	S3_FL
14	70839142	-	hsa-mir-1229	C	T	SYNJ2BP	3utr	S5_tFL
20	44004534	+	hsa-mir-1306	C	T	SYS1	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
12	11285839	-	hsa-mir-3130	A	T	TAS2R30	3utr	S3_tFL
14	92246290	-	hsa-mir-1277	T	A	TC2N	3utr	S4_tFL
6	35442031	-	hsa-mir-22	A	G	TEAD3	3utr	S1_FL,S1_tFL
7	97845855	-	hsa-mir-199a	T	C	TECPR1	3utr	S6_FL,S6_tFL
6	50811115	+	hsa-mir-1237	A	T	TFAP2B	3utr	S3_tFL
8	42692834	-	hsa-mir-548as	C	A	THAP1	3utr	S5_tFL
4	83825929	+	hsa-mir-615	G	A	THAP9	3utr	S3_FL,S3_tFL
10	25314525	+	hsa-mir-4653	G	T	THNSL1	3utr	S1_FL,S1_tFL
5	175386586	-	hsa-mir-371a	A	G	THOC3	3utr	S1_tFL
15	70340941	-	hsa-mir-4715	C	T	TLE3	3utr	S5_tFL
4	164441117	+	hsa-mir-3607	A	G	C4orf43	3utr	S6_FL,S6_tFL
1	93615674	-	hsa-mir-204	A	G	TMED5	3utr	S5_tFL
5	87497359	-	hsa-mir-4774	A	G	TMEM161B	3utr	S1_tFL
1	9664937	+	hsa-mir-15a	A	G	TMEM201	3utr	S5_tFL
14	67938137	-	hsa-mir-4261	A	C	TMEM229B	3utr	S1_FL,S1_tFL
9	104237051	-	hsa-mir-185	C	G	C9orf125	3utr	S1_tFL
12	29659238	-	hsa-mir-129	CT	C	TMTC1	3utr	S2_tFL
11	57505281	+	hsa-mir-1321	A	C	TMX2	3utr	S6_tFL

19	12810789	-	hsa-mir-3176	C	G	TNPO2	3utr	S1_tFL
7	128595317	-	hsa-mir-1289	T	C	TNPO3	3utr	S1_FL,S1_tFL
6	30153021	-	hsa-mir-1224	T	C	TRIM26	3utr	S1_FL,S1_tFL
6	25983973	+	hsa-mir-18a	T	G	TRIM38	3utr	S2_FL3
1	155155030	+	hsa-mir-4475	C	T	TRIM46	3utr	S6_FL
11	4620424	-	hsa-mir-214	T	A	TRIM68	3utr	S2_FL2,S2_FL3
9	135767598	-	hsa-mir-1178	C	A	TSC1	3utr	S3_tFL
2	54480995	-	hsa-mir-34a	A	T	TSPYL6	3utr	S6_tFL
14	89343803	+	hsa-mir-512	GA	G	TTC8	3utr	S2_tFL
6	167771419	+	hsa-mir-4496	T	G	TCP10	3utr	S6_FL
20	43120498	+	hsa-mir-3127	G	C	TTPAL	3utr	S1_tFL
6	30692754	+	hsa-mir-1302	C	CTT	TUBB	3utr	S3_FL
9	139978744	+	hsa-mir-1321	A	G	UAP1L1	3utr	S6_tFL
12	109972748	+	hsa-mir-512	G	A	UBE3B	3utr	S1_tFL
8	59361626	+	hsa-mir-1178	A	G	UBXN2B	3utr	S5_tFL
13	27641060	-	hsa-mir-4762	T	C	USP12	3utr	S5_tFL
7	6200245	+	hsa-mir-4797	A	T	USP42	3utr	S1_tFL
12	6572216	-	hsa-mir-4290	C	T	VAMP1	3utr	S5_tFL
12	48237047	-	hsa-mir-211	C	T	VDR	3utr	S3_FL,S3_tFL
15	91541842	-	hsa-mir-3137	G	A	VPS33B	3utr	S4_FL,S4_tFL
3	184770109	+	hsa-mir-224	C	T	VPS8	3utr	S4_FL,S4_tFL
6	110421523	-	hsa-mir-30a	A	G	WASF1	3utr	S4_FL,S4_tFL
X	54220266	-	hsa-mir-1245b	A	G	WNK3	3utr	S6_tFL
7	120980591	+	hsa-mir-4719	A	T	WNT16	3utr	S1_FL,S1_tFL
X	21876221	+	hsa-mir-448	A	G	YY2	3utr	S4_FL
18	72907534	-	hsa-mir-1229	T	C	ZADH2	3utr	S6_FL
12	57395205	-	hsa-let-7a	G	A	ZBTB39	3utr	S2_FL2,S2_FL3

22	41755337	+	hsa-mir-1261	T	A	ZC3H7B	3utr	S5_tFL
15	42708764	-	hsa-mir-3529	G	A	ZFP106	3utr	S5_tFL
8	144357600	+	hsa-mir-4632	C	G	GLI4	3utr	S2_FL1,S2_FL2,S2_FL3,S2_tFL
11	3380000	-	hsa-mir-1915	T	TCTATTCCTT...	ZNF195	3utr	S2_tFL
19	22152713	-	hsa-mir-193a	A	T	ZNF208	3utr	S2_tFL
19	58638270	-	hsa-mir-1234	T	G	ZNF329	3utr	S3_tFL
5	178510529	+	hsa-mir-34b	G	C	ZNF354C	3utr	S6_FL
X	134481773	+	hsa-mir-1305	G	C	ZNF449	3utr	S2_tFL
19	57093227	+	hsa-mir-1255a	A	G	ZNF470	3utr	S3_FL
19	9720640	-	hsa-mir-31	T	A	ZNF561	3utr	S6_FL
19	58291595	+	hsa-mir-203a	A	C	ZNF586	3utr	S3_tFL
19	12256162	-	hsa-mir-4633	T	C	AC022415.5	3utr	S5_FL
1	247132908	-	hsa-mir-3908	A	ATTCCACAGT...	ZNF670	3utr	S4_FL
1	71530387	-	hsa-mir-145	T	G	ZRANB2	3utr	S1_FL,S1_tFL
6	28402556	-	hsa-mir-1245b	A	C	ZSCAN23	3utr	S3_tFL
11	1912020	+	hsa-mir-4704	C	G	C11orf89	5utr	S4_tFL
17	7221413	-	hsa-mir-432	G	A	GPS2	5utr	S3_tFL
3	11667883	-	hsa-mir-1297	T	G	VGLL4	5utr	S3_FL,S3_tFL
15	25940037	-	hsa-mir-1307	G	A	ATP10A	coding	S5_tFL
15	25958929	-	hsa-mir-184	G	A	ATP10A	coding	S3_FL,S3_tFL
15	25958933	-	hsa-mir-184	G	A	ATP10A	coding	S1_FL,S1_tFL
7	148508727	-	hsa-mir-144	T	A	EZH2	coding	S1_FL,S1_tFL,S2_FL1,S2_FL2,S2_FL3,S2_tFL
7	148508764	-	hsa-mir-4793	T	C	EZH2	coding	S2_FL1,S2_FL2,S2_FL3,S2_tFL
17	37933932	-	hsa-mir-1256	TC	T	IKZF3	coding	S1_FL
17	37933935	-	hsa-mir-1256	T	G	IKZF3	coding	S1_FL
17	37933940	-	hsa-mir-19a	C	G	IKZF3	coding	S1_FL

19	19257600	-	hsa-mir-1911	G	C	MEF2B	coding	S1_FL,S1_tFL
19	19260045	-	hsa-mir-1265	T	A	MEF2B	coding	S1_FL,S1_tFL,S5_tFL
16	2569676	+	hsa-mir-1307	T	C	ATP6V0C	coding	S4_FL,S4_tFL
16	2569708	+	hsa-mir-1226	T	C	ATP6V0C	coding	S4_FL,S4_tFL
11	66130796	-	hsa-mir-1915	G	C	RP11-867G23.8	coding	S6_FL
11	66130901	-	hsa-mir-1254	G	A	RP11-867G23.8	coding	S2_FL1,S2_FL2,S2_tFL
17	76851902	-	hsa-mir-520a	G	A	TIMP2	coding	S5_FL,S5_tFL
16	2569631	+	hsa-mir-1228	T	A	ATP6V0C	coding	S4_FL,S4_tFL
15	45249109	+	hsa-mir-181b	C	T	C15orf43	coding	S1_FL,S1_tFL
17	40948320	-	hsa-mir-3921	GCT	G	WNK4	coding	S4_tFL
3	98250881	-	hsa-mir-1295b	G	A	GPR15	coding	S2_FL2,S2_tFL
2	242814094	+	hsa-mir-4529	C	A	C2orf85	coding	S1_FL,S1_tFL
1	110038447	+	hsa-mir-127	C	T	CYB561D1	coding	S1_tFL
3	121575963	+	hsa-mir-1270	T	C	EAF2	coding	S2_tFL
7	13935600	-	hsa-mir-3657	A	T	ETV1	coding	S1_FL,S1_tFL
1	171123340	+	hsa-mir-2355	A	G	FMO6P	coding	S2_FL2,S2_FL3
15	23686239	-	hsa-mir-1236	T	C	GOLGA6L2	coding	S4_FL
5	149421364	+	hsa-mir-1184	G	A	HMGXB3	coding	S1_FL,S1_tFL
9	34658640	+	hsa-mir-217	G	A	IL11RA	coding	S2_FL1,S2_FL2,S2_FL3,S2_tFL
19	18546106	-	hsa-mir-1301	C	T	ISYNA1	coding	S4_FL,S4_tFL
7	151927358	-	hsa-mir-153	T	TA	MLL3	coding	S3_tFL
22	37425350	+	hsa-mir-1976	G	A	MPST	coding	S2_FL2,S2_FL3,S2_tFL
9	100410462	+	hsa-mir-1256	C	G	NCBP1	coding	S1_tFL
12	18644413	+	hsa-mir-346	T	A	PIK3C2G	coding	S2_FL2,S2_FL3
17	8789893	-	hsa-mir-939	C	T	PIK3R5	coding	S4_FL,S4_tFL
8	18430069	-	hsa-mir-1207	A	T	PSD3	coding	S2_tFL
18	9815126	+	hsa-mir-4659a	C	T	RAB31	coding	S1_tFL

12	85255524	-	hsa-mir-3664	T	G	SLC6A15	coding	S1_FL,S1_tFL
9	102722410	+	hsa-mir-185	G	T	STX17	coding	S1_FL,S1_tFL
3	69075204	-	hsa-mir-1255b	G	A	TMF1	coding	S1_tFL
18	9887493	+	hsa-mir-2110	T	C	TXNDC2	coding	S6_FL
19	44738987	-	hsa-mir-3127	G	A	ZNF227	coding	S1_tFL
19	56953748	-	hsa-mir-127	C	A	ZNF667	coding	S1_FL,S1_tFL

Table 2-Annex: List of the total 544 mutations identified in predicted miRNA-binding sites in the WGS data.

PCSK5 c.7023A>AT; UTR. Target of hsa-mir-5009. VALIDATED

S4FL

S4FLt

BCL2 c.2672C>CT; UTR. Target of hsa-miR-27a. VALIDATED

S3FL

S3FLt

ATP10A c.2236C>CT; p.R746H. Target of miR-184-3p. VALIDATED

S3FL

S3FLt

ATP10A c.2232C>CT; p.S744S Target of miR-184-3p. VALIDATED

S1FL

S1FLt

SLC29A1 c.1017G>GA; UTR; Target of hsa-mir-1254. VALIDATED

S2FL1

S2FLt

SLC29A1 c.1122G>GC; UTR Target of miR-1915-3p. VALIDATED

S6FL

S6FLt

GDI2 c.1450T>TG; UTR. Target of hsa-mir-33a. VALIDATED

S5FL

S5FLt

GDI2 c.1515T>TG; UTR. Target of hsa-mir-15a. VALIDATED

S5FL

S5FLt

GDI2 c.1467T>TC; UTR. Target of miR-1231. VALIDATED

S5FL

S5FLt

EZH2 c.1937A>AT; p.Y646F; Target of miR-144-3p. VALIDATED

S1FL

S1FLt

EZH2 c.1937A>AT; p.Y646F; Target of miR-144-3p. VALIDATED

S2FL1

S2FLt

EZH2 c.1974A>AG; p.K634E; Target of hsa-mir-4793. VALIDATED

S2FL1

S2FLt

MEF2B, c.626G>GC; p.S209X Target of miR-1911-3p. VALIDATED

MEF2B, c.248A>AT; p.D83V Target of miR-1265 VALIDATED

MEF2B, c.248A>AT; p.D83V Target of miR-1265. VALIDATED

ATP10A c.3017C>CT; p.S1006N Target of miR-1307-5p. VALIDATED

PEX19 c.2614A>AG;UTR. Target of miR-125b-2-3p UNDETECTABLE

PEX19 c.2631A>AG;UTR. Target of miR-3189-5p UNDETECTABLE

MYO5B c.5634T>TC; UTR. Target of miR-2681-5p UNDETECTABLE

MYO5B c.5646A>AG; UTR. Target of miR-216b-5p UNDETECTABLE

MYOB5 c.8097G>GT; UTR. Target of hsa-mir-1323. UNDETECTABLE.
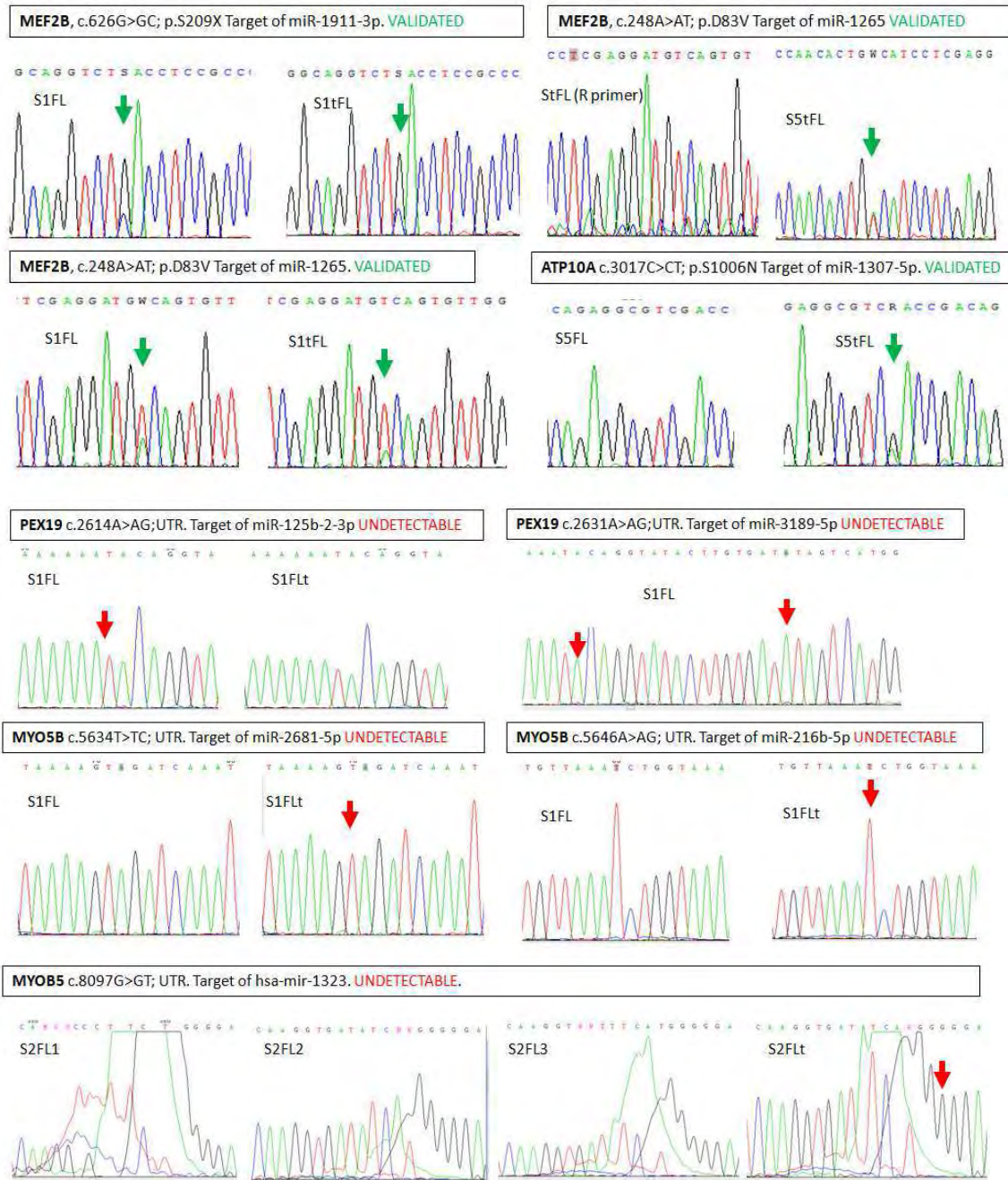
Figure 1-Annex: Results of the Sanger sequencing validation in the preliminary extended cohort.